TEL AVIV UNIVERSITY אוניברסיטת תל-אביב
The Raymond and Beverly Sackler Faculty of Exact Sciences
The Blavatnik School of Computer Science

Machine Learning Algorithms with Applications in Finance

Thesis submitted for the degree of Doctor of Philosophy

by

Eyal Gofer

This work was carried out under the supervision of

Professor Yishay Mansour

Submitted to the Senate of Tel Aviv University

March 2014

To my parents

Acknowledgements

I have had the good fortune to work with extraordinary scientists during my stint as a student at Tel Aviv University. First and foremost, I wish to thank my advisor, Professor Yishay Mansour, for sharing his vast knowledge and experience, and for setting such a clear example of excellence in research. I am also very grateful to Professor Nicolò Cesa-Bianchi and to Professor Claudio Gentile, with whom I had the honor and pleasure of collaborating, for providing that wonderful opportunity.

These memorable years at Tel Aviv University have been made all the more pleasant by the people at the school of computer science. In particular, I would like to thank my fellow students Mariano Schain, Shai Vardi, and Shai Hertz, for their friendly company throughout this time.

Finally, to my dear family, many thanks for everything.

Abstract

Online decision making and learning occur in a great variety of scenarios. The decisions involved may consist of stock trading, ad placement, route planning, picking a heuristic, or making a move in a game. Such scenarios vary also in the complexity of the environment or the opponent, the available feedback, and the nature of possible decisions. Remarkably, in the last few decades, the theory of online learning has produced algorithms that can cope with this rich set of problems. These algorithms have two very desirable properties. First, they make minimal and often worst-case assumptions on the nature of the learning scenario, making them robust. Second, their success is guaranteed to converge to that of the best strategy in a benchmark set, a property referred to as regret minimization.

This work deals both with the general theory of regret minimization, and with its implications for pricing financial derivatives.

One contribution to the theory of regret minimization is a trade-off result, which shows that some of the most important regret minimization algorithms are also guaranteed to have non-negative and even positive levels of regret for any sequence of plays by the environment. Another contribution provides improved regret minimization algorithms for scenarios in which the benchmark set of strategies has a high level of redundancy; these scenarios are captured in a model of dynamically branching strategies.

The contributions to derivative pricing build on a reduction from the problem of pricing derivatives to the problem of bounding the regret of trading algorithms. They comprise regret minimization-based price bounds for a variety of financial derivatives, obtained both by means of existing algorithms and specially designed ones. Moreover, a direct method for converting the performance guarantees of general-purpose regret minimization algorithms into performance guarantees in a trading scenario is developed and used to derive novel lower and upper bounds on derivative prices.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Consider an American company owner who employs workers in the United States and does most of his business in Europe. Naturally, the income of his company is mostly in euros. However, the company's expenses, mainly workers' salaries, are in dollars. As a result, any increase in the dollar-euro exchange rate decreases the income of the company in dollar terms, while the expenses remain the same. If the owner perceives a real chance for an appreciation of the dollar, he should address the risk of not being able to pay salaries on time.

How can this risk be mitigated? Setting aside enough dollars ahead of time would work, but it requires that the company have very large sums at its disposal, which might not be the case. There is obviously a need for some sort of insurance against an appreciation of the dollar. Such insurance exists in the form of an *option*.

Options are securities that allow their holders to buy or sell a certain asset for a given price at some future time. Thus, they act as insurance against a change in the value of an asset, in this case, dollars. A holder of an option on the dollar-euro exchange rate may buy a certain amount of dollars for a set price in euros at some future time. In return for this insurance contract, the company owner would need to pay some premium to the *option writer*, and with this payment his worries would be over.

The technical step of buying insurance clearly does not eliminate the risk. Rather, risk simply changes hands for a cost. This basic transaction also masks a profound problem, that of putting a price on the uncertainty associated with future events. Pricing options and more general types of securities is one aspect of this problem, and it is one of the primary concerns of this work.

Predicting future outcomes is a chief objective of statistics and machine learning. It is therefore reasonable to appeal to those disciplines for methods of coping with uncertainty. For the example above, it would seem natural to suggest a statistical model for the euro-dollar exchange rate that is based on past values. One might also employ sophisticated machine learning algorithms for predicting the future rate using any number of relevant financial indicators as input. Given such tools, one could hope to quantify the risk using a prediction of the exchange rate along with an estimate of the accuracy of the prediction.

There are, however, serious objections to this type of solution. The first one is that the very principle of inducing future behavior from past data, though firmly established in the natural sciences, may be called into question for some scenarios, including financial ones. In the famous words of [78], "The man who has fed the chicken every day throughout its life at last wrings its neck instead, showing that more refined views as to the uniformity of nature would have been useful to the chicken." In this example, the outcome of every day is clearly not randomly chosen, making prediction particularly difficult. Consider then another iconic example, where a large random sample of the swan population is used to estimate the probability of finding a black swan [78, 86]. Predicting the color of a random swan based on this estimate would be accurate with high probability. However, if a sufficiently high *cost* is associated with prediction mistakes, then the expected cost incurred by the prediction might still be unacceptably high. In particular, rare events might not appear in past samples but still have a significant effect on future costs.

A second objection to a statistical solution is that the risk may be successfully mitigated using methods that are entirely non-statistical. In our example, the option writer may eliminate future exchange rate uncertainty by changing euros for dollars immediately. This gives an upper bound on the option price that is independent of any statistical model for the exchange rate that one might come up with. In fact, this method of eliminating risk will work even if the future exchange rate is determined by an adversary and is not statistical at all. Such reasoning is clearly more robust and seems more justifiable than statistical modeling.

1.1 Arbitrage-Free Pricing

The method of setting aside enough dollars to cover future payments is a special case of a whole class of strategies. Let us re-examine the future obligation of the option writer. At a future time, she is obliged to provide a certain amount of dollars for a given amount of euros, and for making this commitment she is paid a fee at the present time. It is now up to the option writer to protect herself against future risk. This can be done not only by setting aside enough dollars immediately but also by trading in euros and dollars up until the day of payment (or option *expiration*). The key requirement of such trading strategies is that no matter how the market fares, the option writer would be left with enough money to cover her obligation to the option holder.

1.1.1 The Arbitrage-Free Assumption

How does the existence of such strategies relate to the price of an option? The answer is surprisingly straightforward. On the one hand, we have an option that guarantees its holder a future payment for a given price. On the other hand, we have a trading strategy that requires a one-time investment to set up and then continues to trade, *always* ending up with more money than the option pays. Thus, the option cannot cost more than the setup cost of the strategy, and this provides an upper bound on the price of the option. Similarly, if we had a strategy that always ended up with less money than the option's payoff, it would give us a lower bound on the option price.

The above statements rely on the intuitive assumption that if one asset is always worth more than another asset in some future time, then it must be worth more at the present time as well. If this were not the case, one might buy the cheaper asset and sell the more expensive asset at the present time, wait until the order of values is reversed, and make a riskless profit on the entire deal. Such a price anomaly, or *arbitrage opportunity*, cannot persist, because traders rush to buy the asset that is cheaper at present and sell the asset that is expensive at present. This process continues until the arbitrage opportunity vanishes. Thus, we rely on the *arbitrage-free assumption*, which is the sensible assertion that such obvious arbitrage opportunities cannot exist.

1.1.2 Regret Minimization

Arbitrage-free pricing requires a trading strategy whose returns always surpass the payment by the option writer. The method of buying dollars immediately is a trivial example of such a strategy, but it would do poorly if the dollar ended up depreciating against the euro. The position of the option writer would obviously be improved if she knew the exchange rate at the day the option expires. If she knew that the dollar would appreciate, her strategy would be to keep only dollars. On the other hand, if she knew that the euro would appreciate, she would keep only euros. Lacking clairvoyance, the option writer might attempt to find a trading strategy that would never fall too far behind the best of those two courses of action. In other words, she could seek to have minimal *regret* over not being able to pick the optimal course of action to start with. Such strategies may be devised within the theory of *online learning*.

1.2 Online Learning

Online learning is a major branch of modern machine learning, with roots going back to the works of Hannan [45] and Blackwell [12] in the 1950's. In a typical online learning scenario, a learner plays a game against an adversary over multiple rounds. Each of the players has a set of allowable actions, and on each round, both players pick their actions simultaneously. Once the actions are chosen, the learner suffers a *loss* which is some fixed function of the pair of actions. At the end of the game, the summed losses of the learner may be compared with the summed losses of playing a single fixed action throughout the game. The *regret* of the learner is the difference between its summed losses and the summed losses of playing the best fixed action. The goal of the learner is to guarantee small regret regardless of the actions taken by the adversary. The exact regret guarantees that may be achieved depend upon the specifics of the game. Nevertheless, for broad classes of games, a learner may always achieve per-round regret that tends to zero as the number of rounds goes to infinity, or *no-regret learning*.

This game formulation may be applied to the case of the option on the exchange rate. The rounds are trading periods, the learner (option writer) has to choose on every round a fraction of funds to be held in dollars, and the adversary (the market) "chooses" a change in the exchange rate. The loss in a single round is the logarithm of the ratio between the asset value of the learner before and after the round, which depends only on the pair of actions chosen on the round itself. The sum of the losses at the end of the game equals minus the logarithm of the relative gain of the learner. The regret is thus the logarithm of the ratio between the gain of the best currency and the gain of the learner. The smaller the regret guarantee achievable by some online learning algorithm, the tighter the upper bound one might get on the price of the option.

The fact that small regret may be guaranteed at all is remarkable and counter-intuitive, even for the simple game between the option writer and the market. It is, after all, impossible for any algorithm to know which of the two asset types would do better in the next round. The solution to this seeming paradox lies in two observations. First, the algorithm is compared to the best fixed asset, not to the best trading strategy with hindsight. Second, while the losses on each *single* round are unpredictable, it is impossible for any adversary to hide the *cumulative* performance of each asset. By allocating funds to each asset in correspondence to its cumulative gains, an online learning algorithm can reduce the fraction placed with a *consistently* bad asset and thus perform comparably to the best asset.

1.2.1 Specific Settings of Online Learning

The properties of an online learning game depend on the specific details of its decision sets, loss function, the exact nature of the information revealed to the learner in each round, and possibly other restrictions. Some important settings and modes are described below.

The *best expert* setting is perhaps the most widely researched type of online game. In this model, the adversary chooses a vector of real bounded values, and the learner chooses a probability vector of the same length. The loss of the learner is then defined as the dot product of the two vectors. The adversary's vector may be seen as the cost of following the advice of several experts (which the adversary may influence). The choice of the learner may be interpreted as a random choice of a single expert. The notion of 'experts' may be used to capture different heuristics, roads to take to work, advertisements to place on a web site, etc.

The *online convex optimization* setting is a strict generalization of the best expert setting. In this model, the learner decides on a point from a fixed convex and compact set in Euclidean space, and the adversary chooses a convex function from this set to the reals. The loss of the learner is the value of the adversary's function on the learner's point.

For both settings, an important distinction may be made regarding the feedback that the learner receives on each round. For example, for a set of experts predicting the weather, the learner has access to the losses of all experts, namely, *full information*. However, in the case of ad placement or choosing from a pool of slot machines, the learner knows only the loss of the choice it made. This feedback mode is known as the *multi-armed bandit* or simply *bandit* setting, in reference to the last example.

Finally, crucial distinctions may also be drawn with respect to the sequence of choices that are made by the adversary. The amount of fluctuation or variability in this sequence (which may be measured in various ways) affects the regret bounds the learner may achieve. A low level of variability has the intuitive effect of helping the learner track the adversary's moves. The work of the learner may also be facilitated by redundancy in the expert sets. Specifically, if there is only a small number of high-quality experts, or if many experts are near-duplicates of other experts, then better regret bounds may be achieved. These aspects and other variants of the online learning theme will be discussed in detail throughout this work, particularly in Chapters 3 and 4.

1.3 Competitive Analysis and Pricing

Recall the company owner who pulled through a euro devaluation using options on the exchange rate. The tide has turned, and an unexpected appreciation in the euro-dollar rate means that he has a surplus in euros. He intends to change these euros for dollars and invest in expanding his business. The exchange rate fluctuates continually, and he would naturally seek to make the change at the highest possible rate. This is the problem of *search for the maximum* or *one-way trading*.

Suppose the owner decides that he will make the change within the next month. Although he does not wish to guess exactly how the prices are generated, he is willing to make the reasonable assumption that the rate will not go above four times or below a quarter of the current rate. The owner then decides on the following strategy: He will change three quarters of the euros right away and another quarter only if the rate exceeds twice the current rate. If this event never occurs, he will simply sell the remaining euros at the end of the month.

Now, if the rate never reaches twice the current rate (suppose this rate is 1), the worst that can happen in terms of the owner's regret is that the rate goes to 2 and then plummets to $\frac{1}{4}$. In that case, the owner could get 2 with the benefit of hindsight, but he ends up with $\frac{3}{4} \cdot 1 + \frac{1}{4} \cdot \frac{1}{4} = \frac{13}{16}$, a ratio of $\frac{32}{13}$. If the rate exceeds 2 at some point, the worst scenario is that it goes all the way to 4, in which case the optimal rate is 4, but the actual sum obtained by the owner is $\frac{3}{4} \cdot 1 + \frac{1}{4} \cdot 2 = \frac{5}{4}$, a ratio of $\frac{16}{5}$, higher than the first case. However, compare this strategy to simply changing all the euros immediately. In case the rate shoots up to 4, the ratio between the optimal rate and the actual return is 4, higher than $\frac{16}{5}$, the highest ratio possible using the strategy chosen by the owner. Thus, using a simple strategy, the owner may improve his *competitive ratio*, which upper bounds the ratio between the outcome of the best possible strategy in hindsight and what his online strategy yields in practice [85].

There is another course of action available to the company owner beside one-way trading, which is to buy insurance against a rise in the value of the euro within the next month. Such insurance is available in the form of a *European fixed-strike lookback call option*. This type of security gives its holder the right to receive the difference between the *maximal* price of an asset over the lifetime of the option, and some pre-agreed price (the *strike price*). If the maximal price of the asset between the issuing of the option and its expiration rises above the strike price, the holder of the option may receive the difference. This dependence on the maximal, rather than the final, asset price constitutes the difference between this option and a standard *European call option*.

The option writer is again faced with the task of ensuring that she is protected in any possible contingency. She has to determine the amount of money required to cover her obligation and thereby know how much to charge for writing the option. Using a one-way trading strategy, the writer can ensure some minimal ratio between the exchange rate she will get eventually and the highest exchange rate by which she might be obliged to pay the holder. By using enough cash for one-way trading, she is guaranteed to be covered, given minimal assumptions, as mentioned in the above example.

However, this strategy ignores the threshold below which no payment takes place at all. If this threshold is very high, for example, a hundred times the current rate, it is practically impossible that any payment will take place. In contrast, if the threshold is zero, payment will always take place. In Chapter 7 it will be shown that the option writer can guard both against the change in rate *and* the event of crossing the threshold, by combining one-way trading and a regret minimization algorithm.

1.4 An Overview of Related Literature

1.4.1 Derivative Pricing in the Finance Literature

There is an immense body of work on derivative pricing in the financial literature, and only a very short glimpse may be offered here. The most influential works on this fundamental problem in finance are the seminal papers of Black and Scholes [11] and Merton [68] on pricing the European call option. Their pricing formula and model assume both an arbitrage-free market and a geometric Brownian motion stochastic process for stock prices. They show that changes in the option price may be exactly replicated by dynamically trading in the stock. By the arbitrage-free assumption, this trading strategy implies a pricing for the option. The Black-Scholes-Merton model has been used to price numerous types of derivatives (see, e.g., [57] for an exposition of derivative pricing). Of specific relevance to this work is the pricing of the fixed-strike lookback option, obtained in [30]. Another relevant work is that of [20], who show, among other things, how the prices of call options determine the price of any stock derivative whose payoff depends only on the final stock price.

The assumptions of The Black-Scholes-Merton model have long been known to conflict with empirical evidence. For example, actual prices exhibit discrete jumps rather than follow a continuous process as implied by the model. To account for that, various models, such as the jump-diffusion model of [69], have incorporated jumps into the stochastic process governing price changes. Empirical evidence shows also that asset prices do not in reality follow a lognormal distribution, as would be implied by a geometric Brownian motion process [67]. A further empirical inconsistency involves the "volatility smile" phenomenon, namely, that the prices of call options with different strike prices on the same asset imply different values for the volatility (standard deviation) parameter of the assumed Brownian process. These problems motivated much research into replacing Brownian motion with more general Lévy processes (see [29] and [79] for a coverage of Lévy processes and their uses in finance and pricing in particular).

While the predominant approach to derivative pricing in the financial community remains that of stochastic modeling, there are some results on robust, model-independent pricing. We mention here the works by Hobson et al., who priced various derivatives in terms of the market prices of call options with various strikes (see [54] for a review of results). These works assume an arbitrage-free market, but otherwise make only minimal, non-stochastic assumptions. Given a derivative, they devise a strategy that involves trading in call options and always has a payoff superior to that of the derivative; thus, the cost of initiating the strategy is an upper bound on the derivative's price. In the specific case of fixed-strike lookback options with zero strike [55], this strategy consists of one-way trading in call options, and the obtained price bound is shown to be tight in the assumed model.

1.4.2 Regret Minimization

Regret minimization research is primarily a creation of the last two or three decades, but its roots can be traced to 1950's works, which were motivated by problems in game theory.

Hannan [45] gave the first no-regret algorithm in the context of a repeated game, where a player wishes to approximate the utility of the best action with hindsight. He suggested that a player use a strategy of adding a random perturbation to the summed past utilities of each action and then choosing the action with the minimal perturbed utility. He then showed that the per-round regret of this strategy (now known generically as Follow the Perturbed Leader) tends to zero as the number of game rounds increases, regardless of the other players' strategies.

Blackwell's classic work on *approachability* [12] considered a generalization of a two-player zero-sum repeated game where the utility (equivalently, loss) matrix contains vector elements. For one-dimensional losses, von Neumann's minimax theorem implies a strategy whereby the minimizing player's average utility may arbitrarily approach the set of reals upper bounded by the value of the game. Blackwell's approachability theorem characterized the high dimensional convex and closed sets that are approachable in the scenario of vector utilities as those for which every containing half-space is approachable. Importantly, the proof gives a constructive algorithm with a convergence rate that is inverse proportional to the square root of the number of rounds. This result may be shown to yield as a special case a regret minimization algorithm for a two-player game, a problem of the type examined by Hannan.

The question whether Blackwell's results somehow stand apart from later work on regret minimization or are subsumed by it was recently answered by the work of [1]. These authors showed an equivalence between the approachability theorem and no-regret learning for a subset of the online convex optimization setting. Namely, any algorithm for a problem in Blackwell's setting may be efficiently converted to an algorithm for a problem in the online convex optimization setting with linear loss functions, and vice versa.

The best experts setting. The most well-known algorithm for this setting is the Hedge or Randomized Weighted Majority algorithm, which was introduced by several authors [38, 65, 89]. This algorithm gives to each expert a weight that decreases exponentially with its cumulative loss, and then normalizes the weights to obtain probability values. The rate of this exponential decrease may be controlled by scaling the cumulative loss with a numeric parameter, called the *learning rate*. This weighting scheme may be implemented by applying on each round a *multiplicative update* that decreases exponentially with the last single-period loss. This algorithm achieves a regret bound of $O(\sqrt{T \ln N})$ for any loss sequence, where T is the horizon, or length of the game, N is the number of experts, and the learning rate is chosen as an appropriate function of both. The bound is optimal since any online learner achieves an expected regret of the same order against a completely random stream of Bernoulli losses.

The above result holds whether or not the horizon is known to the learner; a bound on the horizon may be guessed, and whenever the guess fails the learner may double it and restart the algorithm (changing the learning rate in the process). This "doubling trick" technique [24, 90] changes the regret bound only by a multiplicative constant. An alternative approach for handling the case of an unknown horizon was given in [9], where a time-dependent learning rate was used, yielding a better constant in the bound.

The above bound, while optimal, ignores the actual characteristics of the loss sequences. This fact led to much subsequent research attempting to obtain improved regret bounds given various types of refined scenarios. A scenario in which the best expert has a small cumulative loss was already considered in [38]. It was shown that by choosing a different learning rate, one may obtain an improved bound, where the horizon is replaced by the cumulative loss of the best expert (barring an additive logarithmic factor). This result shares two features with the horizon-based bound. First, a bound on the cumulative loss of the best expert may be guessed using a doubling trick, rather than be known in advance. Second, a matching lower bound may be obtained using a random adversary (a trivial modification of the previous one).

Subsequent works considered replacing the dependence on the horizon with various quantities that measure the variation of the loss sequences. The first such result was given for the Polynomial Weights or Prod algorithm [25], which is a small but significant modification of Hedge. More concretely, the exponential multiplicative update of Hedge is replaced by its first order approximation, namely, a linear function of the last single-period loss. The regret bound for this algorithm has the same form as before, but the horizon is replaced by a known bound on the maximal *quadratic variation* of the losses of any expert, where the quadratic variation is defined as the sum of the squared single-period losses. Using more complicated doubling tricks, this bound on the quadratic variation may be guessed, resulting in a bound that features the maximal quadratic variation of best experts throughout the game. This regret bound, however, contains some additional factors that are logarithmic in the horizon.

The authors of [49] obtain an improved regret bound that depends on an adversarial counterpart of the variance of a random variable. As pointed out in [25], such a dependence is more natural, since for random losses the regret depends on the square root of the variance. Thus, their variation is defined for each expert as the sum of the squared deviations from the average loss of the expert, a quantity that is also necessarily smaller than the quadratic variation used in [25]. Their algorithm, Variation MW, modifies Hedge in a way that decreases the weights of experts exponentially with their variation as well as their cumulative loss. If a bound on the variation of the best experts throughout the game is known in advance, the regret bound of Variation MW substitutes that quantity for the time horizon in the standard regret bound. The authors show that this upper bound on the variation is not necessary, and using complex doubling tricks they obtain a regret bound that is worse by an additive factor logarithmic in the horizon.

More recently, the work of [28] considered a notion of variation appropriate for scenarios in which the single-period loss values for all experts experience only small changes from round to round. The maximal such change for each round is squared and summed over all rounds, defining a new notion of variation. These authors show that running Hedge with a learning rate based on their variation achieves a regret bound featuring this variation in place of the horizon. Thanks to a standard doubling trick, no prior knowledge of the variation is required. They show that their bound is optimal but it seems to be incomparable with the bound of [49].

Online convex optimization. The online convex optimization setting, introduced by Zinkevich [92], deals with sequential decision making problems where the learner chooses a decision from a compact and convex set in Euclidean space and the loss is the value of an adversarially chosen convex function applied to the decision. The regret of the learner is measured with respect to the best fixed decision. This setting generalizes the best experts setting, where the decisions are probability vectors and the loss functions are linear.[1] It also captures additional types of problems. One example is *portfolio selection*, where the learner decides on the proportions of funds to invest in various assets. Here the decision is still a probability vector, but the loss function is logarithmic rather than linear. Another is *online path planning* [59], where on each day a commuter chooses the road to follow from home to work and suffers a loss in the form of delay for each road segment. Here the decisions are vectors with 0/1 entries, rather than probability vectors, indicating whether each road segment is used or not. While a representation of this last problem as an experts problem is possible, with each feasible combination of road segments as an expert, it would require an exponential number of experts. Importantly, Zinkevich's more general problem setting also allows for the application of powerful tools from convex optimization in devising regret minimization algorithms.

For this general new setting, Zinkevich introduced the Online Gradient Descent algorithm and showed that it has vanishing regret. This algorithm as well as Hedge have been shown to be special cases of a more general meta-algorithm, Regularized Follow the Leader [47, 82]. This algorithm is a more subtle version of the strategy that greedily chooses at each time the decision that minimizes the current cumulative loss, or "Follow the Leader". More concretely, it adds a strongly convex regularizing function of the decision to the cumulative loss before minimizing the expression. It

[1] It should be clarified, however, that prediction with expert advice may be defined in a much more general way, which generalizes even online convex optimization. The best expert setting, which is considered here, is just a special case of this general setting, and is sometime referred to as "decision theoretic online learning". See [23] for a thorough exposition of the general theory of prediction with expert advice.

should be commented that the same algorithm, albeit using a random perturbation instead of regularization, had been applied in [59] to optimization problems that are not necessarily convex, but efficiently solvable.[2] Regularized Follow the Leader, in turn, is closely linked to the primal-dual approach of the Mirror Descent algorithm [70] (see [47], for example, for details).

The most basic regret bounds for Online Gradient Descent and more generally, for Regularized Follow the Leader, depend on the square root of the horizon [47, 92]. The constants depend on the diameter of the decision set and the properties of the loss and regularization functions. This dependence on the horizon is optimal in general for linear cost functions [46], a fact proved, as before, by invoking a stochastic adversary and lower bounding the expected regret. This bound can be dramatically improved if the loss functions are strictly convex with a bounded gradient, or more generally, exp-concave. In that case, a regret bound logarithmic in the horizon may be achieved by the Online Newton Step algorithm [51], based on the Newton method for optimization.

The authors of [49] show that in the case of linear loss functions (the dot product with a loss vector), the horizon in the regret bound may be replaced with a certain notion of variation. This variation is defined as the sum of the squared Euclidean distances between the loss vectors and their average. They prove their result for Regularized Follow the Leader used specifically with the squared Euclidean norm as a regularizer. A standard doubling trick negates any need for prior information regarding the magnitude of this variation. Using the same algorithm, a similar variational quantity may also replace the horizon in the logarithmic regret bound, for the special case where all the loss functions are a single convex function applied to arbitrary linear functions [48]. Here the vectors of the arbitrary linear functions are used to calculate the variation.

These results are strengthened by the authors of [28], who use a different type of variation, geared at scenarios where the gradients of the loss functions may change only little in Euclidean norm from round to round. These maximal normed differences are squared and summed over all rounds, yielding their variation. Importantly, for the scenarios considered in [48, 49], this new variation may be upper bounded in terms of the former variation, but can be much smaller in scenarios of gradual change. Using a modification of the Mirror Descent algorithm, the authors of [28] provide regret bounds that replace the horizon with their new variation both for linear costs (square

[2] As commented earlier, this method of random perturbation in turn originates with Hannan.

root regret bound) and for strictly convex losses satisfying some smoothness conditions (logarithmic regret bound).

The multi-armed bandit setting. The multi-armed bandit problem was first considered by Thompson [87], who dealt with the problem of deciding which of two medical treatments is superior, while also minimizing the harm to patients. It was also considered in [76] as an important instance of the general problem of sequential design of experiments in statistical theory. The motivation in both cases comes from the observation that sampling schemes that adapt to the observed data "on the fly" allow for better decision making than fixed schemes decided in advance. In the simplest multi-armed bandit instance described in [76], two biased coins are sampled with the purpose of maximizing the number of heads. If one of the coins is more biased toward heads than the other, it is clearly advantageous to quickly change the sampling policy so as to focus on that coin. A solution to this problem needs to strike an optimal balance between refining the probability estimates (exploration) and acting on them (exploitation).

This problem was generalized to multiple actions with general reward distributions and different types of constraints and has been the subject of much research. More recently, an adversarial version, namely, with an adversary controlling the rewards (equivalently, losses) on each round, was introduced [8]. In the adversarial setting the goal of maximizing the outcome needs to be replaced with the goal of minimizing the regret to the best action (or arm, in a multi-armed slot machine). This problem resembles the best expert setting with the crucial difference that only the loss of the chosen expert (arm) is revealed to the learner and not the full loss information. This difference reflects the situation in many practical scenarios, such as ad placement, where feedback is available only for the chosen action.

In what follows we will mention only selected results and focus primarily on the adversarial setting. For a recent comprehensive survey of the multi-armed bandit problem, see [21].

In the stochastic case, the goal may be formalized as minimizing the *pseudo-regret*, which is the difference between the expected reward of always using the best arm and the expected reward of the sampling policy. The seminal work of [64] introduced the method of maintaining an upper confidence bound for the expected reward of each arm, and always choosing the arm with the highest bound. They show that the pseudo-regret of their policy has asymptotic behavior that is logarithmic in the horizon, and that this dependence is optimal. The UCB1 algorithm of [7] provides a simple upper confidence bound policy that achieves a logarithmic pseudo-regret uniformly over time, rather than just asymptotically.

For the adversarial setting, the authors of [8] introduce the Exp3 algorithm, which is a natural modification of Hedge to the bandit setting.[3] The difference from Hedge stems from the fact that single-period rewards for an action are generally unknown, unless the action is the one chosen. Exp3 overcomes this problem by calculating unbiased estimates for these rewards. This algorithm is shown in [8] to achieve expected regret bounded by $O(\sqrt{TN \ln N})$, where the best possible value is shown to be $\Omega(\sqrt{TN})$ even for pseudo-regret.[4] Variants of Exp3 are shown to obtain this bound for an unknown horizon and achieve a similar bound for the actual regret with high probability.

The gap between the upper and lower bounds on the expected regret was closed by the INF algorithm [4] for oblivious (non-adaptive) adversaries. That algorithm generalizes Exp3 and was shown in [6] to be a special case of Mirror Descent.

The horizon in the Exp3 bound may be replaced by the cumulative reward of the best arm without prior knowledge of that quantity. This was shown in [8] for oblivious (non-adaptive) adversaries and later in [5] for general adversaries and with high probability.

The problem of replacing the horizon with variation was explored in [50]. They examine a generalization of the adversarial bandit problem to the online convex optimization setting with linear costs. The difference from the full information setting is that the learner does not observe the whole loss (or reward) vector after each round, but only the loss corresponding to the chosen decision, namely, the dot product of the decision and the loss vector. The notion of variation considered is the same as in [49], that is, the sum of the squared Euclidean distances between the loss vectors and their average. The authors give an algorithm whose expected regret is bounded by the square root of the variation but is also poly-logarithmic in the horizon. The algorithm is based on Regularized Follow the Leader, and replaces actual losses with loss estimates obtained using the technique of "reservoir sampling".

[3]The original formulation included an additional operation of smoothing the probabilities with a uniform distribution. A subsequent formulation, given in [21, Chapter 3], does not contain it.

[4]It can be shown that the expected regret is never smaller than the pseudo-regret.

1.4.3 Robust Trading and Pricing in the Learning Literature

In contrast to the finance literature, work on financial problems in machine learning has focused primarily on adversarial scenarios, rather than on stochastic ones. In addition, trading is seen primarily as taking place over discrete time periods. The focus of most research has been on devising portfolio selection algorithms with robust performance guarantees. Other results have dealt with pricing derivatives, assuming an arbitrage-free market, but otherwise attempting to keep assumptions about the market to a minimum. It should be stressed, though, that by virtue of the arbitrage-free assumption, trading strategies with robust performance guarantees may be used to obtain bounds on the prices of derivatives. This fact, which will be explained further below, means that the division between trading and pricing is not entirely clear cut.

Portfolio selection. The main financial problem dealt with in the learning literature is *portfolio selection*, where an online algorithm trades in several assets in an adversarial market with the goal of maximizing returns. The optimal algorithm for this problem is simple: On each round, invest all funds in the asset that gains the most in the coming round. However, this algorithm clearly depends on future information, and it is hopeless to compete against it. Instead, the task of maximizing the returns is replaced with that of minimizing the regret with respect to a set of benchmark strategies. A regret minimization algorithm is thus guaranteed to obtain returns comparable to the best strategy in the set. By choosing a rich and powerful set of benchmark investment strategies, one hopes this guarantee would also lead to good returns. It should be emphasized that in financial contexts, the appropriate notion of regret is measured by the *ratio* of the final wealths, rather than the customary difference, since we are interested in percentage returns.

The first such robust portfolio selection algorithm was given by Cover [31], where the benchmark was the set of all *constantly rebalanced portfolios*, namely, strategies that always keep a fixed fraction of funds in each asset. These strategies include holding the best single asset as a trivial case, but the best such strategy may outperform the best asset significantly. Cover's algorithm, the universal portfolio, is equivalent to initially dividing wealth uniformly between all possible constantly rebalanced portfolios, and performing no further action. He shows that the ratio between the final wealths of the best constantly rebalanced portfolio and his algorithm is upper bounded by a polynomial in the number of rounds. In other words, the wealths of the two algorithms have the same asymptotic growth rate. The computational complexity of the algorithm, however, is exponential in the number of assets. Cover's result prompted subsequent work that incorporated side information and transaction costs, proved that the regret of the universal portfolio is optimal, improved computational efficiency, and considered short selling [14, 32, 58, 73, 91]. The work of [52] tackled the problem with a different algorithm using a simple multiplicative update rule. Their algorithm had linear complexity in the number of assets, but gave worse regret bounds compared with the universal portfolio.

Portfolio selection was cast in the online convex optimization setting in the work of [3]. For this problem, the decisions are probability vectors describing the allocation of wealth to assets, and the single-period loss of an asset is minus the logarithm of its single-period price ratio. The single-period loss of the algorithm is minus the logarithm of the dot product of its decision and the asset price ratios. The authors of [3] show that the Online Newton Step algorithm, which is efficiently implementable, achieves regret logarithmic in the number of rounds. This dependence on the horizon is equivalent to that of Cover's algorithm, after standard (additive) regret is translated into multiplicative terms. Furthermore, the decision space is exactly all constantly rebalanced portfolios.

The variational results of both [48] and [28] are applicable to portfolio selection in its online convex optimization representation. The horizon in the logarithmic regret may thus be replaced with variation. In both cases the vectors featuring in the variation are the single-period price ratios of the assets (equivalently, the percentage returns). As noted before, the result of [28] is stronger in some sense, because their variation may be bounded in terms of the variation of [48], but not vice versa. These bounds are a great improvement over the horizon-dependent one under the realistic assumption that the variability is much smaller than the number of trading periods. Furthermore, experiments conducted with the algorithm of [48] show no real change in its regret as the trading frequency increases.

Benchmark sets other than constantly rebalanced portfolios have also been considered. The algorithms of [84] achieve bounded regret with respect to the best switching regime between several fixed investment strategies, with and without transaction costs. This benchmark was also considered in [62] for a portfolio containing two assets: stock and cash. Several results that will be mentioned further down in the context of derivative pricing may also be interpreted as choosing various other benchmark sets.

There are other approaches that rather seek to directly exploit the underlying statistics of the market [17, 44], but without assuming a specific price model. The authors of [44] show that their methods achieve the optimal asymptotic growth rate almost surely, assuming the markets are stationary and ergodic. Both these works do not, however, provide robust adversarial guarantees.

Several of the works mentioned also included experiments on real price data, and these highlight possible gaps in the current theoretical understanding. For example, the robust algorithms of [3, 52, 84] were shown to outperform the (optimal) universal portfolio on real data, with the latter two shown to even outperform the best constantly rebalanced portfolio. The algorithms of [17, 44], while having weak or no guarantees, were shown to achieve extraordinarily high yields.

Derivative pricing. In the Black-Scholes-Merton setup, derivatives are priced by devising a trading strategy that exactly replicates their payoff and applying the arbitrage-free assumption to derive an exact price. The same principle may still be used in an adversarial setting, modeled as a game between the market and an investor [80]. In adversarial settings, exact replication is not necessarily possible, but even then, the arbitrage-free assumption may be used to obtain price *bounds*. To obtain an upper bound, one requires a strategy whose payoff *super-replicates* (always dominates) the payoff of the derivative. The setup cost of the strategy is then an upper bound on the derivative's price. The same holds for *sub-replicating* strategies and lower bounds on the price.

The authors of [80] show that the Black-Scholes-Merton analysis may be extended to an adversarial setting assuming there exists a tradable *variance derivative*. This derivative pays periodic dividends equal to the squared relative change in the price of the stock. The investor's strategy involves trading in the stock as well as the derivative. Their analysis is applied to both discrete and continuous time, but it makes strong assumptions on the smoothness of the price of both the stock and the derivative.

The European call option was priced in a very general adversarial setting in the work of [35]. Their discrete-time model includes two parameters: a bound on the sum of the squared single-period returns of the stock (quadratic variation) and a bound on its absolute single-period returns. The quadratic variation serves as an adversarial counterpart of stochastic volatility. Apart from these two constraining parameters, the model is completely adversarial and allows for price jumps and dependence.

These authors upper bound the price of call options in two different ways. One converts a regret minimization algorithm for the best expert setting into a super-replication strategy and bounds its initial cost through a bound on the regret. The other directly calculates the minimal cost required to super-replicate the payoff of the option, thus obtaining an optimal arbitrage-free upper bound on the price in their model. This is done by producing a recursive expression for the minimax price and strategy for the investor, and showing that it can be approximated efficiently using dynamic programming. It should be commented that the trading strategy for the optimal bound is unrestricted with respect to taking loans and short selling the stock, while the regret minimization-based trading strategy requires neither.[5] The authors also obtain a lower bound on the price of the option with a specific strike price using a tailored sub-replicating strategy.

The optimal upper bound of [35] is only slightly worse than the Black-Scholes-Merton price, and demonstrates a volatility smile behavior as seen in practice. It is also shown that for some settings, the regret-based price depends on the square root of the quadratic variation (like the Black-Scholes-Merton price for small values of the quadratic variation), while their lower bound depends on it linearly (hence, suboptimally).

The regret minimization-based method of [35] will be pursued further and discussed in mathematical detail in the financial part of this thesis. It relies on the crucial observation that the robust pricing of a call option may be cast as a robust portfolio selection problem, where the assets are stock and cash, and the benchmark set is very simple: Hold cash, or hold the stock. Given a lower bound on the ratio between the final wealths of some algorithm and the best strategy in the set, the arbitrage-free assumption implies an upper bound on the price of a derivative that pays the same as the best strategy in the set. The payoff of this derivative and the payoff of a call option differ by only a fixed amount (the strike price) so the same holds for their prices.

It should be commented that the principle used in [35] for a very specific benchmark set may be applied to any benchmark set. Thus, for example, the results for the portfolio selection algorithms of [3, 28, 31, 48] imply arbitrage-free price upper bounds for a

[5]The regret minimization-based strategy does require that the investor possess the strike price in cash before trade begins.

(theoretical) derivative that pays the same as the best constantly rebalanced portfolio.

For the simple benchmark set of holding cash or holding the stock, the authors of [35] required a multiplicative version of an algorithm for the best expert setting, to be used with two experts: one holding cash and one holding the stock. They achieved this by modifying the Polynomial Weights algorithm [25] to obtain multiplicative, rather than additive, regret guarantees. This specific algorithm also has the great advantage of providing regret bounds (and resulting price bounds) that are based on variation.

The work of [34], which is extended in [33], considers pricing a class of American-style lookback options. More specifically, the option holder may choose any time to receive the payoff, which is some known non-negative right-continuous increasing function of the maximal price of the stock at the time of the payoff. They give an exact characterization of optimal strategies for super-replicating the payoff of the options, yielding an arbitrage-free price upper bound that is optimal in their fully-adversarial model. Their strategies are one-way trading, consisting of initially buying a specific amount of stocks and then selling fractions of the amount whenever the stock price reaches specific levels. They also consider extensions to payoff functions that depend on both the maximal price and the current price of the stock.

The methodology of [33] is applied in [63] to options whose payoff is defined in terms of the *upcrossings* of the price of the stock. More specifically, the option's payoff is the maximal value of some function applied to all the price intervals that the stock price crossed upwards. The functions considered take the end points of an interval as inputs and obey some natural regularity conditions.

It should be commented that the results of [33, 63] may also be seen through the lens of portfolio selection (with cash and stock as the assets), since the options guarantee a function of the best payoff of some benchmark set. In [33] it is the best one-way trading strategy, and in [63] it is the best "buy low, sell high" strategy.

Finally, the work of [2] returns to the problem of optimal minimax pricing, which was investigated in [35], and considers the minimax pricing of European options whose payoff is a convex and Lipschitz-continuous function of the final stock price. They show that under strong smoothness conditions on the stock price path, the optimal adversarial strategy for the market converges to a geometric Brownian motion. This implies that the Black-Scholes-Merton price, which is proved by assuming a specific stochastic price process, also holds for their adversarial model. It should be noted, though, that the limitations they place on the adversary do not allow for discontinuous jumps, which are allowed in [35]. It should also be commented that the investor is allowed unbounded loans and short selling, although the authors conjecture that this is not in fact necessary for obtaining their results.

1.4.4 Competitive Analysis and One-Way Trading

Competitive analysis was introduced in the classic work of [85]. It is concerned with obtaining robust performance guarantees for online algorithms with respect to the best offline algorithm, for miscellaneous scenarios (see [16] for a thorough exposition). In this framework, online algorithms are designed to minimize the *competitive ratio*, defined as the worst-case ratio between the performance of the best offline algorithm and the performance of the online algorithm. This is slightly different from the goal of regret minimization, which is to minimize the worst-case *difference* between the performance of the best algorithm in a *benchmark set* and the performance of the online algorithm. Still, despite the long-observed similarities between them, competitive analysis and online learning have developed mostly separately (historically and sociologically). For works that considered a unified framework, the reader is referred to [13] and [22].

The themes of this thesis intersect with competitive analysis on the single problem of one-way trading [36]. In a one-way trading scenario, one asset (such as euros) is converted to another (say, dollars), where the exchange rate fluctuates, and the goal is to maximize returns. The best offline algorithm simply sells the asset at the highest possible rate, and an online algorithm is required to achieve the best possible competitive ratio. This problem is equivalent to a search for the maximum of a series, since a gradual sale by the algorithm may be interpreted as a single selling point chosen randomly. The authors of [36] present an algorithm with an optimal competitive ratio for this problem, assuming upper and lower bounds on the exchange rate are known. They extend their results to cases where only the ratio of these bounds is known, and the horizon is possibly unknown.

It should be noted that for the problem of one-way trading, the gap between competitive analysis and regret analysis all but disappears. In fact, one-way trading may be seen as a special type of portfolio selection problem defined on two assets. The benchmark set is all the one-way trading strategies, or even simpler, strategies that sell the first asset in a single trade.

The work of [36] was related to option pricing in [66], which studied the search problem where the k highest or lowest values should be found. These authors applied their competitive analysis to pricing *floating-strike lookback calls*. This type of security pays the difference between the stock price at some future time, and the minimal stock price over the lifetime of the option.

1.5 Contributions in This Dissertation

This work contains contributions both to the general theory of regret minimization and to its financial applications.

The general theory part of this work deals both with the properties of existing regret minimization algorithms and with developing algorithms for new adversarial scenarios. We prove a surprising regret *guaranteeing* property for important families of regret minimization algorithms, that is, we show lower bounds on the regret for *any* individual sequence of losses. This fundamental property also has interesting pricing implications, as explained below. Another contribution deals with realistic cases in which the adversary does not exercise its full power, and for which general-purpose regret minimization algorithms are suboptimal. We characterize several such scenarios and provide optimal new algorithms for coping with them.

The financial part develops the theory of robust regret minimization-based pricing pioneered in [35]. This theory replaces the classic assumption of a very specific continuous stochastic price process (geometric Brownian motion) with a minimal and adversarial model that allows for sharp price jumps as well as price dependence and momentum, as occur in real markets. Price bounds obtained in this framework are arguably more robust than ones obtained with classic assumptions and are thus of considerable interest.

We apply regret minimization methods to pricing various derivatives. Of particular interest is the pricing of lookback options, which motivated the development of a new family of regret minimization algorithms that include a one-way trading element. Importantly, as in [35], our bounds on the prices of derivatives feature the variation of asset prices.

Most of the pricing results provided are upper bounds on derivative prices. However, we also provide the first optimal lower bound on the price of call options. This bound has the same asymptotic behavior as the stochastically-based Black and Scholes pricing and is derived by applying our new lower bounds on individual sequence regret. Another contribution required for deriving this result is a formula for directly converting general regret bound results to a trading scenario without the need for algorithmic adaptations and new proofs. This technical result and the above results serve to bring robust pricing and the general theory of regret minimization closer together.

1.5.1 Contributions to the Theory of Regret Minimization

Lower bounds on individual sequence regret. This contribution considers the problem of finding lower bounds on the regret that hold for *any* loss sequence. We focus on the online convex optimization setting with linear loss functions, and consider a general class of algorithms whose weight vector at time $t+1$ is determined as the gradient of a concave potential function of cumulative losses at time t. We show that this class includes all linear Regularized Follow the Leader regret minimization algorithms, which in turn include major regret minimization algorithms such as Hedge and Online Gradient Descent.

We first show that algorithms in this class are exactly those that guarantee non-negative regret for any sequence of losses. This result is surprising, given that this class includes algorithms specifically designed to uniformly *minimize* regret.

A much sharper trade-off result is obtained for the *anytime* regret, namely, the maximal regret during the game. We present a bound for the anytime regret that depends on the quadratic variation of the loss sequence, Q_T, and the learning rate. Nevertheless, we show that any learning rate that guarantees a regret upper bound of $O(\sqrt{Q_T})$ necessarily implies an $\Omega(\sqrt{Q_T})$ anytime regret on *any* sequence with quadratic variation Q_T.

We prove our result for the case of potentials with negative definite Hessians, and potentials for the best expert setting satisfying some natural regularity conditions. In the best expert setting, we give our result in terms of a translation-invariant version of the quadratic variation. We apply our lower bounds to Hedge and to linear cost Online Gradient Descent.

Algorithms for scenarios of limited expert set complexity. In many scenarios, assuming a fully malicious adversary may be unrealistically pessimistic, while more refined assumptions seem appropriate. In the best expert setting in particular, an adversary may be limited by various kinds of dependence or redundancy in the set of experts, making the "effective" number of experts smaller than the nominal number of experts. We thus study regret minimization bounds in which the dependence on the number of experts is replaced by measures of the *realized* complexity of the expert class. The measures we consider are defined in retrospect given the realized losses.

We concentrate on two interesting cases. In the first, our measure of complexity is the number of different "leading experts", namely, experts that were best at some point in time. We derive regret bounds that depend only on this measure, independent of the total number of experts. We also consider a case where all experts remain grouped in just a few clusters in terms of their realized cumulative losses. Here too, our regret bounds depend only on the number of clusters determined in retrospect, which serves as a measure of complexity. We show that our regret bounds are optimal by proving matching lower bounds on the expected regret of any algorithm against specially-constructed stochastic adversaries. Our bounds improve on the bounds of general-purpose algorithms such as Hedge for scenarios where the number of clusters or leaders is logarithmic in the number of experts. This means that by applying an extra level of regret minimization to choose between our specialized algorithms and the general-purpose ones, the best regret bound may be achieved (up to constant factors) regardless of the adversary's actual behavior.

Our results are obtained as special cases of a more general analysis for a novel setting of branching experts, where the set of experts may grow over time according to a tree-like structure, determined by an adversary. This setting is of independent interest since it is applicable in online scenarios where new heuristics (experts) that are variants of existing heuristics become available over time. For this setting of branching experts, we give algorithms and analysis that cover both the full information and the bandit scenarios.

1.5.2 Contributions to Derivative Pricing

Pricing a variety of options. We apply a unified regret minimization framework to pricing a variety of options based on the method of [35]. We give variation-based upper bounds on the prices of various known options, namely, the *exchange option*, the *shout option*, and several types of *Asian options*. We derive these bounds by considering a security whose payoff is the maximum of several derivatives. We price this security based on regret bounds with respect to the underlying derivatives and then show how to express the above options in terms of it.

Pricing convex path-independent derivatives. We give robust variation-based upper bounds on the prices of a wide range of derivatives that are characterized as *convex path-independent derivatives*. At their expiration time, these derivatives have a payoff that is a convex function of the price of a certain underlying asset. We derive robust upper bounds on the prices of these derivatives given a robust upper bound on the price of European call options.

Algorithms for pricing lookback options. We present a new family of regret minimization algorithms that combines two algorithmic components. One is a regret minimization component and the other is a one-way trading component, which may only sell assets. The algorithms in this family have regret guarantees not only with respect to the best asset (stock or cash, in this case) at the end of the game but also to the maximal value *throughout* the game. We translate the performance of the regret minimization and one-way trading components into explicit variation-based upper bounds on the price of fixed-strike lookback options. The resulting algorithms are, in general, two-way trading, that is, both buying and selling assets. These algorithms offer a new two-way trading solution to the problem of searching for the maximum of an online sequence studied in the work of [36]. Furthermore, we show that our methods may achieve better competitive ratios than the optimal one-way trading algorithm of [36].

A method for directly applying regret bounds to pricing. We develop a new formula that allows regret bounds to be directly converted to a financial setting. This method saves the need to specially modify existing best expert algorithms to work in a trading setup, where the regret is multiplicative in nature, rather than additive. As a result, existing regret bounds may be used, and there is no need to prove new ones for the modified algorithms.

We apply this method to obtain new variation-based price upper bounds that are applicable in broader settings than that of [35].

An optimal lower bound on the price of "at the money" call options. Based on regret minimization techniques, we give for the first time a robust lower bound on the price of "at the money" call options with the same asymptotic behavior as the robust upper bound given in [35]. (An "at the money" call option is simply one whose strike price is equal to the stock price when the option is issued.) We obtain this bound by combining our theoretic results concerning lower bounds on the anytime regret of individual sequences with our formula for converting regret bounds of best expert algorithms to the financial setting. This price bound has the same asymptotic behavior as the Black-Scholes-Merton price, namely, a dependence on the square root of the quadratic variation, despite the fact that our assumptions are minimal and adversarial, rather than stochastic.

1.6 Outline of This Thesis

This work in organized in two parts. The first part contains contributions to the theory of regret minimization per se. The second part deals with applications of regret minimization to pricing derivatives.

Part I is organized as follows. In Chapter 2 we give some required definitions and background. Chapters 3 and 4 then present two distinct works.

Chapter 3 presents variation-based regret lower bounds that hold for any sequence of losses, and apply for a wide range of learners in the online linear optimization setting. The results in this chapter were published in [42].

Chapter 4 is concerned with improved regret bounds for best expert scenarios with redundant sets of experts. These results are derived in a more general setting, where new experts are allowed to branch from existing experts. This chapter is based on work published in [43].

In Part II, Chapter 5 gives background and definitions. Chapter 6 applies the specially adapted regret minimization algorithm of [35] to derive variation-based upper bounds on the prices of a wide variety of derivatives. These results are based on work that appeared in [41].

Chapter 7 considers the pricing of lookback options in depth, and presents price upper bounds based on algorithms that combine regret minimization and one-way trading. These results are based on the work in [40].

In Chapter 8 we show how to directly translate bounds on the loss and regret of algorithms for the best expert setting into price bounds. This method is then used to prove a variation-based lower bound on the price of a certain call option, as well as new price upper bounds. This chapter is based in part on results in [42].

Part I

Regret Minimization

Chapter 2

Background and Model

This part of the work deals with pure theoretical aspects of regret minimization. Without further ado, we now formally present the basic definitions, facts, and notation to be used later on.

2.1 Regret Minimization Settings

The Best Expert Setting

The *best expert* setting is a simple and well-researched setting that exemplifies the main concepts of regret minimization. In this setting, there are N available experts (or actions), and at each time step $1 \le t \le T$, an online algorithm A (the *learner*) selects a distribution $\mathbf{p}_t$ over the N experts. Simultaneously, an adversary selects a loss vector $\mathbf{l}_t = (l_{1,t}, \ldots, l_{N,t}) \in \mathbb{R}^N$, and the algorithm experiences a loss of $l_{A,t} = \mathbf{p}_t \cdot \mathbf{l}_t$. We denote $L_{i,t} = \sum_{\tau=1}^{t} l_{i,\tau}$ for the cumulative loss of expert i at time t, $\mathbf{L}_t = (L_{1,t}, \ldots, L_{N,t})$, and $L_{A,t} = \sum_{\tau=1}^{t} l_{A,\tau}$ for the cumulative loss of A at time t. The *regret* of A at time T is $R_{A,T} = L_{A,T} - \min_j\{L_{j,T}\}$. The aim of a regret minimization algorithm is to achieve small regret regardless of the sequence of loss vectors chosen by the adversary. The *anytime regret* of A is the maximal regret over time, namely, $\max_t\{R_{A,t}\}$. We will sometimes use the notation $m(t) = \arg\min_i\{L_{i,t}\}$, where we take the smallest such index in case of a tie. We will also denote $L_t^* = L_{m(t),t}$ for the cumulative loss of the best expert after t steps.

We comment that it is customary to impose an additional explicit restriction on the range of single-period losses, namely, that $l_{i,t} \in [0, 1]$ for every i and t. This restriction will be indeed assumed in Chapter 4. However, in Chapter 3 no explicit restriction on losses will be assumed.

The distribution chosen by the learner is naturally interpreted as a random choice over the experts. Crucially, the learner makes this choice given *full information* of past losses incurred by every expert. The most notable algorithm for this setting is the Hedge algorithm [38, 65, 89], which is described below.

The Hedge Algorithm

Parameters: A learning rate $\eta > 0$ and initial weights $w_{i,1} > 0$, $1 \le i \le N$.

For each round $t = 1, \ldots, T$

1. Define probabilities $p_{i,t} = w_{i,t}/W_t$, where $W_t = \sum_{i=1}^{N} w_{i,t}$.
2. For each expert $i = 1, \ldots, N$, let $w_{i,t+1} = w_{i,t} e^{-\eta l_{i,t}}$.

Given a bound on the absolute value of the single-period losses $l_{i,t}$, Hedge may be shown to have bounded regret regardless of the loss sequences chosen by the adversary. If the learning rate η is tuned solely as a function of time, we achieve the so-called zero order regret bounds, which have the form $O(\sqrt{T} \ln N)$.

In contrast to the full information setting, the adversarial *multi-armed bandit* (or bandit) setting limits the learner's observations in every round of play only to the loss incurred by the expert it chose. The Exp3 algorithm [8], which is an adaptation of Hedge for this setting, obtains a zero order bound of $O\big(\sqrt{TN \ln N}\big)$ on the expected regret. Below we give a version of Exp3 taken from [21, Chapter 3]. This version, which is slightly simpler than the original one, achieves the same bound for the weaker notion of pseudo-regret, with a proper choice of η_t.

The Exp3 Algorithm

Parameters: A non-increasing sequence of real numbers $\eta_1, \eta_2, \ldots$.

Let $\mathbf{p}_1$ be the uniform distribution over $\{1, \ldots, N\}$.

For each round $t = 1, \ldots, T$

1. Draw an action I_t from the probability distribution $\mathbf{p}_t$.
2. For each action $i = 1, \ldots, N$, compute the estimated loss $\widetilde{l}_{i,t} = \frac{l_{i,t}}{p_{i,t}} \mathbb{I}\{I_t = i\}$ and update the estimated cumulative loss $\widetilde{L}_{i,t} = \widetilde{L}_{i,t-1} + \widetilde{l}_{i,t}$.
3. Compute the new distribution over actions $\mathbf{p}_{t+1} = \left(p_{1,t+1}, \ldots, p_{N,t+1}\right)$, where

$$p_{i,t+1} = \frac{\exp\left(-\eta_t \widetilde{L}_{i,t}\right)}{\sum_{k=1}^{N} \exp\left(-\eta_t \widetilde{L}_{k,t}\right)} .$$

The Online Linear Optimization Setting

The best expert setting is a special case of the more general setting of *online linear optimization*. In this setting, the online learning algorithm, or *linear forecaster*, chooses at time t a weight vector $\mathbf{x}_t \in \mathcal{K}$, where $\mathcal{K} \subset \mathbb{R}^N$ is a compact and convex set, and incurs a loss of $\mathbf{x}_t \cdot \mathbf{l}_t$. The regret of a linear forecaster A is then defined as $R_{A,T} = L_{A,T} - \min_{\mathbf{u} \in \mathcal{K}}\{\mathbf{u} \cdot \mathbf{L}_T\}$.[1] The best expert setting is simply the case where $\mathcal{K} = \Delta_N$, the probability simplex over N elements.

A no-regret result for the online linear optimization setting is achieved by the Regularized Follow The Leader (RFTL) algorithm, defined below.

Regularized Follow The Leader

Parameters: A learning rate $\eta > 0$ and a strongly convex regularizer function $\mathcal{R} : \mathcal{K} \to \mathbb{R}$.

Let $\mathbf{x}_1 = \arg\min_{\mathbf{x} \in \mathcal{K}}\{\mathcal{R}(\mathbf{x})\}$.

For each round $t = 1, \ldots, T$

Update the weight vector $\mathbf{x}_{t+1} = \arg\min_{\mathbf{x} \in \mathcal{K}} \{\mathbf{x} \cdot \mathbf{L}_t + (1/\eta)\mathcal{R}(\mathbf{x})\}$.

For a continuously twice-differentiable regularizer $\mathcal{R}$, RFTL guarantees, for a proper

[1] We comment that online linear optimization is in turn a special case of the online convex optimization setting. In that more general setting, the algorithm incurs a loss of $f_t(\mathbf{x}_t)$, where f_t is a convex function chosen by the adversary, and the regret is $\sum_t f_t(\mathbf{x}_t) - \min_{\mathbf{u} \in \mathcal{K}}\{\sum_t f_t(\mathbf{u})\}$.

choice of η, a regret bound of $O(\sqrt{\lambda DT})$, where $D = \max_{\mathbf{u}\in\mathcal{K}}\{\mathcal{R}(\mathbf{u}) - \mathcal{R}(\mathbf{x}_1)\}$ and $\lambda = \max_{t,\mathbf{x}\in\mathcal{K}}\{\mathbf{l}_t^\top[\nabla^2\mathcal{R}(\mathbf{x})]^{-1}\mathbf{l}_t\}$ [47].

The RFTL algorithm may be shown to generalize Hedge, where the regularizer function is chosen to be the negative entropy function, namely, $\mathcal{R}(\mathbf{x}) = \sum_{i=1}^N x_i \log x_i$. It also generalizes the Lazy Projection variant of the Online Gradient Descent algorithm (OGD), defined by the rule $\mathbf{x}_{t+1} = \arg\min_{\mathbf{x}\in\mathcal{K}}\{\|\mathbf{x}+\eta\mathbf{L}_t\|_2\}$, by choosing $\mathcal{R}(\mathbf{x}) = \frac{1}{2}\|\mathbf{x}\|_2^2$. For an in-depth discussion of the RFTL algorithm, see [47].

Quadratic variation. We define the *quadratic variation* of the loss sequence $\mathbf{l}_1,\ldots,\mathbf{l}_T$ as $Q_T = \sum_{t=1}^T \|\mathbf{l}_t\|_2^2$. For the best expert setting, we will use the slightly different notion of *relative quadratic variation*, defined as $q_T = \sum_{t=1}^T \delta(\mathbf{l}_t)^2$, where $\delta(\mathbf{v}) = \max_i\{v_i\} - \min_i\{v_i\}$ for any $\mathbf{v} \in \mathbb{R}^N$. Note that it always holds that $q_T \le 2Q_T$. We denote Q for a known lower bound on Q_T and q for a known lower bound on q_T.

2.2 Convex Functions

We mention here some basic facts about convex and concave functions that we will require. For more on convex analysis, see [77], [19], and [71], among others.

We will discuss functions defined on $\mathbb{R}^N$. A function $f : C \to \mathbb{R}$ is *convex*, if C is a convex set and if for every $\lambda \in [0,1]$ and $\mathbf{x},\mathbf{y} \in C$, $f(\lambda\mathbf{x}+(1-\lambda)\mathbf{y}) \le \lambda f(\mathbf{x})+(1-\lambda)f(\mathbf{y})$. The function f is *concave* if $-f$ is convex. The function f is *strictly convex* if the inequality is strict for $\mathbf{x} \ne \mathbf{y}$ and $\lambda \in (0,1)$. The function f is *strongly convex* with parameter $\alpha > 0$, if for every $\mathbf{x},\mathbf{y} \in C$ and $\lambda \in [0,1]$, $f(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) \le \lambda f(\mathbf{x}) + (1-\lambda)f(\mathbf{y}) - (\alpha/2)\lambda(1-\lambda)\|\mathbf{x}-\mathbf{y}\|_2^2$. If f is differentiable on a convex set C, then f is convex iff for every $\mathbf{x},\mathbf{y} \in C$, $\nabla f(\mathbf{y})\cdot(\mathbf{y}-\mathbf{x}) \ge f(\mathbf{y}) - f(\mathbf{x}) \ge \nabla f(\mathbf{x})\cdot(\mathbf{y}-\mathbf{x})$; f is strictly convex iff the above inequalities are strict for $\mathbf{x} \ne \mathbf{y}$. If f is twice differentiable, then it is convex iff its Hessian is positive semi-definite: for every $\mathbf{x} \in C$, $\nabla^2 f(\mathbf{x}) \succeq 0$; f is strongly convex with parameter α iff for every $\mathbf{x} \in C$, every eigenvalue of $\nabla^2 f(\mathbf{x})$ is at least α. The *convex conjugate* of f (defined on $\mathrm{dom} f$) is the function $f^*(\mathbf{y}) = \sup_{\mathbf{x}\in\mathrm{dom} f}\{\mathbf{x}\cdot\mathbf{y} - f(\mathbf{x})\}$, which is convex, and its effective domain is $\mathrm{dom} f^* = \{\mathbf{y} : f^*(\mathbf{y}) < \infty\}$.

Miscellaneous notation. For $\mathbf{x},\mathbf{y} \in \mathbb{R}^N$, we denote $[\mathbf{x},\mathbf{y}]$ for the line segment between $\mathbf{x}$ and $\mathbf{y}$, namely, $\{a\mathbf{x} + (1-a)\mathbf{y} : 0 \le a \le 1\}$. We use the notation $\mathrm{conv}(A)$ for the convex hull of a set $A \subseteq \mathbb{R}^N$, that is, $\mathrm{conv}(A) = \{\sum_{i=1}^{k} \lambda_i \mathbf{x}_i : \mathbf{x}_i \in A,\ \lambda_i \geq 0,\ i = 1, \ldots, k, \sum_{i=1}^{k} \lambda_i = 1\}$.

2.3 Seminorms

A *seminorm* on $\mathbb{R}^N$ is a function $\|\cdot\| : \mathbb{R}^N \to \mathbb{R}$ with the following properties:

- Positive homogeneity: for every $a \in \mathbb{R}$, $\mathbf{x} \in \mathbb{R}^N$, $\|a\mathbf{x}\| = |a|\|\mathbf{x}\|$.
- Triangle inequality: for every $\mathbf{x}$, $\mathbf{x}' \in \mathbb{R}^N$, $\|\mathbf{x} + \mathbf{x}'\| \leq \|\mathbf{x}\| + \|\mathbf{x}'\|$.

Clearly, every norm is a seminorm. A seminorm satisfies $\|\mathbf{x}\| \geq 0$ for every $\mathbf{x}$, and $\|\mathbf{0}\| = 0$. However, unlike a norm, $\|\mathbf{x}\| = 0$ does not imply $\mathbf{x} = \mathbf{0}$. We will not deal with the trivial all-zero seminorm. Thus, there always exists a vector with non-zero seminorm, and by homogeneity, there exists a vector with seminorm a for any $a \in \mathbb{R}^+$.

Chapter 3

Lower Bounds on Individual Sequence Regret

3.1 Introduction

For any sequence of losses, it is trivial to tailor an algorithm that has no regret on that particular sequence. The challenge and the great success of regret minimization algorithms lie in achieving low regret for *every* sequence. This bound may still depend on a measure of sequence smoothness, such as the quadratic variation or variance. The optimality of such regret upper bounds may be demonstrated by proving the existence of some "difficult" loss sequences. Their existence may be implied, for example, by stochastically generating sequences and proving a lower bound on the expected regret of any algorithm (see, e.g., [23]). This type of argument leaves open the possibility that the difficult sequences are, in some way, atypical or irrelevant to actual user needs. In this chapter we address this question by proving lower bounds on the regret of *any* individual sequence, in terms of its quadratic variation.

We first consider the related task of characterizing algorithms that have individual sequence *non-negative* regret. We focus our attention on the online linear optimization setting and on linear forecasters that determine their next weight vector as a function of current cumulative losses. More specifically, if $\mathbf{L}_t \in \mathbb{R}^N$ is the cumulative loss vector at time t, and $\mathbf{x}_{t+1} \in \mathcal{K}$ is the next weight vector, then $\mathbf{x}_{t+1} = g(\mathbf{L}_t)$ for some continuous g. We show that such algorithms have individual sequence non-negative regret if and only if g is the gradient of a concave potential function. We then show that this characteristic is shared by all linear cost Regularized Follow the Leader regret minimization algorithms, which include Hedge and linear cost OGD.

As our main result, we prove a trade-off between the upper bound on an algorithm's regret and a lower bound on its anytime regret, namely, its maximal regret for any prefix of the loss sequence. In particular, if the algorithm has a regret upper bound of $O(\sqrt{Q})$ for any sequence with quadratic variation Q, then it must have an $\Omega(\sqrt{Q})$ anytime regret on any sequence with quadratic variation $\Theta(Q)$.

We prove our result for two separate classes of continuously twice-differentiable potentials. One class has negative definite Hessians in a neighborhood of $\mathbf{L} = \mathbf{0}$, and includes OGD. The other comprises potentials for the best expert setting whose Hessians in a neighborhood of $\mathbf{L} = \mathbf{0}$ have positive off-diagonal entries; in other words, such potentials increase the weights of experts as their performance relatively improves, which is a natural property of regret minimization algorithms. For the first class, we use the quadratic variation Q_T. For the best expert setting, however, we use the more appropriate relative quadratic variation q_T.

We demonstrate our result on linear cost OGD, as an example of a potential with a negative definite Hessian, and on Hedge, as an example of a best expert potential. We comment that it will be shown in Part II, Chapter 8 that the bounds obtained for Hedge in turn imply lower bounds on the price of options.

Related work. As already observed, there are numerous results that provide worst case upper bounds for regret minimization algorithms and show their optimality (see, e.g., [23]). However, to the best of our knowledge, there are no results that lower bound the regret on loss sequences in terms of their individual quadratic variation. A regret trade-off result was given in [37] for the best expert setting; these authors showed that a best expert algorithm with $O(\sqrt{T})$ regret must have a worst case $\Omega(\sqrt{T})$ regret to any fixed average of experts.

3.2 Non-negative Individual Sequence Regret

Our ultimate goal is to prove strictly positive individual sequence regret lower bounds for a variety of algorithms. In this section, we will characterize algorithms for which this goal is achievable, and prove some basic regret lower bounds. This will be done by considering the larger family of algorithms that have *non-negative* regret for any loss sequence. This family, it turns out, can be characterized exactly, and includes the important class of linear cost Regularized Follow the Leader algorithms.

We focus on linear forecasters whose vector at time t is determined as $\mathbf{x}_t = g(\mathbf{L}_{t-1})$, for $1 \leq t \leq T$, where $g : \mathbb{R}^N \to \mathcal{K} \subset \mathbb{R}^N$ is continuous and $\mathcal{K}$ is compact and convex. For such algorithms we can write $L_{A,T} = \sum_{t=1}^{T} g(\mathbf{L}_{t-1}) \cdot (\mathbf{L}_t - \mathbf{L}_{t-1})$.

A *non-negative-regret algorithm* satisfies that $L_{A,T} \geq \min_{\mathbf{u}\in\mathcal{K}}\{\mathbf{u} \cdot (\mathbf{L}_T - \mathbf{L}_0)\}$ for every $\mathbf{L}_0, \ldots, \mathbf{L}_T \in \mathbb{R}^N$. We point out that we allow both positive and negative losses. Synonymously, we will also say that g has non-negative regret. Note that if $\mathbf{L}_0 = \mathbf{L}_T$ (a closed cumulative loss path), then for any $\mathbf{u} \in \mathcal{K}$, it holds that $\mathbf{u}\cdot(\mathbf{L}_T - \mathbf{L}_0) = 0$, and non-negative regret implies that $\sum_{t=1}^{T} g(\mathbf{L}_{t-1}) \cdot (\mathbf{L}_t - \mathbf{L}_{t-1}) \geq 0$. The following theorem gives an exact characterization of non-negative-regret forecasters as the gradients of concave potentials. This theorem is highly similar to Theorem 24.8 in [77], which regards *cyclically monotone mappings*.

Theorem 3.1. *A linear forecaster based on a continuous function g has individual sequence non-negative regret iff there exists a concave potential function $\Phi : \mathbb{R}^N \to \mathbb{R}$ s.t. $g = \nabla\Phi$.*

Proof. Let g have non-negative regret. For any path $\gamma : [0, 1] \to \mathbb{R}^N$ with finite length, we denote by $\int_\gamma \sum_{j=1}^{N} g_j dL_j$ the path integral of the vector field g along γ. For a closed path γ, we have that

$$\begin{aligned}\int_\gamma \sum_{j=1}^{N} g_j dL_j &= \lim_{n\to\infty} \sum_{i=1}^{n} g(\gamma((i-1)/n)) \cdot (\gamma(i/n) - \gamma((i-1)/n)) \\ &\geq \lim_{n\to\infty} \min_{\mathbf{u}\in\mathcal{K}}\{\mathbf{u} \cdot (\gamma(1) - \gamma(0))\} = 0 \ ,\end{aligned}$$

where the inequality is true because g has non-negative regret, and the last equality is true because $\gamma(1) = \gamma(0)$. For the reverse path γ_r, defined by $\gamma_r(t) = \gamma(1-t)$, we have $-\int_\gamma \sum_{j=1}^{N} g_j dL_j = \int_{\gamma_r} \sum_{j=1}^{N} g_j dL_j \geq 0$, where the inequality is true by the fact that γ_r is also a closed path. We thus have that $\int_\gamma \sum_{j=1}^{N} g_j dL_j = 0$ for any closed path, and therefore, g is a conservative vector field. As such, it has a potential function Φ s.t. $g = \nabla\Phi$.

Next, consider the case $\mathbf{L}_0 = \mathbf{L}$, $\mathbf{L}_1 = \mathbf{L}'$,$\mathbf{L}_2 = \mathbf{L}$, for some arbitrary $\mathbf{L}, \mathbf{L}' \in \mathbb{R}^N$. Since g has non-negative regret, we have that $0 \leq g(\mathbf{L}) \cdot (\mathbf{L}' - \mathbf{L}) + g(\mathbf{L}') \cdot (\mathbf{L} - \mathbf{L}') = (g(\mathbf{L}) - g(\mathbf{L}')) \cdot (\mathbf{L}' - \mathbf{L})$. Therefore, $(\nabla\Phi(\mathbf{L}) - \nabla\Phi(\mathbf{L}')) \cdot (\mathbf{L} - \mathbf{L}') \leq 0$ for any $\mathbf{L}, \mathbf{L}' \in \mathbb{R}^N$, which means that Φ is concave (see e.g. [71], Theorem 2.1.3).

In the other direction, let $\mathbf{L}_0, \ldots, \mathbf{L}_T \in \mathbb{R}^N$. We have that

$$\begin{aligned}\sum_{t=1}^{T} g(\mathbf{L}_{t-1}) \cdot (\mathbf{L}_t - \mathbf{L}_{t-1}) &= \sum_{t=1}^{T} \nabla\Phi(\mathbf{L}_{t-1}) \cdot (\mathbf{L}_t - \mathbf{L}_{t-1}) \geq \sum_{t=1}^{T} (\Phi(\mathbf{L}_t) - \Phi(\mathbf{L}_{t-1})) \\ &= \Phi(\mathbf{L}_T) - \Phi(\mathbf{L}_0) \geq \nabla\Phi(\mathbf{L}_T) \cdot (\mathbf{L}_T - \mathbf{L}_0) \\ &\geq \min_{\mathbf{u} \in \mathcal{K}} \{\mathbf{u} \cdot (\mathbf{L}_T - \mathbf{L}_0)\} ,\end{aligned}$$

where the first two inequalities are by the concavity of Φ. □

As a by-product of the proof, we get the following:

Corollary 3.2. *If algorithm A uses a non-negative-regret function $g(\mathbf{L}) = \nabla\Phi(\mathbf{L})$, then it holds that $R_{A,T} \geq \Phi(\mathbf{L}_T) - \Phi(\mathbf{L}_0) - \min_{\mathbf{u}\in\mathcal{K}}\{\mathbf{u} \cdot (\mathbf{L}_T - \mathbf{L}_0)\} \geq 0$.*

If Φ is continuously twice-differentiable, its second order Taylor expansion may be used to derive a similar lower bound, which now includes a non-negative quadratic regret term.

Theorem 3.3. *Let A be an algorithm using a non-negative-regret function $g = \nabla\Phi$, and let $\mathbf{L}_0, \ldots, \mathbf{L}_T \in \mathbb{R}^N$. If Φ is continuously twice-differentiable on the set $\mathrm{conv}(\{\mathbf{L}_0, \ldots, \mathbf{L}_T\})$, then*

$$R_{A,T} = \Phi(\mathbf{L}_T) - \Phi(\mathbf{L}_0) - \min_{\mathbf{u}\in\mathcal{K}}\{\mathbf{u} \cdot (\mathbf{L}_T - \mathbf{L}_0)\} - \frac{1}{2}\sum_{t=1}^{T} \mathbf{l}_t^\top \nabla^2\Phi(\mathbf{z}_t)\mathbf{l}_t ,$$

where $\mathbf{z}_t \in [\mathbf{L}_{t-1}, \mathbf{L}_t]$.

Proof. For every $1 \leq t \leq T$, we have $\Phi(\mathbf{L}_t) - \Phi(\mathbf{L}_{t-1}) = \nabla\Phi(\mathbf{L}_{t-1}) \cdot \mathbf{l}_t + (1/2)\mathbf{l}_t^\top \nabla^2\Phi(\mathbf{z}_t)\mathbf{l}_t$ for some $\mathbf{z}_t \in [\mathbf{L}_{t-1}, \mathbf{L}_t]$. Summing over $1 \leq t \leq T$, we get

$$\begin{aligned}\Phi(\mathbf{L}_T) - \Phi(\mathbf{L}_0) &= \sum_{t=1}^{T} \nabla\Phi(\mathbf{L}_{t-1}) \cdot \mathbf{l}_t + \frac{1}{2}\sum_{t=1}^{T} \mathbf{l}_t^\top \nabla^2\Phi(\mathbf{z}_t)\mathbf{l}_t \\ &= L_{A,T} + \frac{1}{2}\sum_{t=1}^{T} \mathbf{l}_t^\top \nabla^2\Phi(\mathbf{z}_t)\mathbf{l}_t .\end{aligned}$$

Therefore,

$$\begin{aligned} R_{A,T} &= L_{A,T} - \min_{\mathbf{u}\in\mathcal{K}}\{\mathbf{u}\cdot(\mathbf{L}_T - \mathbf{L}_0)\} \\ &= \Phi(\mathbf{L}_T) - \Phi(\mathbf{L}_0) - \min_{\mathbf{u}\in\mathcal{K}}\{\mathbf{u}\cdot(\mathbf{L}_T - \mathbf{L}_0)\} - \frac{1}{2}\sum_{t=1}^{T}\mathbf{l}_t^\top \nabla^2\Phi(\mathbf{z}_t)\mathbf{l}_t \,. \end{aligned}$$

□

Note that the regret is the sum of two non-negative terms. We will sometimes refer to the first one, $\Phi(\mathbf{L}_T) - \Phi(\mathbf{L}_0) - \min_{\mathbf{u}\in\mathcal{K}}\{\mathbf{u}\cdot(\mathbf{L}_T - \mathbf{L}_0)\}$, as *the first order regret term*, and to the second, $-\frac{1}{2}\sum_{t=1}^{T}\mathbf{l}_t^\top \nabla^2\Phi(\mathbf{z}_t)\mathbf{l}_t$, as *the second order regret term*. The second order term is non-negative by the concavity of Φ. These two terms play a key role in our bounds.

3.2.1 Relation to Regularized Follow the Leader

The class of concave potential algorithms contains the important class of linear cost Regularized Follow the Leader algorithms. The linear forecaster $RFTL(\eta, \mathcal{R})$ determines its weights according to the rule $\mathbf{x}_{t+1} = g(\mathbf{L}_t) = \arg\min_{\mathbf{x}\in\mathcal{K}}\{\mathbf{x}\cdot\mathbf{L}_t + \mathcal{R}(\mathbf{x})/\eta\}$. Following [47], we will assume that $\mathcal{R}$ is both strongly convex and continuously twice-differentiable. The next theorem shows that linear RFTL is a concave potential algorithm, with a potential function that is directly related to the convex conjugate of the regularizing function. These are known properties (see, e.g., Lemma 15 in [81]), and a proof, which uses basic calculus, is given in Section 3.5 for completeness.

Theorem 3.4. *If $\mathcal{R}:\mathcal{K}\to\mathbb{R}$ is continuous and strongly convex and $\eta > 0$, then $\Phi(\mathbf{L}) = (-1/\eta)\mathcal{R}^*(-\eta\mathbf{L})$ is concave and continuously differentiable on $\mathbb{R}^N$, and for every $\mathbf{L}\in\mathbb{R}^N$, it holds that $\nabla\Phi(\mathbf{L}) = \arg\min_{\mathbf{x}\in\mathcal{K}}\{\mathbf{x}\cdot\mathbf{L} + \mathcal{R}(\mathbf{x})/\eta\}$ and $\Phi(\mathbf{L}) = \min_{\mathbf{x}\in\mathcal{K}}\{\mathbf{x}\cdot\mathbf{L} + \mathcal{R}(\mathbf{x})/\eta\}$.*

It is now possible to lower bound the regret of $RFTL(\eta, \mathcal{R})$ by applying the lower bounds of Corollary 3.2 and Theorem 3.3.

Theorem 3.5. *The regret of $RFTL(\eta, \mathcal{R})$ satisfies*

$$R_{RFTL(\eta,\mathcal{R}),T} \geq \frac{1}{\eta}\left(\mathcal{R}(\mathbf{x}_{T+1}) - \mathcal{R}(\mathbf{x}_1)\right) + \left(\mathbf{x}_{T+1}\cdot\mathbf{L}_T - \min_{\mathbf{u}\in\mathcal{K}}\{\mathbf{u}\cdot\mathbf{L}_T\}\right) \geq 0\,,$$

and if R^* *is continuously twice-differentiable on* $\mathrm{conv}(\{-\eta\mathbf{L}_0, \ldots, -\eta\mathbf{L}_T\})$*, then*

$$R_{RFTL(\eta,\mathcal{R}),T} \geq \frac{1}{\eta}\left(\mathcal{R}(\mathbf{x}_{T+1}) - \mathcal{R}(\mathbf{x}_1)\right) + \left(\mathbf{x}_{T+1} \cdot \mathbf{L}_T - \min_{\mathbf{u}\in\mathcal{K}}\{\mathbf{u} \cdot \mathbf{L}_T\}\right) + \frac{\eta}{2}\sum_{t=1}^{T} \mathbf{l}_t^\top \nabla^2 \mathcal{R}^*(-\eta\mathbf{z}_t)\mathbf{l}_t \geq 0 \ ,$$

where $\mathbf{z}_t \in [\mathbf{L}_{t-1}, \mathbf{L}_t]$.

Proof. Let Φ be the potential function of $RFTL(\eta, \mathcal{R})$ according to Theorem 3.4. We have that $\Phi(\mathbf{L}_t) = \mathbf{x}_{t+1} \cdot \mathbf{L}_t + \mathcal{R}(\mathbf{x}_{t+1})/\eta$ and $\mathbf{x}_{t+1} = \nabla\Phi(\mathbf{L}_t)$ for every t. Therefore,

$$\begin{aligned}\Phi(\mathbf{L}_T) - \Phi(\mathbf{L}_0) &= \mathbf{x}_{T+1} \cdot \mathbf{L}_T + \frac{1}{\eta}\mathcal{R}(\mathbf{x}_{T+1}) - \mathbf{x}_1 \cdot \mathbf{L}_0 - \frac{1}{\eta}\mathcal{R}(\mathbf{x}_1)\\ &= \frac{1}{\eta}\left(\mathcal{R}(\mathbf{x}_{T+1}) - \mathcal{R}(\mathbf{x}_1)\right) + \mathbf{x}_{T+1} \cdot \mathbf{L}_T \ ,\end{aligned}$$

where we used the fact that $\mathbf{L}_0 = \mathbf{0}$. Therefore,

$$\begin{aligned}R_{RFTL(\eta,\mathcal{R}),T} &\geq \Phi(\mathbf{L}_T) - \Phi(\mathbf{L}_0) - \min_{\mathbf{u}\in\mathcal{K}}\{\mathbf{u} \cdot \mathbf{L}_T\}\\ &= \frac{1}{\eta}\left(\mathcal{R}(\mathbf{x}_{T+1}) - \mathcal{R}(\mathbf{x}_1)\right) + \left(\mathbf{x}_{T+1} \cdot \mathbf{L}_T - \min_{\mathbf{u}\in\mathcal{K}}\{\mathbf{u} \cdot \mathbf{L}_T\}\right) \geq 0 \ ,\end{aligned}$$

where the first inequality is by Corollary 3.2, and the second inequality is by the fact that $\mathbf{x}_1 = \arg\min_{\mathbf{u}\in\mathcal{K}}\{\mathcal{R}(\mathbf{u})\}$ and $\mathbf{x}_{T+1} \in \mathcal{K}$. This concludes the first part of the proof.

By Theorem 3.4, $\Phi(\mathbf{L}) = -(1/\eta)\mathcal{R}^*(-\eta\mathbf{L})$. Thus, Φ is continuously twice-differentiable on $\mathrm{conv}(\{\mathbf{L}_0, \ldots, \mathbf{L}_T\})$, and we have $\nabla^2\Phi(\mathbf{L}) = -\eta\nabla^2\mathcal{R}^*(-\eta\mathbf{L})$. By the first part and by Theorem 3.3, for some $\mathbf{z}_t \in [\mathbf{L}_{t-1}, \mathbf{L}_t]$,

$$\begin{aligned}R_{RFTL(\eta,\mathcal{R}),T} &= \Phi(\mathbf{L}_T) - \Phi(\mathbf{L}_0) - \min_{\mathbf{u}\in\mathcal{K}}\{\mathbf{u} \cdot (\mathbf{L}_T - \mathbf{L}_0)\} - \frac{1}{2}\sum_{t=1}^{T} \mathbf{l}_t^\top \nabla^2\Phi(\mathbf{z}_t)\mathbf{l}_t\\ &= \frac{1}{\eta}\left(\mathcal{R}(\mathbf{x}_{T+1}) - \mathcal{R}(\mathbf{x}_1)\right) + \mathbf{x}_{T+1} \cdot \mathbf{L}_T - \min_{\mathbf{u}\in\mathcal{K}}\{\mathbf{u} \cdot \mathbf{L}_T\}\\ &\quad + \frac{\eta}{2}\sum_{t=1}^{T} \mathbf{l}_t^\top \nabla^2\mathcal{R}^*(-\eta\mathbf{z}_t)\mathbf{l}_t \ .\end{aligned}$$

□

Note that the first order regret term is split into two new non-negative terms, namely, $\left(\mathcal{R}(\mathbf{x}_{T+1}) - \mathcal{R}(\mathbf{x}_1)\right)/\eta$, and $\mathbf{x}_{T+1} \cdot \mathbf{L}_T - \min_{\mathbf{u}\in\mathcal{K}}\{\mathbf{u} \cdot \mathbf{L}_T\}$.

The first order regret lower bound given in Theorem 3.5 may be proved directly by extending the "follow the leader, be the leader" (FTL-BTL) lemma ([59], see also [47]). The extension and proof are given in Subsection 3.5.1.

3.3 Strictly Positive Individual Sequence Anytime Regret

In this section we give lower bounds on the anytime regret for two classes of potentials. These are the class of potentials with negative definite Hessians, and a rich class of best expert potentials that includes Hedge.

We start by describing our general argument, which is based on the lower bounds of Theorem 3.3 and Corollary 3.2. Note first that these bounds hold for any $1 \leq T' \leq T$. Now, let C be some convex set on which Φ is continuously twice-differentiable. If $\mathbf{L}_0, \ldots, \mathbf{L}_T \in C$, then we may lower bound the regret by the second order term of Theorem 3.3, and further lower bound that term over any $\mathbf{L}_0, \ldots, \mathbf{L}_T \in C$. Otherwise, there is some $1 \leq T' \leq T$ s.t. $\mathbf{L}_{T'} \notin C$, and we may lower bound the regret at time T' by the first order term, using Corollary 3.2. We further lower bound that term over any $\mathbf{L}_{T'} \notin C$. The minimum of those two lower bounds gives a lower bound on the anytime regret. By choosing C properly, we will be able to prove that these two lower bounds are strictly positive.

We now present our analysis in more detail. Observe that for every $\eta > 0$ and concave potential $\Phi : \mathbb{R}^N \to \mathbb{R}$, $\Phi_\eta(\mathbf{L}) = (1/\eta)\Phi(\eta\mathbf{L})$ is also concave, with $\nabla\Phi_\eta(\mathbf{L}) = \nabla\Phi(\eta\mathbf{L})$. Let $\|\cdot\|$ be a non-trivial seminorm on $\mathbb{R}^N$. In addition, let $a > 0$ be such that Φ is continuously twice-differentiable on the set $\{\|\mathbf{L}\| \leq a\}$, and let $\mathbf{L}_0 = \mathbf{0}$.

Suppose algorithm A uses Φ_η and encounters the loss path $\mathbf{L}_0, \ldots, \mathbf{L}_T$. If there is some T' s.t. $\|\mathbf{L}_{T'}\| > a/\eta$, then applying Corollary 3.2 to Φ_η, we get

$$R_{A,T'} \geq \frac{1}{\eta}\left(\Phi(\eta\mathbf{L}_{T'}) - \Phi(\mathbf{0}) - \min_{\mathbf{u}\in\mathcal{K}}\{\mathbf{u} \cdot \eta\mathbf{L}_{T'}\}\right) .$$

Defining $\rho_1(a) = \inf_{\|\mathbf{L}\|\geq a}\{\Phi(\mathbf{L}) - \Phi(\mathbf{0}) - \min_{\mathbf{u}\in\mathcal{K}}\{\mathbf{u}\cdot\mathbf{L}\}\}$, we have that $R_{A,T'} \geq \rho_1(a)/\eta$.

We next assume that $\|\mathbf{L}_t\| \leq a/\eta$ for every t. It is easily verified that the set $\{\|\mathbf{L}\| \leq a/\eta\}$ is convex, and since it contains $\mathbf{L}_t$ for every t, it also contains $\mathrm{conv}(\{\mathbf{L}_0, \ldots, \mathbf{L}_T\})$. This means that $\Phi_\eta(\mathbf{L}) = (1/\eta)\Phi(\eta\mathbf{L})$ is continuously twice-differentiable on $\mathrm{conv}(\{\mathbf{L}_0, \ldots, \mathbf{L}_T\})$. Applying Theorem 3.3 to Φ_η and dropping the non-negative first order term, we have

$$R_{A,T} \geq -\frac{1}{2}\sum_{t=1}^{T} \mathbf{l}_t^\top \nabla^2 \Phi_\eta(\mathbf{z}_t)\mathbf{l}_t = -\frac{\eta}{2}\sum_{t=1}^{T} \mathbf{l}_t^\top \nabla^2 \Phi(\eta\mathbf{z}_t)\mathbf{l}_t \ ,$$

where $\mathbf{z}_t \in [\mathbf{L}_{t-1}, \mathbf{L}_t]$. We now define $\rho_2(a) = \inf_{\|\mathbf{L}\|\leq a, \|\mathbf{l}\|=1}\{-\mathbf{l}^\top \nabla^2\Phi(\mathbf{L})\mathbf{l}\}$. If $\|\mathbf{l}_t\| \neq 0$, then

$$-\mathbf{l}_t^\top \nabla^2\Phi(\eta\mathbf{z}_t)\mathbf{l}_t = -(\mathbf{l}_t/\|\mathbf{l}_t\|)^\top \nabla^2\Phi(\eta\mathbf{z}_t)(\mathbf{l}_t/\|\mathbf{l}_t\|)\|\mathbf{l}_t\|^2 \geq \rho_2(a)\|\mathbf{l}_t\|^2 \ ,$$

where we used the fact that $\|\eta\mathbf{z}_t\| \leq a$, which holds since $\mathbf{z}_t \in \mathrm{conv}(\{\mathbf{L}_0, \ldots, \mathbf{L}_T\})$. Otherwise, $-\mathbf{l}_t^\top \nabla^2\Phi(\eta\mathbf{z}_t)\mathbf{l}_t \geq 0 = \rho_2(a)\|\mathbf{l}_t\|^2$, so in any case,

$$R_{A,T} \geq -\frac{\eta}{2}\sum_{t=1}^{T} \mathbf{l}_t^\top \nabla^2\Phi(\eta\mathbf{z}_t)\mathbf{l}_t \geq \frac{\eta}{2}\sum_{t=1}^{T}\rho_2(a)\|\mathbf{l}_t\|^2 = \frac{\eta}{2}\rho_2(a)\sum_{t=1}^{T}\|\mathbf{l}_t\|^2 \ .$$

Thus, we have

Lemma 3.6. *If $\|\mathbf{L}_t\| \leq a/\eta$ for every t, then $R_{A,T} \geq \frac{\eta}{2}\rho_2(a)\sum_{t=1}^{T}\|\mathbf{l}_t\|^2$. Otherwise, for any t s.t. $\|\mathbf{L}_t\| > a/\eta$, $R_{A,t} \geq \rho_1(a)/\eta$. Therefore,*

$$\max_t\{R_{A,t}\} \geq \min\left\{\frac{\rho_1(a)}{\eta}, \frac{\eta}{2}\rho_2(a)\sum_{t=1}^{T}\|\mathbf{l}_t\|^2\right\} \ .$$

Note that $\rho_1(a)$ is non-decreasing, $\rho_2(a)$ is non-increasing, and $\rho_1(a), \rho_2(a) \geq 0$. Lemma 3.6 therefore highlights a trade-off in the choice of a.

It still remains to bound ρ_1 and ρ_2 away from zero, and that will be done in two different ways for the cases of negative definite Hessians and best expert potentials. Nevertheless, the next technical lemma, which is instrumental in bounding ρ_1 away from zero, still holds in general. Essentially, it says that in the definition of $\rho_1(a)$, it suffices to take the infimum over $\{\|\mathbf{L}\| = a\}$ instead of $\{\|\mathbf{L}\| \geq a\}$.

Lemma 3.7. *It holds that $\rho_1(a) = \inf_{\|\mathbf{L}\|=a}\{\Phi(\mathbf{L}) - \Phi(\mathbf{0}) - \min_{\mathbf{u}\in\mathcal{K}}\{\mathbf{u}\cdot\mathbf{L}\}\}$.*

Proof. Let $\|\mathbf{L}\| \geq a$, and define $\mathbf{L}' = a\mathbf{L}/\|\mathbf{L}\|$, so $\|\mathbf{L}'\| = a$. Since Φ is concave, we have that $\Phi(\mathbf{L}) \geq \Phi(\mathbf{L}') + \nabla\Phi(\mathbf{L})(\mathbf{L} - \mathbf{L}')$, and thus

$$\begin{aligned}
\Phi(\mathbf{L}) - \min_{\mathbf{u}\in\mathcal{K}}\{\mathbf{u}\cdot\mathbf{L}\} &\geq \Phi(\mathbf{L}') + \nabla\Phi(\mathbf{L})\cdot(\mathbf{L} - \mathbf{L}') - \min_{\mathbf{u}\in\mathcal{K}}\{\mathbf{u}\cdot\mathbf{L}\} \\
&\geq \Phi(\mathbf{L}') + \left(\frac{\|\mathbf{L}\|}{a} - 1\right)\nabla\Phi(\mathbf{L})\cdot\mathbf{L}' - \frac{\|\mathbf{L}\|}{a}\min_{\mathbf{u}\in\mathcal{K}}\{\mathbf{u}\cdot\mathbf{L}'\} \\
&= \Phi(\mathbf{L}') - \min_{\mathbf{u}\in\mathcal{K}}\{\mathbf{u}\cdot\mathbf{L}'\} \\
&\quad + \left(\frac{\|\mathbf{L}\|}{a} - 1\right)\left(\nabla\Phi(\mathbf{L})\cdot\mathbf{L}' - \min_{\mathbf{u}\in\mathcal{K}}\{\mathbf{u}\cdot\mathbf{L}'\}\right) \\
&\geq \Phi(\mathbf{L}') - \min_{\mathbf{u}\in\mathcal{K}}\{\mathbf{u}\cdot\mathbf{L}'\}\ ,
\end{aligned}$$

where the last inequality uses the fact that $\nabla\Phi(\mathbf{L})\cdot\mathbf{L}' - \min_{\mathbf{u}\in\mathcal{K}}\{\mathbf{u}\cdot\mathbf{L}'\} \geq 0$. Subtracting $\Phi(\mathbf{0})$, the result follows. □

We now present the main result of this section, assuming we have shown $\rho_1, \rho_2 > 0$. In what follows we denote $Q_T = Q_T(\mathbf{l}_1, \ldots, \mathbf{l}_T) = \sum_{t=1}^{T} \|\mathbf{l}_t\|^2$ for the generic quadratic variation w.r.t. the seminorm $\|\cdot\|$ (not to be confused with specific notions of quadratic variation). We denote $Q > 0$ for a given lower bound on Q_T.

Theorem 3.8. *Let $a > 0$ satisfy $\rho_1(a), \rho_2(a) > 0$.*

(i) For every $\eta > 0$, $\max_t\{R_{A,t}\} \geq \min\{\frac{\rho_1(a)}{\eta}, \frac{\eta}{2}\rho_2(a)Q\}$, and for $\eta = \sqrt{\frac{2\rho_1(a)}{\rho_2(a)Q}}$,

$$\max_t\{R_{A,t}\} \geq \sqrt{\rho_1(a)\rho_2(a)/2} \cdot \sqrt{Q}\ .$$

(ii) If for any sequence with quadratic variation $Q_T \leq Q'$ we have $R_{A,T} \leq c\sqrt{Q'}$, then for any such sequence,

$$\max_t\{R_{A,t}\} \geq \frac{\rho_1(a)\rho_2(a)}{2c} \cdot \frac{Q_T}{\sqrt{Q'}}\ .$$

In particular, if $Q_T = \Theta(Q')$, then $\max_t\{R_{A,t}\} = \Omega(\sqrt{Q_T})$.

Proof. (i) By Lemma 3.6,

$$\max_t\{R_{A,t}\} \geq \min\left\{\frac{\rho_1(a)}{\eta}, \frac{\eta}{2}\rho_2(a)\sum_{t=1}^{T}\|\mathbf{l}_t\|^2\right\} \geq \min\left\{\frac{\rho_1(a)}{\eta}, \frac{\eta}{2}\rho_2(a)Q\right\}\ .$$

Picking $\eta = \sqrt{\frac{2\rho_1(a)}{\rho_2(a)Q}}$ implies that $\frac{\rho_1(a)}{\eta} = \frac{\eta}{2}\rho_2(a)Q = \sqrt{\frac{1}{2}Q\rho_1(a)\rho_2(a)}$.

(ii) Given a and η, let $0 < \epsilon < a/\eta$ satisfy that $T = Q'/\epsilon^2$ is an integer. In addition, let $\mathbf{x} \in \mathbb{R}^N$ be such that $\|\mathbf{x}\| = 1$. The loss sequence $\mathbf{l}_t = (-1)^{t+1}\epsilon\mathbf{x}$, for $1 \le t \le T$, satisfies that $Q_T = \epsilon^2 T = Q'$ and that $\|\mathbf{L}_t\| = \|(1-(-1)^t)\epsilon\mathbf{x}/2\| \le \epsilon < a/\eta$ for every t. Therefore,

$$c\sqrt{Q'} \ge R_{A,T} \ge \frac{\eta}{2}\rho_2(a)\sum_{t=1}^{T}\|\mathbf{l}_t\|^2 = \frac{\eta}{2}\rho_2(a)Q' ,$$

where the second inequality is by Lemma 3.6. This implies that $\eta \le \frac{2c}{\rho_2(a)\sqrt{Q'}}$.

On the other hand, let $T > \frac{a^2}{\eta^2 Q'}$, define $\epsilon = \sqrt{Q'/T}$, and consider the loss sequence $\mathbf{l}_t = \epsilon\mathbf{x}$. Then $Q_T = \epsilon^2 T = Q'$, and $\|\mathbf{L}_T\| = \epsilon T = \sqrt{Q'T} > a/\eta$. Thus, again using Lemma 3.6, we have $c\sqrt{Q'} \ge R_{A,T} \ge \rho_1(a)/\eta$, which means that $\eta \ge \frac{\rho_1(a)}{c\sqrt{Q'}}$. Together, we have that $\frac{\rho_1(a)}{c\sqrt{Q'}} \le \eta \le \frac{2c}{\rho_2(a)\sqrt{Q'}}$, implying that $c \ge \sqrt{\rho_1(a)\rho_2(a)/2}$.[1] Given any sequence with $Q_T \le Q'$, we have that $\frac{\rho_1(a)}{\eta} \ge \frac{\rho_1(a)\rho_2(a)\sqrt{Q'}}{2c}$ and $\frac{\eta}{2}\rho_2(a)Q_T \ge \frac{\rho_1(a)\rho_2(a)Q_T}{2c\sqrt{Q'}}$, so by Lemma 3.6, $\max_t\{R_{A,t}\} \ge \frac{\rho_1(a)\rho_2(a)}{2c} \cdot \frac{Q_T}{\sqrt{Q'}}$, concluding the proof. □

3.3.1 Potentials with Negative Definite Hessians

For this case, we pick $\|\cdot\|_2$ as our seminorm. Let $a > 0$ and let $\nabla^2\Phi(\mathbf{L}) \prec 0$ for $\mathbf{L}$ s.t. $\|\mathbf{L}\|_2 \le a$. In this setting, the infimum in the definitions of $\rho_1(a)$ and $\rho_2(a)$ is equivalent to a minimum, using continuity and the compactness of L^2 balls and spheres.

Lemma 3.9. *If $\|\cdot\| = \|\cdot\|_2$, then $\rho_1(a) = \min_{\|\mathbf{L}\|=a}\{\Phi(\mathbf{L}) - \Phi(\mathbf{0}) - \min_{\mathbf{u}\in\mathcal{K}}\{\mathbf{u}\cdot\mathbf{L}\}\}$ and $\rho_2(a) = \min_{\|\mathbf{L}\|\le a, \|\mathbf{l}\|=1}\{-\mathbf{l}^\top\nabla^2\Phi(\mathbf{L})\mathbf{l}\}$.*

Proof. By Lemma 3.7, $\rho_1(a) = \inf_{\|\mathbf{L}\|=a}\{\Phi(\mathbf{L}) - \Phi(\mathbf{0}) - \min_{\mathbf{u}\in\mathcal{K}}\{\mathbf{u}\cdot\mathbf{L}\}\}$. The function $h(\mathbf{L}) = \min_{\mathbf{u}\in\mathcal{K}}\{\mathbf{u}\cdot\mathbf{L}\}$ is continuous on $\mathbb{R}^N$ (Lemma 3.20). Therefore, $\Phi(\mathbf{L}) - \Phi(\mathbf{0}) - \min_{\mathbf{u}\in\mathcal{K}}\{\mathbf{u}\cdot\mathbf{L}\}$ is continuous on the compact set $\{\|\mathbf{L}\| = a\}$ and attains a minimum there. The second statement follows from the fact that $-\mathbf{l}^\top\nabla^2\Phi(\mathbf{L})\mathbf{l}$ is a continuous function on the compact set $\{\|\mathbf{L}\| \le a\} \times \{\|\mathbf{l}\| = 1\}$, and attains a minimum there. □

By Lemma 3.9, $\rho_2(a) = \min_{\|\mathbf{L}\|\le a, \|\mathbf{l}\|=1}\{-\mathbf{l}^\top\nabla^2\Phi(\mathbf{L})\mathbf{l}\} > 0$, where the inequality is true since the Hessians are negative definite, so we are taking the minimum of positive values. In addition, if $\mathbf{L} \ne \mathbf{0}$, then $\Phi(\mathbf{L}) - \Phi(\mathbf{0}) - \min_{\mathbf{u}\in\mathcal{K}}\{\mathbf{u}\cdot\mathbf{L}\} > \nabla\Phi(\mathbf{L})\cdot\mathbf{L} - \min_{\mathbf{u}\in\mathcal{K}}\{\mathbf{u}\cdot\mathbf{L}\} \ge 0$, since Φ is strictly concave. Thus, again by Lemma 3.9, $\rho_1(a) =$

[1] Note that this means we cannot guarantee a regret upper bound of $c\sqrt{Q'}$ for $c < \sqrt{\rho_1(a)\rho_2(a)/2}$.

$\min_{\|\mathbf{L}\|=a}\{\Phi(\mathbf{L})-\Phi(\mathbf{0})-\min_{\mathbf{u}\in\mathcal{K}}\{\mathbf{u}\cdot\mathbf{L}\}\} > 0$. The following statement is an immediate consequence of Theorem 3.8:

Theorem 3.10. *If $\nabla^2\Phi(\mathbf{L}) \prec 0$ for every $\mathbf{L}$ s.t. $\|\mathbf{L}\|_2 \le a$, for some $a > 0$, then*

(i) For every $\eta > 0$, it holds that $\max_t\{R_{A,t}\} \ge \min\{\frac{\rho_1(a)}{\eta}, \frac{\eta}{2}\rho_2(a)Q\}$, and for $\eta = \sqrt{\frac{2\rho_1(a)}{\rho_2(a)Q}}$, $\max_t\{R_{A,t}\} \ge \sqrt{\rho_1(a)\rho_2(a)/2}\cdot\sqrt{Q}$.

(ii) If for any sequence with quadratic variation $Q_T \le Q'$ we have $R_{A,T} \le c\sqrt{Q'}$, then for any such sequence, $\max_t\{R_{A,t}\} \ge \frac{\rho_1(a)\rho_2(a)}{2c}\cdot\frac{Q_T}{\sqrt{Q'}}$. In particular, if $Q_T = \Theta(Q')$, then $\max_t\{R_{A,t}\} = \Omega(\sqrt{Q_T})$.

3.3.2 The Best Expert Setting

In the best expert setting, where $\mathcal{K} = \Delta_N$, potentials can never be strictly concave, let alone have negative definite Hessians. To see that, let $\mathbf{L} \in \mathbb{R}^N$, $c \in \mathbb{R}$, and define $\mathbf{L}' = \mathbf{L} + c\cdot\mathbf{1}$, where $\mathbf{1}$ is the all-one vector. We will say that $\mathbf{L}'$ is a *uniform translation* of $\mathbf{L}$. Then

$$c = \nabla\Phi(\mathbf{L})\cdot(\mathbf{L}'-\mathbf{L}) \ge \Phi(\mathbf{L}') - \Phi(\mathbf{L}) \ge \nabla\Phi(\mathbf{L}')\cdot(\mathbf{L}'-\mathbf{L}) = c\ ,$$

where we use the concavity of Φ, the fact that $\nabla\Phi$ is a probability vector, and the fact that $\mathbf{L}'-\mathbf{L} = c\cdot\mathbf{1}$. For a strictly concave Φ, the above inequalities would be strict if $c \ne 0$, but instead, they are equalities. Thus, the conditions for strict concavity are not fulfilled at any point $\mathbf{L}$.

We will replace the negative definite assumption with the assumption that for every $i \ne j$, $\frac{\partial^2\Phi}{\partial L_i \partial L_j} > 0$. This condition is natural for regret minimization algorithms, because $\frac{\partial^2\Phi(\mathbf{L})}{\partial L_i \partial L_j} = \frac{\partial p_i(\mathbf{L})}{\partial L_j}$, where p_i is the weight of expert i. Thus, we simply require that an increase in the cumulative loss of expert j results in an increase in the weight of every other expert (and hence a decrease in its own weight). A direct implication of this assumption is that $\frac{\partial\Phi(\mathbf{L})}{\partial L_i} > 0$ for every i and $\mathbf{L}$. To see that, observe that $p_i(\mathbf{L}) = 1 - \sum_{j\ne i} p_j(\mathbf{L})$, so $\frac{\partial p_i(\mathbf{L})}{\partial L_i} = -\sum_{j\ne i}\frac{\partial p_j(\mathbf{L})}{\partial L_i} < 0$. Since $p_i(\mathbf{L}) \ge 0$ and it is strictly decreasing in L_i, it follows that $p_i(\mathbf{L}) > 0$, or $\frac{\partial\Phi(\mathbf{L})}{\partial L_i} > 0$.

Using the above assumption we proceed to bound ρ_1 and ρ_2 away from zero. We first prove some general properties of best expert potentials.

Lemma 3.11. *(i) Every row and column of* $\nabla^2\Phi$ *sum up to zero. (ii)* $\Phi(\mathbf{L}) - \Phi(\mathbf{0}) - \min_{\mathbf{u}\in\mathcal{K}}\{\mathbf{u}\cdot\mathbf{L}\}$ *is invariant w.r.t. a uniform translation of* $\mathbf{L}$*. (iii)* $\mathbf{l}^\top\nabla^2\Phi(\mathbf{L})\mathbf{l}$ *is invariant w.r.t. a uniform translation of either* $\mathbf{l}$ *or* $\mathbf{L}$*.*

Proof. (i) Defining $h(\mathbf{L}) = \sum_{i=1}^N \frac{\partial\Phi(\mathbf{L})}{\partial L_i}$, we have that $h \equiv 1$, and thus $0 = \frac{\partial h(\mathbf{L})}{\partial L_j} = \sum_{i=1}^N \frac{\partial^2\Phi(\mathbf{L})}{\partial L_i \partial L_j}$ for every j.

(ii) Let $\mathbf{L}' = \mathbf{L} + c\cdot\mathbf{1}$. As already seen, $\Phi(\mathbf{L}') - \Phi(\mathbf{L}) = c$, and we also have that

$$\min_{\mathbf{u}\in\mathcal{K}}\{\mathbf{u}\cdot\mathbf{L}'\} = \min_{\mathbf{u}\in\mathcal{K}}\{\mathbf{u}\cdot\mathbf{L} + c\cdot\mathbf{u}\cdot\mathbf{1}\} = \min_{\mathbf{u}\in\mathcal{K}}\{\mathbf{u}\cdot\mathbf{L}\} + c\,.$$

Therefore,

$$\Phi(\mathbf{L}') - \Phi(\mathbf{0}) - \min_{\mathbf{u}\in\mathcal{K}}\{\mathbf{u}\cdot\mathbf{L}'\} = \Phi(\mathbf{L}) - \Phi(\mathbf{0}) - \min_{\mathbf{u}\in\mathcal{K}}\{\mathbf{u}\cdot\mathbf{L}\}\,.$$

(iii) Note that in this context, $\mathbf{l}, \mathbf{L} \in \mathbb{R}^N$ are unrelated vectors.

$$\begin{aligned}(\mathbf{l}+c\mathbf{1})^\top\nabla^2\Phi(\mathbf{L})(\mathbf{l}+c\mathbf{1}) &= \mathbf{l}^\top\nabla^2\Phi(\mathbf{L})\mathbf{l} + c\mathbf{1}^\top\nabla^2\Phi(\mathbf{L})(\mathbf{l}+c\mathbf{1}) + \mathbf{l}^\top\nabla^2\Phi(\mathbf{L})(c\mathbf{1})\\ &= \mathbf{l}^\top\nabla^2\Phi(\mathbf{L})\mathbf{l}\,,\end{aligned}$$

where we have that $\mathbf{1}^\top\nabla^2\Phi(\mathbf{L}) = \mathbf{0}^\top$ and $\nabla^2\Phi(\mathbf{L})\mathbf{1} = \mathbf{0}$ by (i). To show the invariance for $\mathbf{L}$, recall that for $\mathbf{L}' = \mathbf{L}+c\cdot\mathbf{1}$, $\Phi(\mathbf{L}')-\Phi(\mathbf{L}) = c$, and thus, $\nabla^2\Phi(\mathbf{L}') = \nabla^2\Phi(\mathbf{L})$. □

We now consider $\rho_1(a)$ and $\rho_2(a)$ where we use the seminorm $\|\mathbf{v}\| = \delta(\mathbf{v})$. (The fact that $\delta(\mathbf{v}) = \max_i\{v_i\} - \min_i\{v_i\}$ is a seminorm is proved in Lemma 3.21 in Section 3.5.) Under this seminorm, $\sum_{t=1}^T \|\mathbf{l}_t\|^2$ becomes q_T, the relative quadratic variation. Note that $\delta(\mathbf{v})$ is invariant to uniform translation. In particular, for every $\mathbf{v} \in \mathbb{R}^N$ we may consider its "normalized" version, $\hat{\mathbf{v}} = \mathbf{v} - \min_i\{v_i\}\cdot\mathbf{1}$. We have that $\delta(\hat{\mathbf{v}}) = \delta(\mathbf{v})$, $\hat{\mathbf{v}} \in [0, \delta(\mathbf{v})]^N$, and there exist entries i and j s.t. $\hat{\mathbf{v}}_i = 0$ and $\hat{\mathbf{v}}_j = \delta(\mathbf{v})$. Denoting $\mathcal{N}(a)$ for the set of normalized vectors with seminorm a, we thus have that $\mathcal{N}(a) = \{\mathbf{v} \in [0,a]^N : \exists i, j \text{ s.t. } v_i = a,\ v_j = 0\}$. The set $\mathcal{N}(a)$ is bounded and also closed, as a finite union and intersection of closed sets.

Using invariance to uniform translation, we can now show that the infima in the expressions for ρ_1 and ρ_2 may be taken over compact sets, and thus be replaced with minima. Using the requirement that $\frac{\partial^2\Phi}{\partial L_i \partial L_j} > 0$ for every $i \neq j$, we can then show that the expressions inside the minima are positive. This is summarized in the next lemma.

Lemma 3.12. *For the best expert setting, it holds that* $\rho_1(a) = \min_{\mathbf{L}\in\mathcal{N}(a)}\{\Phi(\mathbf{L})\} - \Phi(\mathbf{0}) > 0$ *and* $\rho_2(a) = \min_{\mathbf{L}\in[0,a]^N,\mathbf{l}\in\mathcal{N}(1)}\{-\mathbf{l}^\top\nabla^2\Phi(\mathbf{L})\mathbf{l}\} > 0$.

Proof. We start with $\rho_1(a)$. By Lemma 3.7, we have that $\rho_1(a) = \inf_{\|\mathbf{L}\|=a}\{\Phi(\mathbf{L}) - \Phi(\mathbf{0}) - \min_{\mathbf{u}\in\mathcal{K}}\{\mathbf{u}\cdot\mathbf{L}\}\}$. By part (ii) of Lemma 3.11, $\Phi(\mathbf{L}) - \Phi(\mathbf{0}) - \min_{\mathbf{u}\in\mathcal{K}}\{\mathbf{u}\cdot\mathbf{L}\}$ is invariant to uniform translation, so $\rho_1(a) = \inf_{\mathbf{L}\in\mathcal{N}(a)}\{\Phi(\mathbf{L}) - \Phi(\mathbf{0}) - \min_{\mathbf{u}\in\mathcal{K}}\{\mathbf{u}\cdot\mathbf{L}\}\}$. For every normalized vector $\mathbf{L}$, $\min_{\mathbf{u}\in\mathcal{K}}\{\mathbf{u}\cdot\mathbf{L}\} = 0$, by placing all the weight of $\mathbf{u}$ on the zero entry of $\mathbf{L}$. In addition, as already observed, $\Phi(\mathbf{L}) - \Phi(\mathbf{0}) - \min_{\mathbf{u}\in\mathcal{K}}\{\mathbf{u}\cdot\mathbf{L}\}$ is continuous, and by the compactness of $\mathcal{N}(a)$, $\rho_1(a) = \min_{\mathbf{L}\in\mathcal{N}(a)}\{\Phi(\mathbf{L}) - \Phi(\mathbf{0}) - \min_{\mathbf{u}\in\mathcal{K}}\{\mathbf{u}\cdot\mathbf{L}\}\} = \min_{\mathbf{L}\in\mathcal{N}(a)}\{\Phi(\mathbf{L}) - \Phi(\mathbf{0})\}$. Now, $\Phi(\mathbf{L}) - \Phi(\mathbf{0}) \geq \nabla\Phi(\mathbf{L})\cdot\mathbf{L} = \sum_{i=1}^N p_i(\mathbf{L})L_i$ by the concavity of Φ. Recall that $p_i(\mathbf{L}) > 0$, and for $\mathbf{L}\in\mathcal{N}(a)$, we have that $L_i \geq 0$ for every i, and there is at least one index with $L_i > 0$. Thus, $\Phi(\mathbf{L}) - \Phi(\mathbf{0}) > 0$, proving the statement for ρ_1.

We now move on to ρ_2. By Lemma 3.11, $-\mathbf{l}^\top\nabla^2\Phi(\mathbf{L})\mathbf{l}$ is invariant to a uniform translation of either $\mathbf{l}$ or $\mathbf{L}$. Therefore, $\rho_2(a) = \inf_{\mathbf{L}\in[0,a]^N,\mathbf{l}\in\mathcal{N}(1)}\{-\mathbf{l}^\top\nabla^2\Phi(\mathbf{L})\mathbf{l}\}$. Since $-\mathbf{l}^\top\nabla^2\Phi(\mathbf{L})\mathbf{l}$ is continuous on the compact set $[0,a]^N \times \mathcal{N}(1)$, $\rho_2(a) = \min_{\mathbf{L}\in[0,a]^N,\mathbf{l}\in\mathcal{N}(1)}\{-\mathbf{l}^\top\nabla^2\Phi(\mathbf{L})\mathbf{l}\}$. To show that $\rho_2(a) > 0$, it therefore suffices to show that $-\mathbf{l}^\top\nabla^2\Phi(\mathbf{L})\mathbf{l} > 0$ for every $\mathbf{L}\in[0,a]^N$ and $\mathbf{l}\in\mathcal{N}(1)$. Fix $\mathbf{L}$, and denote $A = -\nabla^2\Phi(\mathbf{L})$. For any $\mathbf{l}\in\mathcal{N}(1)$, one entry is 1, another is 0, and the rest are in $[0,1]$. If $N=2$, we have that $-\mathbf{l}^\top\nabla^2\Phi(\mathbf{L})\mathbf{l}$ is either $-\nabla^2\Phi(\mathbf{L})_{1,1}$ or $-\nabla^2\Phi(\mathbf{L})_{2,2}$, and in both cases positive. We next assume $N\geq 3$ and minimize $f(\mathbf{l}) = \mathbf{l}^\top A\mathbf{l} = \sum_{i,j} a_{i,j}l_il_j$ for $l_k\in[0,1]$, $k\geq 3$, where $l_1 = 1$, $l_2 = 0$. For every k,

$$\frac{\partial f}{\partial l_k} = 2a_{k,k}l_k + 2\sum_{i\neq k}a_{i,k}l_i = 2\sum_{i\neq k}a_{i,k}(l_i - l_k)\ ,$$

where the last equality is by part (i) of Lemma 3.11. Note that $a_{i,k} < 0$ if $i\neq k$ and $a_{k,k} > 0$. Thus, for $l_k = 0$, $\frac{\partial f}{\partial l_k} < 0$, and for $l_k = 1$, $\frac{\partial f}{\partial l_k} > 0$, so by the linearity of $\frac{\partial f}{\partial l_k}$, there is $l_k^* \in [0,1]$ satisfying $\frac{\partial f}{\partial l_k} = 0$, and it is a minimizing choice for f. Thus, at a minimum point, it holds that $\frac{\partial f}{\partial l_k} = 0$, or $\sum_{i\neq k}a_{i,k}l_i = -a_{k,k}l_k$ for every $k\geq 3$. We have that

$$\begin{aligned}
f(\mathbf{l}) &= \sum_{i=1}^{N} a_{i,i} l_i^2 + \sum_{i=1}^{N} \sum_{j\neq i} a_{i,j} l_i l_j = \sum_{i=1}^{N} a_{i,i} l_i^2 + \sum_{i=1}^{2} l_i \sum_{j\neq i} a_{i,j} l_j + \sum_{i=3}^{N} l_i \sum_{j\neq i} a_{i,j} l_j \\
&= \sum_{i=1}^{N} a_{i,i} l_i^2 + \sum_{j\neq 1} a_{1,j} l_j + \sum_{i=3}^{N} l_i(-a_{i,i} l_i) = \sum_{i=1}^{N} a_{i,i} l_i^2 + \sum_{j\neq 1} a_{1,j} l_j - \sum_{i=3}^{N} a_{i,i} l_i^2 \\
&= \sum_{i=1}^{2} a_{i,i} l_i^2 + \sum_{j\neq 1} a_{1,j} l_j = a_{1,1} + \sum_{j\neq 1} a_{1,j} l_j = -\sum_{j\neq 1} a_{1,j} + \sum_{j\neq 1} a_{1,j} l_j \\
&= \sum_{j\neq 1} a_{1,j}(l_j - 1) \geq a_{1,2}(l_2 - 1) = -a_{1,2} \, .
\end{aligned}$$

Since our choice of $l_1 = 1$ and $l_2 = 0$ was arbitrary, we have that $f(\mathbf{l}) \geq \min_{i\neq j}\{-a_{i,j}\} > 0$. Thus, $-\mathbf{l}^\top \nabla^2 \Phi(\mathbf{L})\mathbf{l} > 0$, completing the proof. □

We can now apply Theorem 3.8 to the best expert setting.

Theorem 3.13. *If $\frac{\partial^2 \Phi}{\partial L_i \partial L_j} > 0$ for every $i \neq j$ and every $\mathbf{L}$ s.t. $\delta(\mathbf{L}) \leq a$, then*

(i) For every $\eta > 0$, it holds that $\max_t\{R_{A,t}\} \geq \min\{\frac{\rho_1(a)}{\eta}, \frac{\eta}{2}\rho_2(a)q\}$, and for $\eta = \sqrt{\frac{2\rho_1(a)}{\rho_2(a)q}}$, $\max_t\{R_{A,t}\} \geq \sqrt{\rho_1(a)\rho_2(a)/2} \cdot \sqrt{q}$.

(ii) If for any sequence with relative quadratic variation $q_T \leq q'$ we have $R_{A,T} \leq c\sqrt{q'}$, then for any such sequence, $\max_t\{R_{A,t}\} \geq \frac{\rho_1(a)\rho_2(a)}{2c} \cdot \frac{q_T}{\sqrt{q'}}$. In particular, if $q_T = \Theta(q')$, then $\max_t\{R_{A,t}\} = \Omega(\sqrt{q_T})$.

3.4 Application to Specific Regret Minimization Algorithms

3.4.1 Online Gradient Descent with Linear Costs

In this subsection, we deal with the Lazy Projection variant of the OGD algorithm ([92]) with a fixed learning rate η and linear costs. In this setting, for each t, OGD selects a weight vector according to the rule $\mathbf{x}_{t+1} = \arg\min_{\mathbf{x}\in\mathcal{K}}\{\|\mathbf{x} + \eta\mathbf{L}_t\|_2\}$, where $\mathcal{K} \subset \mathbb{R}^N$ is compact and convex. As observed in [49] and [47], this algorithm is equivalent to $RFTL(\eta, \mathcal{R})$, where $\mathcal{R}(\mathbf{x}) = \frac{1}{2}\|\mathbf{x}\|_2^2$, namely, setting $\mathbf{x}_{t+1} = \arg\min_{\mathbf{x}\in\mathcal{K}}\{\mathbf{x} \cdot \mathbf{L}_t + (1/2\eta)\|\mathbf{x}\|_2^2\}$. In what follows we will make the assumption that $\mathcal{K} \supseteq B(\mathbf{0}, a)$, where $B(\mathbf{0}, a)$ is the closed ball with radius a centered at $\mathbf{0}$, for some $a > 0$.

Note that solving the above minimization problem without the restriction $\mathbf{x} \in \mathcal{K}$ yields $\mathbf{x}'_{t+1} = -\eta \mathbf{L}_t$. However, if $\|\mathbf{L}_t\|_2 \leq a/\eta$, then $\mathbf{x}'_{t+1} \in \mathcal{K}$, and then, in fact, $\mathbf{x}_{t+1} = -\eta \mathbf{L}_t$. By Theorem 3.4,

$$\begin{aligned}\Phi_\eta(\mathbf{L}_t) &= \mathbf{x}_{t+1} \cdot \mathbf{L}_t + (1/\eta)\mathcal{R}(\mathbf{x}_{t+1}) = -\eta \mathbf{L}_t \cdot \mathbf{L}_t + (1/2\eta)\| - \eta \mathbf{L}_t\|_2^2 \\ &= -(\eta/2)\|\mathbf{L}_t\|_2^2 \ .\end{aligned}$$

Thus, if $\|\mathbf{L}\|_2 \leq a$, then $\Phi(\mathbf{L}) = -(1/2)\|\mathbf{L}\|_2^2$ and also $\nabla^2\Phi(\mathbf{L}) = -I$, where I is the identity matrix. By Lemma 3.9,

$$\rho_1(a) = \min_{\|\mathbf{L}\|_2 = a} \{-\frac{1}{2}\|\mathbf{L}\|_2^2 - \min_{\mathbf{u}\in\mathcal{K}}\{\mathbf{u}\cdot\mathbf{L}\}\} \geq \min_{\|\mathbf{L}\|_2 = a}\{-\frac{1}{2}\|\mathbf{L}\|_2^2 - (-\mathbf{L})\cdot\mathbf{L}\} = \frac{1}{2}a^2 \ ,$$

where we used the fact that $-\mathbf{L} \in \mathcal{K}$ if $\|\mathbf{L}\|_2 = a$. In addition, by Lemma 3.9,

$$\rho_2(a) = \min{}_{\|\mathbf{L}\|_2 \leq a, \|\mathbf{l}\|_2 = 1}\{-\mathbf{l}^\top(-I)\mathbf{l}\} = 1 \ .$$

By Theorem 3.10, we have that $\max_t\{R_{A,t}\} \geq \min\{a^2/(2\eta), (\eta/2)Q\}$, and for $\eta = \frac{a}{\sqrt{Q}}$, $\max_t\{R_{A,t}\} \geq \frac{a}{2}\sqrt{Q}$.

3.4.2 The Hedge Algorithm

Hedge is the most notable regret minimization algorithm for the expert setting. We have $\mathcal{K} = \Delta_N$, and the algorithm gives a weight $p_{i,t+1} = \frac{p_{i,0}e^{-\eta L_{i,t}}}{\sum_{j=1}^N p_{j,0}e^{-\eta L_{j,t}}}$ to expert i at time $t+1$, where the initial weights $p_{i,0}$ and the learning rate η are parameters, and $\sum_{i=1}^N p_{i,0} = 1$.

It is easy to see that for the potential $\Phi_\eta(\mathbf{L}) = -(1/\eta)\ln(\sum_{i=1}^N p_{i,0}e^{-\eta L_i})$, we have that $\mathbf{p} = (p_1, \ldots, p_N) = \nabla\Phi_\eta(\mathbf{L})$. The Hessian $\nabla^2\Phi_\eta$ has the following simple form:

Lemma 3.14. *Let $\mathbf{L} \in \mathbb{R}^N$ and denote $\mathbf{p} = \nabla\Phi_\eta(\mathbf{L})$. Then $\nabla^2\Phi_\eta(\mathbf{L}) = \eta \cdot (\mathbf{p}\mathbf{p}^\top - diag(\mathbf{p})) \preceq 0$, where $diag(\mathbf{p})$ is the diagonal matrix with $\mathbf{p}$ as its diagonal.*

The proof is given in Section 3.5. We will assume $p_{1,0} = \ldots = p_{N,0} = 1/N$, and write $Hed(\eta)$ for Hedge with parameters η and the uniform distribution. Thus, $\Phi(\mathbf{L}) = -\ln((1/N)\sum_{i=1}^N e^{-L_i})$, and we have by Lemma 3.14 that $\frac{\partial^2\Phi}{\partial L_i \partial L_j} > 0$ for every $i \neq j$. Therefore, by Lemma 3.12, $\rho_1, \rho_2 > 0$. We now need to calculate ρ_1 and ρ_2. This is straightforward in the case of ρ_1, but for ρ_2 we give the value only for $N = 2$.

Lemma 3.15. *For any $N \geq 2$, $\rho_1(a) = \ln \frac{N}{N-1+e^{-a}}$. For $N = 2$, it holds that $\rho_2(a) = \left(e^{a/2} + e^{-a/2}\right)^{-2}$.*

Proof. By Lemma 3.12,

$$\begin{aligned}\rho_1(a) &= \min_{\mathbf{L}\in\mathcal{N}(a)}\{\Phi(\mathbf{L})\} - \Phi(\mathbf{0}) = \min_{\mathbf{L}\in\mathcal{N}(a)}\{\Phi(\mathbf{L})\} = -\ln \max_{\mathbf{L}\in\mathcal{N}(a)}\left\{\frac{1}{N}\sum_{i=1}^{N} e^{-L_i}\right\} \\ &= -\ln\left(\frac{1}{N}((N-1)e^{-0} + e^{-a})\right) = \ln\frac{N}{N-1+e^{-a}}\,.\end{aligned}$$

In addition, $\rho_2(a) = \min_{\mathbf{L}\in[0,a]^N, \mathbf{l}\in\mathcal{N}(1)}\{-\mathbf{l}^\top\nabla^2\Phi(\mathbf{L})\mathbf{l}\} > 0$. For $N = 2$, we have $\mathcal{N}(1) = \{(1,0),(0,1)\}$. Therefore, $-\mathbf{l}^\top\nabla^2\Phi(\mathbf{L})\mathbf{l}$ is either $-\nabla^2\Phi(\mathbf{L})_{1,1}$ or $-\nabla^2\Phi(\mathbf{L})_{2,2}$. Denoting $\mathbf{p} = \mathbf{p}(\mathbf{L}) = \nabla\Phi(\mathbf{L})$, we have that $-\nabla^2\Phi(\mathbf{L})_{1,1} = p_1 - p_1^2$ and $-\nabla^2\Phi(\mathbf{L})_{2,2} = p_2 - p_2^2$ by Lemma 3.14. Since $p_1 = 1 - p_2$, we have $p_1 - p_1^2 = p_2 - p_2^2$. Now,

$$\begin{aligned}p_1(1-p_1) &= \frac{e^{-L_1}}{e^{-L_1}+e^{-L_2}}\cdot\frac{e^{-L_2}}{e^{-L_1}+e^{-L_2}} = e^{-(L_1+L_2)}\left(e^{-L_1}+e^{-L_2}\right)^{-2} \\ &= \left(e^{(L_2-L_1)/2} + e^{(L_1-L_2)/2}\right)^{-2}\,.\end{aligned}$$

The function $(e^x + e^{-x})^{-2}$ is decreasing for $x \geq 0$, so the smallest value is attained for $|L_1 - L_2| = a$. Thus, for $N = 2$, $\rho_2(a) = (e^{a/2} + e^{-a/2})^{-2}$. □

Picking $a = 1.2$, we have by Theorem 3.13 that

Theorem 3.16. *For $N = 2$, there exists η s.t.*

$$\max_t\{R_{Hed(\eta),t}\} \geq \sqrt{\rho_1(a)\rho_2(a)q/2} \geq 0.195\sqrt{q}\,.$$

For a general N we may still lower bound $\max_t\{R_{Hed(\eta),t}\}$ by providing a lower bound for $\rho_2(a)$. We use the fact that the term $-\mathbf{l}^\top\nabla^2\Phi(\mathbf{L})\mathbf{l}$, which is minimized in the definition of $\rho_2(a)$, may be interpreted as the variance of a certain discrete and bounded random variable. Thus, the key element that will be used is a general lower bound on the variance of such variables. This bound, possibly of separate interest, is given in the following lemma, whose proof can be found in Section 3.5.

Lemma 3.17. *Let $x_1 < \ldots < x_N \in \mathbb{R}$, $0 < p_1, \ldots, p_N < 1$, and $\sum_{k=1}^{N} p_k = 1$, and let X be a random variable that obtains the value x_k with probability p_k, for every $1 \leq k \leq N$. Then for every $1 \leq i < j \leq N$, $Var(X) \geq \frac{p_i p_j}{p_i + p_j}\cdot(x_i - x_j)^2$, and equality* *is attained iff* $N = 2$*, or* $N = 3$ *with* $i = 1$*,* $j = 3$*, and* $x_2 = \frac{p_1 x_1 + p_3 x_3}{p_1 + p_3}$*. In particular,* $Var(X) \geq \frac{p_1 p_N}{p_1 + p_N} \cdot (x_N - x_1)^2 \geq \frac{1}{2} \min_k\{p_k\}(x_N - x_1)^2$.

We comment that for the special case $p_1 = \ldots = p_N = 1/N$, Lemma 3.17 yields the inequality $Var(X) \geq (x_N - x_1)^2/(2N)$, sometimes referred to as the Szőkefalvi Nagy inequality [88].

We now proceed to lower bound $\rho_2(a)$ for any N.

Lemma 3.18. *For any* $N \geq 2$*, it holds that* $\rho_2(a) \geq \frac{1}{2(N-1)e^a + 2}$*.*

Proof. Recall that $\rho_2(a) = \min_{\mathbf{L} \in [0,a]^N, \mathbf{l} \in \mathcal{N}(1)}\{-\mathbf{l}^\top \nabla^2 \Phi(\mathbf{L}) \mathbf{l}\}$. By Lemma 3.14, $\nabla^2 \Phi(\mathbf{L}) = \mathbf{p}\mathbf{p}^\top - diag(\mathbf{p})$, where $\mathbf{p} = \mathbf{p}(\mathbf{L}) = \nabla \Phi(\mathbf{L})$, and thus,

$$-\mathbf{l}^\top \nabla^2 \Phi(\mathbf{L}) \mathbf{l} = \mathbf{l}^\top diag(\mathbf{p}) \mathbf{l} - \mathbf{l}^\top \mathbf{p}\mathbf{p}^\top \mathbf{l} = \sum_{i=1}^N p_i l_i^2 - \left(\sum_{i=1}^N p_i l_i\right)^2 .$$

Assume for now that $l_1, \ldots, l_N$ are distinct. Then $-\mathbf{l}^\top \nabla^2 \Phi(\mathbf{L}) \mathbf{l} = Var(X)$, where X is a random variable obtaining the value l_i with probability p_i, $i = 1, \ldots, N$. By Lemma 3.17, $Var(X) \geq \frac{1}{2} \min_k\{p_k\}(\max_i\{l_i\} - \min_i\{l_i\})^2$. For $\mathbf{l} \in \mathcal{N}(1)$ it holds that $\max_i\{l_i\} = 1$ and $\min_i\{l_i\} = 0$. In addition, for $\mathbf{L} \in [0,a]^N$, the smallest value for $\min_k\{p_k\}$ is obtained when one entry of $\mathbf{L}$ is a and the rest are 0, yielding that $\min_k\{p_k\} \geq \frac{e^{-a}}{N-1+e^{-a}}$. Together we have that $-\mathbf{l}^\top \nabla^2 \Phi(\mathbf{L}) \mathbf{l} \geq \frac{1}{2} \cdot \frac{e^{-a}}{N-1+e^{-a}} = \frac{1}{2(N-1)e^a+2}$. Since the expression $-\mathbf{l}^\top \nabla^2 \Phi(\mathbf{L}) \mathbf{l}$ is continuous in $\mathbf{l}$, this inequality holds even if the assumption that the entries of $\mathbf{l}$ are distinct numbers is dropped. The result follows. □

Theorem 3.19. *For any* $0 < \alpha < 1/N$ *and* $\eta > 0$*, it holds that* $\max_t\{R_{Hed(\eta),t}\} \geq \min\{\frac{1}{\eta} \ln(\frac{N(1-\alpha)}{N-1}), \frac{1}{4}\eta\alpha q\}$*, and for* $\alpha = 1/(2N)$ *and* $\eta = \sqrt{(8N/q) \ln \frac{2N-1}{2N-2}}$*,* $\max_t\{R_{Hed(\eta),t}\} \geq \frac{\sqrt{q}}{4N} \cdot \sqrt{\frac{2N}{2N-1}} = \Omega(\sqrt{q}/N)$*.*

Proof. By Theorem 3.13, for every $\eta > 0$, it holds that $\max_t\{R_{Hed(\eta),t}\} \geq \min\{\frac{\rho_1(a)}{\eta}, \frac{\eta}{2}\rho_2(a) q\}$. We fix a, observing that any $a \in (0, \infty)$ may be used, and denote $\alpha = \frac{1}{(N-1)e^a+1}$. Note that α may thus obtain any value in $(0, 1/N)$. By Lemma 3.18, we have that $\rho_2(a) \geq \alpha/2$. By Lemma 3.15, $\rho_1(a) = \ln \frac{N}{N-1+e^{-a}}$, and since

$$\frac{N(1-\alpha)}{N-1} = \frac{N}{N-1} \cdot \frac{(N-1)e^a}{(N-1)e^a+1} = \frac{Ne^a}{(N-1)e^a+1} = \frac{N}{N-1+e^{-a}} ,$$

we have that $\rho_1(a) = \ln \frac{N(1-\alpha)}{N-1}$. Thus, $\max_t\{R_{Hed(\eta),t}\} \geq \min\{\frac{1}{\eta} \ln \frac{N(1-\alpha)}{N-1}, \frac{1}{4}\eta\alpha q\}$. We may maximize the r.h.s. by picking $\eta = \sqrt{\frac{4}{\alpha q} \ln \frac{N(1-\alpha)}{N-1}}$, yielding $\max_t\{R_{Hed(\eta),t}\} \geq$

$\sqrt{\frac{1}{4}\alpha q \ln \frac{N(1-\alpha)}{N-1}}$. We may further pick a s.t. $\alpha = 1/(2N)$ and obtain

$$\begin{aligned}\max_t\{R_{Hed(\eta),t}\} &\geq \sqrt{\frac{q}{8N}\ln\frac{N-1/2}{N-1}} = \sqrt{\frac{q}{8N}\ln\left(1+\frac{1}{2N-2}\right)} \\ &\geq \sqrt{\frac{q}{8N}\cdot\frac{\frac{1}{2N-2}}{1+\frac{1}{2N-2}}} = \sqrt{\frac{q}{8N}\cdot\frac{1}{2N-1}} \\ &= \frac{\sqrt{q}}{4N}\cdot\sqrt{\frac{2N}{2N-1}}\,,\end{aligned}$$

where the second inequality uses the fact that $\ln(1+x) \geq \frac{x}{x+1}$ for $x \in (-1, \infty)$. $\square$

3.5 Appendix: Additional Claims and Missing Proofs

Lemma 3.20. *If $\mathcal{K} \subset \mathbb{R}^N$ is compact, then $h(\mathbf{L}) = \min_{\mathbf{u}\in\mathcal{K}}\{\mathbf{u}\cdot\mathbf{L}\}$ is continuous on $\mathbb{R}^N$.*

Proof. First of all, $h(\mathbf{L})$ is well-defined, because for every $\mathbf{L} \in \mathbb{R}^N$, $\mathbf{u}\cdot\mathbf{L}$ is continuous and $\mathcal{K}$ is compact. Similarly, $\|\cdot\|_2$ is continuous on $\mathcal{K}$, so $D = \max_{\mathbf{u}\in\mathcal{K}}\{\|\mathbf{u}\|_2\}$ is well-defined. Let $\mathbf{L}, \mathbf{L}' \in \mathbb{R}^N$, and let $\mathbf{u}, \mathbf{u}' \in \mathcal{K}$ satisfy $h(\mathbf{L}) = \mathbf{u}\cdot\mathbf{L}$ and $h(\mathbf{L}') = \mathbf{u}'\cdot\mathbf{L}'$. Therefore,

$$\begin{aligned}h(\mathbf{L}') - h(\mathbf{L}) &= \mathbf{u}'\cdot\mathbf{L}' - \mathbf{u}\cdot\mathbf{L} \leq \mathbf{u}\cdot\mathbf{L}' - \mathbf{u}\cdot\mathbf{L} = \mathbf{u}\cdot(\mathbf{L}'-\mathbf{L}) \\ &\leq \|\mathbf{u}\|_2\cdot\|\mathbf{L}'-\mathbf{L}\|_2 \leq D\|\mathbf{L}'-\mathbf{L}\|_2\,,\end{aligned}$$

where the second inequality is by the Cauchy-Schwarz inequality. Reversing the roles of $\mathbf{L}$ and $\mathbf{L}'$, we get $h(\mathbf{L}) - h(\mathbf{L}') \leq D\|\mathbf{L}-\mathbf{L}'\|_2$, and together we get $|h(\mathbf{L}') - h(\mathbf{L})| \leq D\|\mathbf{L}-\mathbf{L}'\|_2$, which means that h is continuous. $\square$

Lemma 3.21. *It holds that $\delta : \mathbb{R}^N \to \mathbb{R}$, $\delta(\mathbf{v}) = \max_i\{v_i\} - \min_i\{v_i\}$, is a seminorm.*

Proof. Positive homogeneity: if $a \geq 0$, then $\delta(a\mathbf{v}) = \max_i\{av_i\} - \min_i\{av_i\} = a\max_i\{v_i\} - a\min_i\{v_i\} = |a|\delta(\mathbf{v})$. Otherwise, $\delta(a\mathbf{v}) = \max_i\{av_i\} - \min_i\{av_i\} = -\min_i\{-av_i\} + \max_i\{-av_i\} = \delta(-a\mathbf{v})$, and as already seen, $\delta(-a\mathbf{v}) = |a|\delta(\mathbf{v})$.

Triangle inequality: given $\mathbf{u}, \mathbf{v} \in \mathbb{R}^N$, it holds for every j that $\min_i\{u_i\}+\min_i\{v_i\} \leq u_j + v_j \leq \max_i\{u_i\} + \max_i\{v_i\}$. Thus, $\max_i\{u_i + v_i\} \leq \max_i\{u_i\} + \max_i\{v_i\}$, and

$\min_i\{u_i\} + \min_i\{v_i\} \le \min_i\{u_i + v_i\}$. Together we get

$$\begin{aligned}\delta(\mathbf{u}+\mathbf{v}) &= \max_i\{u_i + v_i\} - \min_i\{u_i + v_i\} \\ &\le \max_i\{u_i\} + \max_i\{v_i\} - \min_i\{u_i\} - \min_i\{v_i\} \\ &= \delta(\mathbf{u}) + \delta(\mathbf{v}) \ .\end{aligned}$$

□

Proof of Theorem 3.4: We start by observing that

$$\begin{aligned}(-1/\eta)\mathcal{R}^*(-\eta\mathbf{L}) &= (-1/\eta)\sup_{\mathbf{x}\in\mathcal{K}}\{\mathbf{x}\cdot(-\eta\mathbf{L}) - \mathcal{R}(\mathbf{x})\} \\ &= \inf_{\mathbf{x}\in\mathcal{K}}\{\mathbf{x}\cdot\mathbf{L} + \mathcal{R}(\mathbf{x})/\eta\} \\ &= \min_{\mathbf{x}\in\mathcal{K}}\{\mathbf{x}\cdot\mathbf{L} + \mathcal{R}(\mathbf{x})/\eta\} \ .\end{aligned}$$

We used the fact that $\mathbf{x}\cdot\mathbf{L} + \mathcal{R}(\mathbf{x})/\eta$ is continuous and attains a minimum on a compact set. Consequently, Φ is defined for every $\mathbf{L}$. Since $\mathcal{R}^*$ is convex, $\Phi(\mathbf{L}) = (-1/\eta)\mathcal{R}^*(-\eta\mathbf{L})$ is concave.

Denote $h(\mathbf{u},\mathbf{L}) = \mathbf{u}\cdot\mathbf{L} + \mathcal{R}(\mathbf{u})/\eta$, and let $\mathbf{L},\mathbf{L}' \in \mathbb{R}^N$, where $\mathbf{L} \neq \mathbf{L}'$. In addition, let $\mathbf{x} = \arg\min_{\mathbf{u}\in\mathcal{K}}\{h(\mathbf{u},\mathbf{L})\}$ and $\mathbf{x}' = \arg\min_{\mathbf{u}\in\mathcal{K}}\{h(\mathbf{u},\mathbf{L}')\}$, and denote $\Delta\mathbf{L} = \mathbf{L}' - \mathbf{L}$ and $\Delta\mathbf{x} = \mathbf{x}' - \mathbf{x}$. We have

$$\Phi(\mathbf{L}') - \Phi(\mathbf{L}) = h(\mathbf{x}',\mathbf{L}') - h(\mathbf{x},\mathbf{L}) \le h(\mathbf{x},\mathbf{L}') - h(\mathbf{x},\mathbf{L}) = \mathbf{x}\cdot\Delta\mathbf{L} \ ,$$

where the inequality is by definition of $\mathbf{x}'$. Reversing the roles of $\mathbf{L}$ and $\mathbf{L}'$, we also get $\Phi(\mathbf{L}) - \Phi(\mathbf{L}') \le -\mathbf{x}'\cdot\Delta\mathbf{L}$. Therefore, $\mathbf{x}'\cdot\Delta\mathbf{L} \le \Phi(\mathbf{L}') - \Phi(\mathbf{L}) \le \mathbf{x}\cdot\Delta\mathbf{L}$, or $\Delta\mathbf{x}\cdot\Delta\mathbf{L} \le \Phi(\mathbf{L}') - \Phi(\mathbf{L}) - \mathbf{x}\cdot\Delta\mathbf{L} \le 0$, implying that

$$|\Phi(\mathbf{L}') - \Phi(\mathbf{L}) - \mathbf{x}\cdot\Delta\mathbf{L}| \le |\Delta\mathbf{x}\cdot\Delta\mathbf{L}| \le \|\Delta\mathbf{x}\|_2\|\Delta\mathbf{L}\|_2 \ ,$$

or $\frac{|\Phi(\mathbf{L}+\Delta\mathbf{L}) - \Phi(\mathbf{L}) - \mathbf{x}\cdot\Delta\mathbf{L}|}{\|\Delta\mathbf{L}\|_2} \le \|\Delta\mathbf{x}\|_2$. If we showed that $\lim_{\Delta\mathbf{L}\to\mathbf{0}}\|\Delta\mathbf{x}\|_2 = 0$, then

$$\lim_{\Delta\mathbf{L}\to\mathbf{0}} \frac{\Phi(\mathbf{L}+\Delta\mathbf{L}) - \Phi(\mathbf{L}) - \mathbf{x}\cdot\Delta\mathbf{L}}{\|\Delta\mathbf{L}\|_2} = 0 \ ,$$

that is, $\nabla\Phi(\mathbf{L}) = \mathbf{x}$, as required. Furthermore, the fact that Φ is differentiable on $\mathbb{R}^N$

would imply that it is continuously differentiable (see, e.g., Corollary 25.5.1 in [77]).

Since $\mathcal{R}$ is strongly convex (with some parameter $\alpha > 0$), we have that $\mathcal{R}(\frac{1}{2}\mathbf{x} + \frac{1}{2}\mathbf{x}') \leq \frac{1}{2}\mathcal{R}(\mathbf{x}) + \frac{1}{2}\mathcal{R}(\mathbf{x}') - \frac{\alpha}{2} \cdot \frac{1}{2} \cdot \frac{1}{2}\|\Delta\mathbf{x}\|_2^2$. Dividing by η and adding $\mathbf{L} \cdot (\frac{1}{2}\mathbf{x} + \frac{1}{2}\mathbf{x}')$ to both sides, we get

$$\begin{aligned}\frac{1}{\eta}\mathcal{R}(\frac{\mathbf{x}+\mathbf{x}'}{2}) + \mathbf{L} \cdot (\frac{\mathbf{x}+\mathbf{x}'}{2}) \leq \frac{1}{2\eta}\mathcal{R}(\mathbf{x}) + \mathbf{L} \cdot \frac{1}{2}\mathbf{x} + \frac{1}{2\eta}\mathcal{R}(\mathbf{x}') \\ + \mathbf{L} \cdot \frac{1}{2}\mathbf{x}' - \frac{\alpha}{8\eta}\|\Delta\mathbf{x}\|_2^2 \,,\end{aligned}$$

or

$$h(\frac{\mathbf{x}+\mathbf{x}'}{2}, \mathbf{L}) \leq \frac{1}{2}h(\mathbf{x}, \mathbf{L}) + \frac{1}{2}h(\mathbf{x}', \mathbf{L}) - \frac{\alpha}{8\eta}\|\Delta\mathbf{x}\|_2^2 \,.$$

By definition of $\mathbf{x}$, $h(\frac{\mathbf{x}+\mathbf{x}'}{2}, \mathbf{L}) \geq h(\mathbf{x}, \mathbf{L})$, so, rearranging, we get

$$h(\mathbf{x}', \mathbf{L}) - h(\mathbf{x}, \mathbf{L}) \geq \frac{\alpha}{4\eta}\|\Delta\mathbf{x}\|_2^2 \,.$$

Since

$$\begin{aligned}h(\mathbf{x}', \mathbf{L}) &= \mathbf{x}' \cdot \mathbf{L} + \frac{1}{\eta}\mathcal{R}(\mathbf{x}') = \mathbf{x}' \cdot \mathbf{L}' + \frac{1}{\eta}\mathcal{R}(\mathbf{x}') - \mathbf{x}' \cdot \Delta\mathbf{L} \\ &= h(\mathbf{x}', \mathbf{L}') - \mathbf{x}' \cdot \Delta\mathbf{L} \leq h(\mathbf{x}, \mathbf{L}') - \mathbf{x}' \cdot \Delta\mathbf{L} \,,\end{aligned}$$

we have that

$$\begin{aligned}\frac{\alpha}{4\eta}\|\Delta\mathbf{x}\|_2^2 &\leq h(\mathbf{x}, \mathbf{L}') - \mathbf{x}' \cdot \Delta\mathbf{L} - h(\mathbf{x}, \mathbf{L}) \\ &= \mathbf{x} \cdot \mathbf{L}' - \mathbf{x}' \cdot \Delta\mathbf{L} - \mathbf{x} \cdot \mathbf{L} = -\Delta\mathbf{x} \cdot \Delta\mathbf{L} \\ &\leq \|\Delta\mathbf{x}\|_2\|\Delta\mathbf{L}\|_2 \,,\end{aligned}$$

and thus, $\|\Delta\mathbf{x}\|_2 \leq \frac{4\eta}{\alpha}\|\Delta\mathbf{L}\|_2$. Therefore, $\lim_{\Delta\mathbf{L}\to\mathbf{0}} \|\Delta\mathbf{x}\|_2 = 0$, completing the proof. □

Proof of Lemma 3.14: We assume w.l.o.g. that $p_{i,0} > 0$ for every i. Let $1 \leq i, j \leq N$. We have that

$$\frac{\partial}{\partial L_i}\left(\ln\frac{\partial\Phi_\eta}{\partial L_j}\right) = \frac{\partial^2\Phi_\eta}{\partial L_i \partial L_j} \cdot \left(\frac{\partial\Phi_\eta}{\partial L_j}\right)^{-1} = \frac{\partial^2\Phi_\eta}{\partial L_i \partial L_j}p_j^{-1} \,.$$

Since

$$\ln \frac{\partial \Phi_\eta}{\partial L_j} = \ln \left(\frac{p_{i,0} e^{-\eta L_j}}{\sum_{k=1}^N p_{k,0} e^{-\eta L_k}} \right) = \ln p_{i,0} - \eta L_j + \eta \Phi_\eta \ ,$$

we also have that

$$\frac{\partial}{\partial L_i} \left(\ln \frac{\partial \Phi_\eta}{\partial L_j} \right) = -\eta \frac{\partial L_j}{\partial L_i} + \eta \frac{\partial \Phi_\eta}{\partial L_i} = -\eta \cdot 1_{i=j} + \eta p_i \ ,$$

and therefore, $\frac{\partial^2 \Phi_\eta}{\partial L_i \partial L_j} p_j^{-1} = -\eta \cdot 1_{i=j} + \eta p_i$, or, equivalently, $\frac{\partial^2 \Phi_\eta}{\partial L_i \partial L_j} = \eta(p_i p_j - 1_{i=j} p_j)$, as required.

For any $\mathbf{p} \in \Delta_N$ and $\mathbf{z} \in \mathbb{R}^N$, we denote by $X_{\mathbf{p},\mathbf{z}}$ a random variable which satisfies $\mathbb{P}(X_{\mathbf{p},\mathbf{z}} = z_i) = \sum_{\{j: z_j = z_i\}} p_j$ for every $1 \le i \le N$. Note that even though the values z_i are not necessarily distinct, nevertheless, $\mathbb{E}[X_{\mathbf{p},\mathbf{z}}] = \sum_{i=1}^N p_i z_i$ and $Var(X_{\mathbf{p},\mathbf{z}}) = \sum_{i=1}^N p_i z_i^2 - (\sum_{i=1}^N p_i z_i)^2$ as usual, by continuity. Thus, for any $\mathbf{z} \in \mathbb{R}^N$, we have that

$$\begin{aligned} 0 \le Var(X_{\mathbf{p},\mathbf{z}}) &= \sum_{i=1}^N p_i z_i^2 - \left(\sum_{i=1}^N p_i z_i \right)^2 = \mathbf{z}^\top diag(\mathbf{p}) \mathbf{z} - \sum_{i=1}^N \sum_{j=1}^N p_i z_i p_j z_j \\ &= \mathbf{z}^\top diag(\mathbf{p}) \mathbf{z} - \mathbf{z}^\top \mathbf{p}\mathbf{p}^\top \mathbf{z} = \mathbf{z}^\top (diag(\mathbf{p}) - \mathbf{p}\mathbf{p}^\top) \mathbf{z} \ . \end{aligned}$$

We therefore have that $diag(\mathbf{p}) - \mathbf{p}\mathbf{p}^\top = -(1/\eta) \nabla^2 \Phi_\eta(\mathbf{L})$ is positive semi-definite, and the result follows. □

Proof of Lemma 3.17: Denoting $y_k = x_k - \mathbb{E}[X]$ for $1 \le k \le N$ and $p = \frac{p_i}{p_i + p_j}$, we have that

$$\begin{aligned} Var(X) &\ge p_i y_i^2 + p_j y_j^2 \ge p_i y_i^2 + p_j y_j^2 - (p_i + p_j)(p y_i + (1-p) y_j)^2 \\ &= (p_i + p_j)(p y_i^2 + (1-p) y_j^2 - (p y_i + (1-p) y_j)^2) \\ &= (p_i + p_j)((p - p^2) y_i^2 + (p - p^2) y_j^2 - 2(p - p^2) y_i y_j) \\ &= (p_i + p_j)(p - p^2)(y_i - y_j)^2 = \frac{p_i p_j}{p_i + p_j} (y_i - y_j)^2 \\ &= \frac{p_i p_j}{p_i + p_j} (x_i - x_j)^2 \ . \end{aligned}$$

To obtain an equality, the first inequality must become an equality, necessitating that $x_k = \mathbb{E}[X]$ for every $k \ne i, j$, and thus, $N \le 3$. If $N = 2$, the first inequality is an equality, and $p y_i + (1-p) y_j = p_1 x_1 + p_2 x_2 - \mathbb{E}[X] = 0$, so the second above inequality is also an equality. For $N = 3$, if $x_k = \mathbb{E}[X]$, then the first inequality becomes an equality, and in addition,

$$\begin{aligned} py_i + (1-p)y_j &= \frac{p_i x_i + p_j x_j}{p_i + p_j} - \mathbb{E}[X] = \frac{\mathbb{E}[X] - p_k x_k}{p_i + p_j} - \mathbb{E}[X] \\ &= \frac{\mathbb{E}[X](1-p_k)}{p_i + p_j} - \mathbb{E}[X] = 0 \ , \end{aligned}$$

so the second inequality becomes an equality. Thus, we have an equality iff $x_k = \mathbb{E}[X]$, or equivalently, $i = 1$, $k = 2$, $j = 3$, and $x_2 = p_1 x_1 + p_2 x_2 + p_3 x_3$.

Finally, we have in particular that

$$\begin{aligned} Var(X) &\geq \frac{p_1 p_N}{p_1 + p_N} \cdot (x_N - x_1)^2 = \frac{1}{1/p_1 + 1/p_N} \cdot (x_N - x_1)^2 \\ &\geq \frac{1}{2} \min_k \{p_k\} (x_N - x_1)^2 \ , \end{aligned}$$

completing the proof. □

3.5.1 An Extension of the FTL-BTL Lemma

The upper bound on the regret of the RFTL algorithm relies on the "follow the leader, be the leader" (FTL-BTL) lemma (see [47]). We extend that lemma to facilitate the proof of a lower bound as well. The following lemma gives the core inductive argument of the extension as well as the original, for completeness.

Lemma 3.22. *Let X be a non-empty set, let $1 \leq n \in \mathbb{N}$, and let $g_i : X \to \mathbb{R}$ for $1 \leq i \leq n$. Denote $G_i = \sum_{j=1}^{i} g_j$, and assume G_i has at least one global minimum on X. Denote x_i for such a minimum for $1 \leq i \leq n$, and let $x_0 \in X$ be an arbitrary point. Then $\sum_{j=1}^{n} g_j(x_j) \leq G_n(x_n) \leq \sum_{j=1}^{n} g_j(x_{j-1})$.*

Proof. By induction on n. For $n = 1$, $\sum_{j=1}^{n} g_j(x_j) = G_1(x_1) \leq G_1(x_0) = g_1(x_0) = \sum_{j=1}^{n} g_j(x_{j-1})$, where the inequality is by definition of x_1. We assume correctness for n and proceed to prove for $n+1$. We have on the one hand that

$$\begin{aligned} G_{n+1}(x_{n+1}) &\leq G_{n+1}(x_n) = G_n(x_n) + g_{n+1}(x_n) \\ &\leq \sum_{j=1}^{n} g_j(x_{j-1}) + g_{n+1}(x_n) = \sum_{j=1}^{n+1} g_j(x_{j-1}) \ , \end{aligned}$$

where the first inequality is by definition of x_{n+1}, and the second one is by the induction assumption. On the other hand,

$$G_{n+1}(x_{n+1}) = G_n(x_{n+1}) + g_{n+1}(x_{n+1}) \geq G_n(x_n) + g_{n+1}(x_{n+1})$$
$$\geq \sum_{j=1}^{n} g_j(x_j) + g_{n+1}(x_{n+1}) = \sum_{j=1}^{n+1} g_j(x_j) \ ,$$

where the first inequality is by definition of x_n, and the second inequality is by the induction assumption. $\square$

We now give the extended FTL-BTL lemma. We assume the RFTL setting, so that $\mathcal{K} \subset \mathbb{R}^N$ is non-empty, compact, and convex, $\mathcal{R} : \mathcal{K} \to \mathbb{R}$ is strongly convex, and $\eta > 0$.

Lemma 3.23. *Let $\mathbf{l}_t \in \mathbb{R}^N$ for $1 \leq t \leq T$, and let $\mathbf{u}_0 = \arg\min_{\mathbf{u}\in\mathcal{K}}\{\mathbf{u} \cdot \sum_{t=1}^{T} \mathbf{l}_t\}$. Then*

$$R_{RFTL(\eta,\mathcal{R}),T} \leq \sum_{t=1}^{T} \mathbf{l}_t \cdot (\mathbf{x}_t - \mathbf{x}_{t+1}) + (\mathcal{R}(\mathbf{u}_0) - \mathcal{R}(\mathbf{x}_1))/\eta$$

and

$$R_{RFTL(\eta,\mathcal{R}),T} \geq (\mathcal{R}(\mathbf{x}_{T+1}) - \mathcal{R}(\mathbf{x}_1))/\eta + \sum_{t=1}^{T} \mathbf{l}_t \cdot (\mathbf{x}_{T+1} - \mathbf{u}_0) \geq 0 \ .$$

Proof. To fit the terminology of Lemma 3.22, set $X = \mathcal{K}$, $n = T+1$, $g_1(\mathbf{x}) = \mathcal{R}(\mathbf{x})/\eta$, and $g_{t+1}(\mathbf{x}) = \mathbf{l}_t \cdot \mathbf{x}$ for $1 \leq t \leq T$. Thus, $G_{t+1}(\mathbf{x}) = \mathcal{R}(\mathbf{x})/\eta + \mathbf{x} \cdot \sum_{\tau=1}^{t} \mathbf{l}_\tau$ for $0 \leq t \leq T$, and $\mathbf{x}_{t+1} = \arg\min_{\mathbf{u}\in\mathcal{K}} \{\mathcal{R}(\mathbf{u})/\eta + \mathbf{u} \cdot \sum_{\tau=1}^{t} \mathbf{l}_\tau\}$. In addition, we set $\mathbf{x}_0 = \mathbf{x}_1$. By Lemma 3.22,

$$\mathcal{R}(\mathbf{x}_1)/\eta + \sum_{t=1}^{T} \mathbf{l}_t \cdot \mathbf{x}_{t+1} \leq \mathcal{R}(\mathbf{x}_{T+1})/\eta + \sum_{t=1}^{T} \mathbf{l}_t \cdot \mathbf{x}_{T+1} \leq \mathcal{R}(\mathbf{x}_1)/\eta + \sum_{t=1}^{T} \mathbf{l}_t \cdot \mathbf{x}_t \ ,$$

and thus

$$\sum_{t=1}^{T} \mathbf{l}_t \cdot \mathbf{x}_{t+1} - \sum_{t=1}^{T} \mathbf{l}_t \cdot \mathbf{x}_t \leq (\mathcal{R}(\mathbf{x}_{T+1}) - \mathcal{R}(\mathbf{x}_1))/\eta + \sum_{t=1}^{T} \mathbf{l}_t \cdot \mathbf{x}_{T+1} - \sum_{t=1}^{T} \mathbf{l}_t \cdot \mathbf{x}_t \leq 0 \ .$$

We focus first on the left inequality. By definition of $\mathbf{x}_{T+1}$, $G_{T+1}(\mathbf{x}_{T+1}) \leq G_{T+1}(\mathbf{u}_0)$, implying that $\mathcal{R}(\mathbf{x}_{T+1})/\eta + \sum_{t=1}^{T} \mathbf{l}_t \cdot \mathbf{x}_{T+1} \leq \mathcal{R}(\mathbf{u}_0)/\eta + \sum_{t=1}^{T} \mathbf{l}_t \cdot \mathbf{u}_0$. Therefore, $\sum_{t=1}^{T} \mathbf{l}_t \cdot \mathbf{x}_{t+1} - \sum_{t=1}^{T} \mathbf{l}_t \cdot \mathbf{x}_t \leq (\mathcal{R}(\mathbf{u}_0) - \mathcal{R}(\mathbf{x}_1))/\eta + \sum_{t=1}^{T} \mathbf{l}_t \cdot \mathbf{u}_0 - \sum_{t=1}^{T} \mathbf{l}_t \cdot \mathbf{x}_t$, and rearranging,

$$R_{RFTL(\eta,\mathcal{R}),T} \leq \sum_{t=1}^{T} \mathbf{l}_t \cdot (\mathbf{x}_t - \mathbf{x}_{t+1}) + (\mathcal{R}(\mathbf{u}_0) - \mathcal{R}(\mathbf{x}_1))/\eta \ ,$$

which is the known upper bound expression on the regret of RFTL.

Moving on to the other inequality, we have that

$$\begin{aligned} R_{RFTL(\eta,\mathcal{R}),T} &= \sum_{t=1}^{T} \mathbf{l}_t \cdot \mathbf{x}_t - \sum_{t=1}^{T} \mathbf{l}_t \cdot \mathbf{u}_0 \\ &\geq (\mathcal{R}(\mathbf{x}_{T+1}) - \mathcal{R}(\mathbf{x}_1))/\eta + \left(\sum_{t=1}^{T} \mathbf{l}_t \cdot \mathbf{x}_{T+1} - \sum_{t=1}^{T} \mathbf{l}_t \cdot \mathbf{u}_0 \right) \ . \end{aligned}$$

It holds that $(\mathcal{R}(\mathbf{x}_{T+1}) - \mathcal{R}(\mathbf{x}_1))/\eta \geq 0$ by definition of $\mathbf{x}_1$, and $\sum_{t=1}^{T} \mathbf{l}_t \cdot \mathbf{x}_{T+1} - \sum_{t=1}^{T} \mathbf{l}_t \cdot \mathbf{u}_0 \geq 0$ by definition of $\mathbf{u}_0$. □

We note that a lower bound on the regret of RFTL may also be shown more directly. Specifically, setting $\mathbf{L}_0 = \mathbf{0}$ and $\mathbf{x}_0 = \mathbf{x}_1$, it holds that

$$\begin{aligned} \sum_{t=1}^{T} \mathbf{x}_t \cdot \mathbf{l}_t &= \sum_{t=1}^{T} (\mathbf{x}_t \cdot \mathbf{L}_t + \mathcal{R}(\mathbf{x}_t)/\eta) - \sum_{t=1}^{T} (\mathbf{x}_t \cdot \mathbf{L}_{t-1} + \mathcal{R}(\mathbf{x}_t)/\eta) \\ &\geq \sum_{t=1}^{T} (\mathbf{x}_t \cdot \mathbf{L}_t + \mathcal{R}(\mathbf{x}_t)/\eta) - \sum_{t=1}^{T} (\mathbf{x}_{t-1} \cdot \mathbf{L}_{t-1} + \mathcal{R}(\mathbf{x}_{t-1})/\eta) \\ &= (\mathbf{x}_T \cdot \mathbf{L}_T + \mathcal{R}(\mathbf{x}_T)/\eta) - (\mathbf{x}_0 \cdot \mathbf{L}_0 + \mathcal{R}(\mathbf{x}_0)/\eta) \\ &= (\mathcal{R}(\mathbf{x}_T) - \mathcal{R}(\mathbf{x}_1))/\eta + (\mathbf{x}_T \cdot \mathbf{L}_T - \mathbf{u}_0 \cdot \mathbf{L}_T) + \mathbf{u}_0 \cdot \mathbf{L}_T \ , \end{aligned}$$

where the inequality is by definition of $\mathbf{x}_t$. Since $(\mathcal{R}(\mathbf{x}_T) - \mathcal{R}(\mathbf{x}_1))/\eta \geq 0$ by definition of $\mathbf{x}_1$ and $\mathbf{x}_T \cdot \mathbf{L}_T \geq \mathbf{u}_0 \cdot \mathbf{L}_T$ by definition of $\mathbf{u}_0$, the regret of RFTL must be non-negative. We point out that the above lower bound features $\mathbf{x}_T$ instead of $\mathbf{x}_{T+1}$, which appears in the extended FTL-BTL lemma. This difference may be eliminated by using $T - 1$ instead of T in the proof of the lemma.

Chapter 4

Regret Minimization for Branching Experts

4.1 Introduction

In this chapter, we design algorithms for the best expert setting with refined regret bounds in terms of the number of experts, N. Our algorithms are able to control regret when N may grow over time, but the experts' cumulative losses enjoy specific patterns that naturally occur in practical scenarios. These patterns are captured by the notion of "branching experts", in which the addition of new experts to the pool creates a tree structure. Although our algorithms are designed for this branching experts setting, when applied to the standard N-expert setting they deliver regret bounds where the "complexity" term $\log N$ associated with the set of N experts does not occur.

For the sake of illustration, consider the following sequential path-planning problem on a graph [59]. The number N of available paths from source to destination (i.e., experts in the game) is very large, and the loss of a path picked at time t is the sum over the current costs (at the same time t) of the edges included in the path. One may expect that in T prediction steps only a small number Λ_T of paths have become "leaders" at any time $t \leq T$, where a path i is leader at time t if its cumulative loss $L_{i,t}$ is the smallest over all paths. We show that a variant of Hedge, played over a growing pool of leaders,[1] achieves a regret bound that only depends on the number of leaders, rather than on the number of experts. In general, in a game with N experts and Λ_T leaders, we prove that the regret of our modified Hedge is $O\big(\Lambda_T(1+\ln L_T^*) + \sqrt{L_T^*\Lambda_T}\big)$,

[1] Note that in this example a leader can be found efficiently by solving a shortest path problem.

independent of N. Our result is actually phrased in a more general model, where the notion of leader is parameterized by a value $\alpha \geq 0$ and Λ_T is replaced by $\Lambda_{\alpha,T}$. This value represents the edge an expert must have over the loss of the previous leaders in order to become leader itself.

A second natural scenario we consider is one where the cumulative losses of all experts remain clustered around the loss values of a few experts. Intuitively, as far as regret is concerned experts that have similar cumulative losses are interchangeable. Hence, working with one representative in each cluster is a convenient approximation of the original problem, and one expects the regret to be controlled by the number of clusters, rather than by the overall number of experts. As before, we make a reduction to the setting of a growing pool of experts. We start with a single cluster (all experts start off with zero loss) and then split a cluster (i.e., increase the pool of experts by at least one) whenever the largest cumulative loss difference within the cluster exceeds some threshold that depends on a size parameter $\alpha \geq 0$. We prove that the regret of our Hedge variant is at most of order $\mathcal{N}_{\alpha,T}(1 + \alpha\mathcal{N}_{\alpha,T})(1 + \ln L_T^*) + \sqrt{L_T^* \mathcal{N}_{\alpha,T}}$, if the experts may indeed be divided into $\mathcal{N}_{\alpha,T}$ α-similar subsets after T steps.

In both of the above settings, of few leading experts and of clustered experts, our algorithm is essentially optimal: We prove that the main terms of both regret bounds, $\sqrt{L_T^* \Lambda_{\alpha,T}}$ and $\sqrt{L_T^* \mathcal{N}_{\alpha,T}}$, are only improvable by constant factors. The same result is also proven for the general case of a growing set of experts.

We then turn our attention to the bandit setting. Here, as a motivating scenario for the growing set of experts, consider a framework where we apply heuristics to solve a sequence of instances of a hard optimization problem. This task can be naturally cast in a bandit setting, where just a single heuristic is tested on each instance. Now suppose that new heuristics become available as time goes by, and we add them to the pool of candidate heuristics in order to improve our chances. Some of them might be variants of other heuristics in the pool, and some others might be completely new. In all cases, we would like to control the regret of the bandit algorithm with respect to the growing pool of heuristics. Now, if a variant i' of some heuristic i already in the pool becomes available at time t, then it is reasonable to compute the regret against a pair of compound experts that use i up to time t, and from then on either i or its variant i'. On the other hand, if a heuristic k unrelated to any other in the pool is added at time t, then we want to compare to a pair of compound experts that use the best heuristic up to time t and then either stick to it, or switch to k.

In this context, we introduce a new non-trivial modification of Exp3, and show that its expected regret is at most of order $\bigl(1+(\ln f)/(\ln N_T)\bigr)\sqrt{TN_T \ln N_T}$, where N_T is the final number of experts, and f stands for the product of the degrees of nodes along the branch leading to the best expert. This factor may be as small as $\Theta(1)$ and is always bounded by K^{d_T}, where d_T is the number of time steps in which some new expert was added to the pool, and K is the branching factor of the tree of experts.

4.2 Branching Experts with Full Information

We consider a modification of prediction with expert advice where the adversary gradually reveals the experts to the learner. Specifically, the game between adversary and learner starts with one known expert. As the game progresses, the adversary may choose to reveal the existence of more experts. Once an expert is revealed, the learner may start using it by placing some weight on its decisions. The regret of a learner at the end of the game is measured w.r.t. all revealed experts.

Each newly revealed expert is given some history, in the form of a starting loss closely related to the cumulative loss of one of the previously revealed experts. In our setup we therefore consider each newly revealed expert as an approximate clone of an existing expert in terms of its cumulative loss. From that point on, the new expert is allowed to freely diverge from its parent. Let $1, \dots, N_t$ be the indices of the experts revealed after the first t rounds, with $N_0 = 1$. This process of approximate cloning results in a tree structure where, at any time $t \geq 1$, the root is the initial expert (the one at time $t = 0$) and the leaves correspond to the N_t experts participating in the game at time t. For convenience, when an expert is cloned, one of the clones is used to represent the continuation of the original expert (i.e., it is a perfect clone). We now describe the game in more detail.

For each round $t = 1, 2, \dots, T$

1. For each expert $i = 1, \dots, N_{t-1}$, the adversary reveals a set $C(i,t)$ of experts containing i itself and possibly additional approximate clones. The adversary also reveals the past losses $L_{j,t-2}, l_{j,t-1}$ for each $j \in C(i,t)$. The new experts are indexed by $N_{t-1}+1, \dots, N_t$.

2. The learner chooses $\mathbf{p}_t = \big(p_{1,t}, \dots, p_{N_t,t}\big)$.

3. The adversary reveals losses $\mathbf{l}_t = \big(l_{1,t}, \dots, l_{N_t,t}\big)$, and the learner suffers a loss $\mathbf{p}_t \cdot \mathbf{l}_t$.

The reason why the pair $L_{j,t-2}, l_{j,t-1}$ is revealed instead of its sum $L_{j,t-2} + l_{j,t-1}$ is explained in Footnote 4. We use $m(t) = \arg\min_{i=1,\dots,N_t}\{L_{i,t}\}$ to denote the index of the best expert in the first t steps (where we take the smallest such index in case of a tie), so that $L^*_T = L_{m(T),T}$. The regret of an algorithm A is then defined as usual, $R_{A,T} = L_{A,T} - L^*_T$. In order to have the standard expert scenario as a special case of the branching one, we set $L_{j,t}, l_{j,t} = 0$ for $t \le 0$ and for all j. In the standard expert scenario $N_1 = |C(1,1)|$ corresponds to the number N of experts. In order to facilitate a comparison to the standard expert bounds, we express our branching experts bounds in terms of this quantity N_1.

In order to elucidate the combinatorial tree structure underlying this notion of regret, consider the simple case when $L_{j,t-1} = L_{i,t-1}$ for all $j \in C(i,t)$, all $i = 1, \dots, N_{t-1}$, and all t. Hence, each new expert j starts off as a perfect clone of its parent expert i. Now, at each time T the learner is competing with the set of N_T compound experts associated with the paths $i_0, i_1, \dots, i_T$, where i_0 is the "root expert" and $i_t \in C(i_{t-1}, t)$ for all $t = 1, \dots, T$. This combinatorial interpretation will help in comparing our results to other settings of compound experts.

When clones are not perfect, we use $\alpha_{i,t-1} = \max\big\{\big|L_{j,t-2} - L_{k,t-2}\big| \,:\, j,k \in C(i,t)\big\}$ to denote the "diameter" of a split.[2] Since the learner has no control over $\alpha_{i,t}$ at each split, it may suffer the sum of those elements from the root to the leaf of the best expert in the form of regret. Furthermore, the splits increase the number of experts, which means that upper bounds on the regret must also grow. Interestingly, there are natural scenarios where the above quantities are small. In these scenarios, described in

[2] Note that perfect cloning is not equivalent to a split with diameter zero. This discrepancy is due to technical issues that arise in Section 4.5.

Section 4.5, the role of revealing new experts is taken from the adversary and *given to the learner*. Specifically, all the experts are known at the beginning of the game, and it is the learner that decides to gradually start using some of them. Generally speaking, in such scenarios the vast majority of experts are either poor candidates for being the best expert, or perform comparably to some other experts, making them safe for the learner to ignore. The important observation to be made is that any algorithm and analysis that hold when the splits are determined by the adversary may also be used if the splits are determined by the learner. Although referring to the best overall expert rather than the best revealed expert clearly increases the regret, this difference will be bounded for the scenarios we consider, and easily taken into account.

4.3 Related Work

Prediction with expert advice when the pool of experts varies over time has been investigated in the past. The model of "sleeping experts", or "specialists" [15, 39], allows a subset A_t of the experts to be awake at each time step t, where A_t is determined by a time-selection function. The regret is then measured against each expert only on the time steps when it was awake. Although we may view our setting of branching experts as a special case of sleeping experts (where experts are progressively woken up), their notion of regret is different from ours, since we compete against *compound* experts, associated with paths on the expert tree. Therefore, the two settings are incomparable.

Branching experts can also be viewed as special cases of more general combinatorial constructions, like the permutation experts of [60] or the shifting experts of [53]. Although in some cases these more general settings can be extended to accommodate growing sets of experts —see, e.g., [83], the resulting regret bounds are much worse than ours, mainly due to the dependence on $\ln(N_T T)$, where N_T is the total number of experts in the pool after T steps. Our bounds, instead, depend in a more detailed way on the structure of the expert tree, and replace $\ln(N_T T)$ with $\ln \Pi$, where Π depends on the splits on the branches that the leading experts belong to, and can thus be much smaller than N_T —see discussion in the next paragraph.

In the case of perfect cloning, which was mentioned in Section 4.2, Hedge may directly simulate the branching experts if we limit the number of rounds in which splits occur to at most d, and the maximal degree of a split to at most K, where d and K are preliminarily available. Taking $N_1 = 1$, this limits the final number of compound experts to at most K^d. An application of standard Hedge then leads to the regret bound $d \ln K + \sqrt{2L_T^* d \ln K}$. Our modified Hedge, called Hed_C, virtually obtains the same bound (with slightly worse constants) without preliminary knowledge of the tree structure. However, we may do much better in cases where there are few splits along the branches where the set of leading experts occur. In fact, the main term in the regret bound of Hed_C is of order $\sqrt{L_T^* \ln \Pi}$, where Π can be $\Theta(1)$ even when N_T is exponential in T.

It is also instructive to consider a type of scenario in which any advantage of Hed_C over Hedge is removed. Specifically, assume that all the splits are announced in the initial rounds of the game. In these initial rounds no losses are incurred, and afterwards, the game proceeds with a fixed number of experts. It is possible to design such a scenario with $N+1$ experts (for a large enough N) such that the regret of Hed_C is $\Omega\big(\sqrt{L_T^* N}\big)$ while Hedge has the usual regret bound of $O\big(\ln N + \sqrt{L_T^* \ln N}\big)$ (see Claim 4.13 in Section 4.8). Since Theorem 4.3 in Section 4.4 gives an $O\big(N + \sqrt{L_T^* N}\big)$ regret for this case, it is arguably a worst-case scenario for Hed_C. Clearly enough, in order to make Hed_C achieve a similar regret as Hedge in this hard case, without losing much of its advantage in the others, it suffices to add a meta-Hedge algorithm aggregating the two.

In the bandit setting, our modification of Exp3 for branching experts, called Exp3.C, has an expected regret bound of the order $\big(1 + (\ln f)/(\ln N_T)\big)\sqrt{T N_T \ln N_T}$, containing the factor $\sqrt{N_T}$ (see Section 4.7). It seems plausible that the modification of Exp3 for shifting experts, Exp3.S, could be extended to a growing pool of experts along the lines of [83]. The resulting regret bound would be of the order $\sqrt{T N_T S_T \ln(N_T T)}$, where S_T is the number of times $i_t \neq i_{t+1}$ in the path $i_0, i_1, \ldots, i_T$ from the root to the best action $i_T = m(T)$. These two bounds appear to be incomparable.

We now move on to discussing related work in the standard N-expert setting, with fixed N. There are a few examples of expert algorithms whose regret bound does not depend on the number of experts. The most trivial example is Follow the Leader (FTL), namely, the algorithm that deterministically picks the current best expert. It is easy to see that the regret of FTL is bounded by the number of times the leader changes, *no matter if the same few leaders* keep alternating. Another trivial example is Hedge run with uneven initial weights, which implicitly assumes using a prior that peaks on a small set containing the best expert. A substantially less trivial example is the NormalHedge algorithm of [26], and its refinement due to [27]. Except for constant factors, a bound on the cumulative loss of these algorithms can be written as $\inf_{0\le\epsilon\le 1}\{L_{\sigma(\epsilon N),T} + \sqrt{T\ln(1/\epsilon)}\}$, where $\sigma(1),\ldots,\sigma(N)$ is a permutation of expert indices such that $L_{\sigma(1),T} \le \cdots \le L_{\sigma(N),T}$. This bound is incomparable to our bounds for few leaders and clustered experts. Indeed, it is easy to find examples where our bounds dominate NormalHedge's. (Think of one expert with constant total loss, and the remaining $N-1$ experts with linear losses always clustered around a single value: Then ϵ must be $1/N$ to ensure sublinear regret, while Λ_T and $\mathcal{N}_{\alpha,T}$ can be made constant.)

Finally, we could apply the Fixed Share algorithm of [53] for shifting experts to the few leaders setting. By doing so, we would get a regret bound of the order of $\sqrt{TS_T\ln(NT)}$, where S_T is the number of times the current leader changes. Now, even in the best case for Fixed Share (i.e., when $S_T = \Lambda_T - 1$) our bound for Hed_C is still better by a factor of at least $\sqrt{\ln(NT)}$. Using more sophisticated shifting algorithms, like Mixing Past Posteriors [18], may improve on the $S_T \ln N$ term in the Fixed Share bound, but it does not affect the other term $S_T \ln T$. Recently, [61] gave a surprising Bayesian interpretation to Mixing Past Posteriors. Finding a similarly efficient Bayesian formulation of our branching experts construction is an open problem.

The few leaders and the clustered experts settings take advantage of specific "suboptimalities" in the loss sequence chosen by the adversary. Adaptive procedures able to take advantage of such suboptimalities were also proposed by [75]. Relating those results to the ones derived in this chapter remains an interesting open problem.

4.4 Adapting Hedge for the Branching Setup

The main change that is required to handle the new setup is deciding on weights for newly revealed experts. We will handle this problem by applying a general mechanism we term a *partial restart*. A partial restart redistributes the weights $w_{i,t}$ of existing experts among all experts, old and new, without changing the sum of the weights. Following this redistribution, the usual exponential update step takes place.

Hedge with Partial Restarts

For each round $t = 1, 2, \ldots, T$

1. Get the new experts $N_{t-1} + 1, \ldots, N_t$ together with their past losses $L_{j,t-2}, l_{j,t-1}$ for $j = N_{t-1} + 1, \ldots, N_t$.
2. From $w_{1,t-1}, \ldots, w_{N_{t-1},t-1}$ compute new weights $w'_{1,t-1}, \ldots, w'_{N_t,t-1} \geq 0$ such that $W'_{t-1} = W_{t-1}$, where $W'_{t-1} = \sum_{i=1}^{N_t} w'_{i,t-1}$.
3. Update the new weights: $w_{i,t} = w'_{i,t-1} e^{-\eta l_{i,t-1}}$, and set $p_{i,t} = w_{i,t}/W_t$, for each $i = 1, \ldots, N_t$.
4. Observe losses $\mathbf{l}_t = \big(l_{1,t}, \ldots, l_{N_t,t}\big)$ and suffer loss $\mathbf{p}_t \cdot \mathbf{l}_t$.

Note that an ordinary (full) restart is equivalent to a partial restart where all experts are assigned equal weights. A partial restart, in contrast to a full restart, may preserve more information about the preceding run, depending on how $w'_{i,t}$ are defined.

If an algorithm is an augmentation of Hedge with partial restarts, its loss is upper bounded by an expression that depends explicitly on the number of restarts, and only implicitly on their exact nature. This bound is given in the next lemma.

Lemma 4.1. *Let $0 < \eta \leq 1$ and let A be an augmentation of Hedge with at most n partial restarts. Then $\ln(W_{T+1}/W_1) \leq -(1 - e^{-\eta})(L_{A,T} - n)$.*

Proof. Recall that $w_{i,t+1} = w'_{i,t} e^{-\eta l_{i,t}}$ and $W'_t = W_t$ for every $i = 1, \ldots, N_{t+1}$, $t = 1, \ldots, T$ and note that $W'_t > 0$. Let $p'_{i,t} = w'_{i,t}/W'_t$ for all $i = 1, \ldots, N_{t+1}$, so that $\mathbf{p}'_t = (p'_{1,t}, \ldots, p'_{N_{t+1},t})$ is a probability vector. Hence, for every t,

$$\ln \frac{W_{t+1}}{W_t} = \ln \frac{\sum_{i=1}^{N_{t+1}} w_{i,t+1}}{W'_t} = \ln \frac{\sum_{i=1}^{N_{t+1}} w'_{i,t} e^{-\eta l_{i,t}}}{W'_t} = \ln \sum_{i=1}^{N_{t+1}} p'_{i,t} e^{-\eta l_{i,t}} \; .$$

Now, if there is no restart at time $t + 1$, that is $|C(i, t + 1)| = 1$ and $w'_{i,t} = w_{i,t}$ for all $i = 1, \ldots, N_t$, then $N_{t+1} = N_t$ and $p'_{i,t} = p_{i,t}$, $i = 1, \ldots, N_t$. In this case, by the convexity of $f(x) = e^x$, we have $e^{-\eta l_{i,t}} \le 1 - \big(1 - e^{-\eta}\big)\, l_{i,t}$. Hence we can write

$$\begin{aligned}
\ln \sum_{i=1}^{N_{t+1}} p'_{i,t} e^{-\eta l_{i,t}} &= \ln \sum_{i=1}^{N_t} p_{i,t} e^{-\eta l_{i,t}} \\
&\le \ln \sum_{i=1}^{N_t} p_{i,t}\big(1 - (1 - e^{-\eta}) l_{i,t}\big) \\
&\le -(1 - e^{-\eta}) \mathbf{p}_t \cdot \mathbf{l}_t \\
&= -(1 - e^{-\eta}) l_{A,t} \; .
\end{aligned}$$

On the other hand, if a restart takes place, then we have

$$\ln \sum_{i=1}^{N_{t+1}} p'_{i,t} e^{-\eta l_{i,t}} \le \ln \sum_{i=1}^{N_{t+1}} p'_{i,t}\big(1 - (1 - e^{-\eta}) l_{i,t}\big) \le -(1 - e^{-\eta}) \sum_{i=1}^{N_{t+1}} p'_{i,t} l_{i,t}$$

which is trivially upper bounded by $-(1-e^{-\eta})(l_{A,t}-1)$, since $\sum_{i=1}^{N_{t+1}} p'_{i,t} l_{i,t} \ge 0 \ge l_{A,t}-1$.

Thus, if $n' \le n$ is the number of restarts,

$$\ln \frac{W_{T+1}}{W_1} = \sum_{t=1}^{T} \ln \frac{W_{t+1}}{W_t} \le -(1 - e^{-\eta})\big(L_{A,T} - n'\big) \le -(1 - e^{-\eta})\big(L_{A,T} - n\big) \; .$$

□

We now describe a specific way of defining the weights $w'_{i,t}$ out of the weights $w_{i,t}$, and call Hed$_C$ the resulting partially restarting variant of Hedge. The algorithm starts by setting $w_{j,1} = 1$ for every expert $j = 1, \ldots, N_1$, where $N_1 = |C(1,1)|$.[3] At time $t+1$, its partial restart stage distributes the weight of a parent expert i among experts in $C(i, t+1)$ while maintaining the same weight proportions as ordinary Hedge. Namely,

$$w'_{j,t} = w_{i,t} \, \frac{e^{-\eta L_{j,t-1}}}{\sum_{k \in C(i,t+1)} e^{-\eta L_{k,t-1}}} \; , \qquad \text{for every } j \in C(i, t+1)$$

(recall that $L_{j,t-1}$, for $j \in C(i, t+1)$, are all revealed to the algorithm).[4] Note that when $|C(i, t+1)| = 1$ —that is, expert i is not cloned at time $t+1$, then $w'_{i,t} = w_{i,t}$. As a step towards bounding the regret of Hed$_C$, we first lower bound the weights it assigns to experts.

[3] Alternatively, we may set $w_{1,0} = 1$ for the single expert existing at time $t = 0$. This causes a minor difference in the regret bounds.

[4] The two-stage update $w_{i,t} \to w'_{i,t} \to w_{i,t+1}$ is the reason why pairs of losses $L_{j,t-1}$ and $l_{j,t}$ are provided for each new expert $i = N_t + 1, \ldots, N_{t+1}$.

Lemma 4.2. *If i is an expert at time $t \geq 1$, and $i_0, \ldots, i_t = i$ are the experts along the path from the root to i in the branching experts tree, then*

$$w_{i_t,t} \geq \exp\left(-\eta L_{i_t,t-1} - \eta \sum_{\tau=1}^{t-1} \alpha_{i_\tau,\tau}\right) \prod_{\tau=1}^{t-1} \max\Big\{1, 2\big|C(i_\tau, \tau+1)\big| - 2\Big\}^{-1} .$$

Proof. Recall that $i_0, \ldots, i_t$ is a path from the root i_0 to expert i_t in the branching experts tree, where $i_\tau \in C(i_{\tau-1}, \tau)$ for all $\tau = 1, \ldots, t$. We first prove by induction that

$$w_{i_t,t} \geq e^{-\eta L_{i_t,t-1}} \prod_{\tau=1}^{t-1} \Big(1 + \big(|C(i_\tau, \tau+1)| - 1\big) e^{\eta \alpha_{i_\tau,\tau}}\Big)^{-1} .$$

For $t = 1$ both sides equal 1 and the claim is trivial. We next assume the claim holds for t and prove it for $t+1$. For every t we have that

$$\begin{aligned}\sum_{j \in C(i_t,t+1)} e^{-\eta L_{j,t-1}} &= e^{-\eta L_{i_t,t-1}} \sum_{j \in C(i_t,t+1)} e^{-\eta (L_{j,t-1} - L_{i_t,t-1})} \\ &\leq e^{-\eta L_{i_t,t-1}} \Big(1 + \big(|C(i_t, t+1)| - 1\big) e^{\eta \alpha_{i_t,t}}\Big) ,\end{aligned}$$

where we used $\alpha_{i_t,t} = \max\left\{\big|L_{j,t-1} - L_{k,t-1}\big| \,:\, j,k \in C(i_t, t+1)\right\}$ and also $i_t \in C(i_t, t+1)$. Thus

$$\begin{aligned}w_{i_{t+1},t+1} = w'_{i_{t+1},t} e^{-\eta l_{i_{t+1},t}} &= w_{i_t,t} \frac{e^{-\eta L_{i_{t+1},t}}}{\sum_{j \in C(i_t,t+1)} e^{-\eta L_{j,t-1}}} \\ &\geq w_{i_t,t} e^{-\eta \left(L_{i_{t+1},t} - L_{i_t,t-1}\right)} \Big(1 + \big(|C(i_t, t+1)| - 1\big) e^{\eta \alpha_{i_t,t}}\Big)^{-1} \\ &\geq e^{-\eta L_{i_{t+1},t}} \prod_{\tau=1}^{t} \Big(1 + \big(|C(i_\tau, \tau+1)| - 1\big) e^{\eta \alpha_{i_\tau,\tau}}\Big)^{-1}\end{aligned}$$

completing the induction. It is easy to verify that

$$1 + \big(|C(i_\tau, \tau+1)| - 1\big) e^{\eta \alpha_{i_\tau,\tau}} \leq \max\big\{1, 2|C(i_\tau, \tau+1)| - 2\big\} e^{\eta \alpha_{i_\tau,\tau}}$$

and thus

$$\begin{aligned}w_{i_t,t} &\geq e^{-\eta L_{i_t,t-1}} \prod_{\tau=1}^{t-1} \Big(\max\big\{1, 2|C(i_\tau, \tau+1)| - 2\big\} e^{\eta \alpha_{i_\tau,\tau}}\Big)^{-1} \\ &= \exp\left(-\eta L_{i_t,t-1} - \eta \sum_{\tau=1}^{t-1} \alpha_{i_\tau,\tau}\right) \prod_{\tau=1}^{t-1} \max\big\{1, 2|C(i_\tau, \tau+1)| - 2\big\}^{-1} .\end{aligned}$$

□

We can now combine Lemmas 4.1 and 4.2 to upper bound the regret of Hed_C. In what follows we denote $\Pi_t = N_1 \prod_{\tau=1}^{t-1} \max\Big\{1, 2\big|C(i_\tau, \tau+1)\big| - 2\Big\}$ and $\mathcal{A}_t = \sum_{\tau=1}^{t-1} \alpha_{i_\tau,\tau}$, where $i_0, \ldots, i_t = m(t)$ is a path from the root to the best expert at time t. Note that neither Π_t nor $\mathcal{A}_t$ are monotone in t because the index of the best expert $m(t)$ changes over time. We also denote d_t for the number of rounds in $2, \ldots, t$ in which splits occurred.

Theorem 4.3. *Let $\mathcal{L}'$ and Π' be known upper bounds on $L_T^* + \mathcal{A}_{T+1}$ and Π_{T+1}, respectively. If $\eta = \ln\big(1 + \sqrt{(2\ln\Pi')/\mathcal{L}'}\big)$, then the regret of* Hed_C *satisfies*

$$R_{Hed_C,T} \le d_T + \mathcal{A}_{T+1} + \ln\Pi' + \sqrt{2\mathcal{L}'\ln\Pi'}\ .$$

Proof. By Lemma 4.2 we have that

$$\ln\frac{W_{T+1}}{W_1} \ge \ln\frac{w_{m(T),T+1}}{N_1} \ge -\eta L_T^* - \eta\mathcal{A}_{T+1} - \ln\Pi_{T+1} \ge -\eta\big(L_T^* + \mathcal{A}_{T+1}\big) - \ln\Pi'\ .$$

In addition, by Lemma 4.1, $\ln(W_{T+1}/W_1) \le -(1-e^{-\eta})(L_{Hed_C,T} - d_T)$. Combining the upper and lower bounds for $\ln(W_{T+1}/W_1)$ and rearranging, we have

$$L_{Hed_C,T} - d_T \le \frac{\eta\big(L_T^* + \mathcal{A}_{T+1}\big) + \ln\Pi'}{1 - e^{-\eta}}\ .$$

The rest of the proof is similar to the standard proof of Hedge and is given here for completeness. Denote $\nu = \ln\Pi'$ and $\mathcal{L} = L_T^* + \mathcal{A}_{T+1}$. It is easily verified that $\eta \le \frac{1}{2}(e^\eta - e^{-\eta})$ for every $\eta \ge 0$, and therefore,

$$L_{Hed_C,T} - d_T \le \frac{\frac{1}{2}(e^\eta - e^{-\eta})\mathcal{L} + \nu}{1 - e^{-\eta}} = \frac{\nu}{1 - e^{-\eta}} + \frac{\mathcal{L}}{2}(e^\eta + 1)\ .$$

Recall that we set η such that $e^\eta = 1 + \sqrt{2\nu/\mathcal{L}'}$, and therefore

$$\begin{aligned} R_{Hed_C,T} &= L_{Hed_C,T} + \mathcal{A}_{T+1} - \mathcal{L} \le d_T + \mathcal{A}_{T+1} + \frac{\nu}{1 - e^{-\eta}} + \frac{\mathcal{L}}{2}(e^\eta - 1) \\ &\le d_T + \mathcal{A}_{T+1} + \frac{\nu}{1 - e^{-\eta}} + \frac{\mathcal{L}'}{2}(e^\eta - 1) \\ &= d_T + \mathcal{A}_{T+1} + \Big(1 + \sqrt{\mathcal{L}'/(2\nu)}\Big)\nu + \frac{\mathcal{L}'}{2}\sqrt{2\nu/\mathcal{L}'} \\ &= d_T + \mathcal{A}_{T+1} + \nu + \sqrt{2\mathcal{L}'\nu} \end{aligned}$$

completing the proof. $\square$

In the standard N-expert setting, $\alpha_{i,t} = 0$ for all i and t, implying $\mathcal{A}_{T+1} = 0$. Moreover, $d_T = 0$ and $\Pi_{T+1} = N$. Therefore, in this special case Theorem 4.3 recovers the standard regret bound of Hedge.

The requirement in Theorem 4.3 that bounds be known in advance may be relaxed by using a doubling trick. Since the index of the best expert $m(t)$ changes over time, the doubling is applied to a bound $\mathcal{L}'_\tau$ on the quantity $\max_{t\le\tau}\{L^*_t + \mathcal{A}_{t+1}\}$ and to a bound ν_τ on the quantity $\ln\bigl(\max_{t\le\tau}\{\Pi_{t+1}\}\bigr)$. For any single value of $\mathcal{L}'_\tau$, ν_τ is doubled $O(\ln \nu_T)$ times, and the total regret of runs is $O\Bigl(\bigl(1 + d_T + \mathcal{A}\bigr)\ln \nu_T + \nu_T + \sqrt{\mathcal{L}'_\tau \nu_T}\Bigr)$, where we denote $\mathcal{A} = \max_{t\le T}\{\mathcal{A}_{t+1}\}$. Adding up these values for all doubled values of $\mathcal{L}'_\tau$ yields a regret bound of

$$R_{Hed_C,T} = O\left(\Bigl(\bigl(1 + d_T + \mathcal{A}\bigr)\ln \nu_T + \nu_T\Bigr)\ln \mathcal{L}'_{T+1} + \sqrt{\mathcal{L}'_{T+1}\nu_T}\right) . \qquad (4.1)$$

Note that $\mathcal{L}'_{T+1} \le 2\max_{t\le T}\{L^*_t + \mathcal{A}_{t+1}\} \le 2(L^*_T + \mathcal{A})$, where the first inequality holds since $\mathcal{L}'_{T+1}$ is doubled only until it exceeds $\max_{t\le T}\{L^*_t + \mathcal{A}_{t+1}\}$. This gives $\sqrt{(\mathcal{L}'_{T+1}/2)\nu_T} \le \sqrt{L^*_T\nu_T} + \frac{1}{2}(\mathcal{A} + \nu_T)$. Therefore, (4.1) gives the following.

Corollary 4.4. *Let* $\mathcal{A} = \max_{t\le T}\{\mathcal{A}_{t+1}\}$ *and* $\Pi = \max_{t\le T}\{\Pi_{t+1}\}$. *Applying a doubling trick to* Hed_C *yields the regret bound*

$$R_{Hed_C,T} = O\left(\Bigl(1 + \ln\bigl(L^*_T + \mathcal{A}\bigr)\Bigr)\bigl((d_T + \mathcal{A})\ln\ln\Pi + \ln\Pi\bigr) + \sqrt{L^*_T \ln \Pi}\right) .$$

Finally, we point out that if $K \ge 2$ is the maximal degree of splits in the tree for $t > 1$, then $\Pi \le N_1(2K-2)^{d_T}$, and the main term $\sqrt{L^*_T \ln \Pi}$ in the above regret bound becomes of order $\sqrt{\bigl(\ln N_1 + d_T \ln K\bigr)L^*_T}$.

4.5 Applications

Few leading experts. We consider a best expert scenario with N experts, where the set of experts that happen to be "leaders" throughout the game is small. The set of all-time leaders (leader set, for short) includes initially only the first expert. In every round the current best expert is added to the set iff its current cumulative loss is strictly smaller than the cumulative loss of all experts *in the leader set*. We generalize this definition by requiring that the advantage over all experts in the leader set must be strictly greater than $\alpha \geq 0$, where α is a parameter. Formally, the leader set starts as $S_1 = \{1\}$, and at the beginning of each round $t > 1$, $S_t = S_{t-1} \cup \{m(t-1)\}$ iff $L_{m(t-1),t-1} + \alpha < L_{j,t-1}$ for every expert $j \in S_{t-1}$. In adversarial branching terms, we will consider such a new leader as "revealed" at time t. It will branch off the former best expert i in S_{t-1}, namely, the one that satisfies $L_{i,t-2} \leq L_{j,t-2}$ for every $j \in S_{t-1}$ (where the smallest index is taken in a tie). Thus a split will always have two children: a previous leader and the new one.

We denote the number of leaders in the first T steps by $\Lambda_{\alpha,T}$. We assume that upper bounds $\Lambda'_\alpha \geq \Lambda_{\alpha,T}$ and $L' \geq L^*_T$ are known in advance (although the identities of the leaders are not), where a doubling trick is used to guess both quantities. The learner may run Hed_C while simulating an adversary that reveals experts gradually, making Theorem 4.3 applicable with the following settings: $N_1 = 1$, $d_T = \Lambda_{\alpha,T} - 1$, $\mathcal{A}_{t+1} \leq d_T \max\{\alpha, 1-\alpha\}$, and $\Pi_{t+1} \leq 2^{d_T}$. This provides a bound on the regret w.r.t. the best *revealed* expert. Since the cumulative loss of the best overall expert is smaller by at most $\alpha + 1$, we simply need to add $\alpha + 1$ to this bound to get a bound on the regret. We thus obtain the following.

Theorem 4.5. *Let $\alpha_1 = \max\{\alpha, 1-\alpha\}$. In the few leading experts scenario, the regret of* Hed_C *run with parameters $\eta = \ln\Big(1 + \sqrt{(2\ln 2)(\Lambda'_\alpha - 1)/(L' + (\Lambda'_\alpha - 1)\alpha_1)}\Big)$, $\mathcal{L}' = L' + (\Lambda'_\alpha - 1)\alpha_1$, and $\Pi' = 2^{\Lambda'_\alpha - 1}$ satisfies $R_{Hed_C,T} = O\Big(\Lambda'_\alpha(\alpha+1) + \sqrt{(\Lambda'_\alpha - 1)L'}\Big)$.*

We point out that there is a trade-off in the choice of α, since an increase in α causes a decrease in $\Lambda_{\alpha,T}$. A doubling trick may again be applied to guess both Λ'_α and L': When either bound is violated, the bound is doubled and the algorithm is restarted.

Corollary 4.6. *Applying a doubling trick to* Hed_C *in the few leading experts scenario yields a regret bound of $O\Big(\Lambda_{\alpha,T}(\alpha+1)(1+\ln L^*_T) + \sqrt{L^*_T \Lambda_{\alpha,T}}\Big)$.*

Clustered experts. We next consider a best expert scenario where experts may be divided into a small number of subsets such that the cumulative losses inside each subset are "similar" at all times. Intuitively, working with one representative of each subset instead of the individual experts is a good approximation for the original problem. An important difference is that the number of representatives may be much smaller than the number of experts, making the regret bound better. Given the approximated regret bound, the maximal "diameter" of the subsets may be added to obtain a regret bound for the original problem.

Formally, let $\alpha \geq 0$ be predetermined by the learner, and let $\mathcal{N}_{\alpha,T}$ be the number of subsets of experts that are *α-similar* after T steps. Namely, for every $t = 1, \ldots, T$ and every experts i and j in the same subset, $|L_{i,t} - L_{j,t}| \leq \alpha$. As before, we start by assuming we know upper bounds $\mathcal{N}'_\alpha \geq \mathcal{N}_{\alpha,T}$ and $L' \geq L^*_T$, and eventually relax this assumption using a doubling trick.

The learner will implement Hed_C in conjunction with the following splitting scheme. Initially, all experts reside in the same cluster. For every t, a cluster is split at the beginning of time $t+1$ iff the difference between the cumulative losses of any two experts inside it at time t exceeds $\beta = (2\mathcal{N}'_\alpha - 1)\alpha$. To split a cluster, we first sort its members by their cumulative loss at time t. Then we find the largest gap between cumulative loss values and split there. (If more than one maximal gap exists, we pick one arbitrarily.) Since $\beta \geq (2\mathcal{N}_{\alpha,T} - 1)\alpha$ and all the subsets are α-similar, this gap must be larger than α. Furthermore, members of any given subset cannot be on both sides of the gap. Next, if the gap in either of the two parts is larger than β, the process is repeated. Thus, β is an upper bound on the diameter of each cluster (i.e., the largest difference $|L_{i,T} - L_{j,T}|$ over pairs of experts i, j in the cluster) after any number T of steps. In addition, since clusters always contain entire subsets, the total number of splits after T steps does not exceed $\mathcal{N}_{\alpha,T} - 1$.

We apply Theorem 4.3 with the settings $N_1 = 1$, $d_T \leq \mathcal{N}'_\alpha - 1$, and $\mathcal{A}_{t+1} \leq \beta d_T \leq (2\mathcal{N}'_\alpha - 1)(\mathcal{N}'_\alpha - 1)\,\alpha$. As for Π', recall that it upper bounds $\Pi_{T+1} = \prod_{\tau=1}^{T} \max\{1, 2|C(i_\tau, \tau+1)| - 2\}$, since $N_1 = 1$. Let $n_1, \ldots, n_k$ be the values of $2|C(i_\tau, \tau+1)| - 2$ in the product that are not zero. We have

$$\Pi_{T+1} = \prod_{i=1}^{k} n_i \leq \left(\frac{1}{k}\sum_{i=1}^{k} n_i\right)^k \leq \left(\frac{2\mathcal{N}_{\alpha,T} - 2}{k}\right)^k \leq \exp\{(2/e)(\mathcal{N}_{\alpha,T} - 1)\}\ ,$$

where the last inequality holds since the function $(a/x)^x$ is maximized at $x = a/e$ for every $a > 0$. We may thus set $\Pi' = \exp\{(2/e)(\mathcal{N}'_\alpha - 1)\}$. Theorem 4.3 now yields a bound on the regret w.r.t. the best revealed expert. We still need to add to this bound the quantity $\beta + 1$, which bounds the difference between the cumulative losses of the best revealed expert and the best overall expert. We obtain the following result for the case of α-similar subsets.

Theorem 4.7. *In the clustered experts scenario, the regret of* Hed_C *run with parameters* $\eta = \ln\Big(1+\sqrt{(4/e)(\mathcal{N}'_\alpha - 1)/\big(L' + (2\mathcal{N}'_\alpha - 1)(\mathcal{N}'_\alpha - 1)\alpha\big)}\Big)$ *and* $\mathcal{L}' = L' + (2\mathcal{N}'_\alpha - 1)(\mathcal{N}'_\alpha - 1)\alpha$ *satisfies* $R_{Hed_C,T} = O\left(\mathcal{N}'_\alpha(1 + (2\mathcal{N}'_\alpha - 1)\alpha) + \sqrt{L'(\mathcal{N}'_\alpha - 1)}\right)$.

If both $\mathcal{N}'_\alpha$ and L' are unknown, a doubling trick once again may be used.

Corollary 4.8. *Applying a doubling trick to* Hed_C *in the clustered experts setting yields a regret bound of* $O\left(\mathcal{N}_{\alpha,T}(1 + \alpha\mathcal{N}_{\alpha,T})(1 + \ln L^*_T) + \sqrt{L^*_T \mathcal{N}_{\alpha,T}}\right)$.

Remark 4.9. *If losses are stochastic (or contain a stochastic element), the diameter of a set of experts grows gradually over time, rather than remaining constant. Fix a time horizon* T *and consider* N *experts with i.i.d. Bernoulli random losses. For* $\delta \in (0, 1)$, *the diameter of this set is* $O(\sqrt{T \ln(N/\delta)})$ *with probability at least* $1 - \delta$. *This is shown by combining a "maximal" concentration inequality with the union bound. Picking* $\alpha = \Theta(\sqrt{T \ln(N/\delta)})$ *for this case thus yields a single cluster and* $O(\sqrt{T \ln(N/\delta)})$ *regret. A similar argument applies to the few leading experts scenario.*

4.6 Lower Bounds

In this section we prove lower bounds for the branching setup, as well as the few leaders and clustered experts scenarios of Section 4.5. We show that the key term in the regret bound of Hed_C for the branching setting, $\sqrt{L^*_T \ln \Pi}$ (see Corollary 4.4), may not be improved in general. The same holds for the corresponding terms $\sqrt{L^*_T \Lambda_{\alpha,T}}$ and $\sqrt{L^*_T \mathcal{N}_{\alpha,T}}$ for the other two scenarios (Corollaries 4.6 and 4.8, respectively) if the number of leaders or similar subsets is at most logarithmic in the number of experts. This condition is clearly necessary, since otherwise Hedge itself guarantees better regret than Hed_C.

We use a single construction for all the above scenarios. It involves an oblivious stochastic adversary whose branching tree is a highly unbalanced comb-shaped tree, that is, with splits occurring only in a single branch. This construction and accompanying lemma are geared towards the case of subsets of identical experts, but are useful for the other scenarios as well. The construction proceeds as follows. Given N for the number of experts and K for the number of unique experts, we define sets $S_1 \supset S_2 \supset \cdots \supset S_K$ of experts, where $S_1 = \{1, \ldots, N\}$, and S_{i+1} is a random half of S_i, which contains the best expert. The K distinct subsets are $S_j \setminus S_{j+1}$, for $j = 1, \ldots, K$, where we define $S_{K+1} = \emptyset$. Just as in proofs for the standard best expert setting (see, e.g., Theorems 4.7 and 4.8 in Chapter 4 of [72]), this construction prevents any learner from doing better than random. We comment, however, that additional care is required to control the number of distinct experts. We make use of the following lemma.

Lemma 4.10. *Let $l_{i,t} \in \{0,1\}$ for all i and t, where exactly K loss sequences $(l_{i,1}, \ldots, l_{i,T})$ are distinct. Even if K is known to the learner, it holds that*

(i) For every $T \leq \lfloor \log_2 N \rfloor$, there is an oblivious stochastic adversary that generates N loss sequences of length T of which $K = T+1$ are unique, such that the expected regret of any algorithm A satisfies $\mathbb{E}[R_{A,T}] \geq T/2$.

(ii) For every $T \geq 1 + \lfloor \log_2 N \rfloor$ and $K \leq 1 + \lfloor \log_2 N \rfloor$ there is an oblivious stochastic adversary that generates N loss sequences of length T, of which K are unique, such that the expected regret of any algorithm A satisfies $\mathbb{E}[R_{A,T}] \geq \frac{1}{4}\sqrt{\lfloor T/(K-1) \rfloor}(K-1) = \Omega\big(\sqrt{T(K-1)}\big)$.

Proof. (i) Consider the following loss sequence generation. At each time t, the adversary focuses only on the set S_t of experts, where $S_t = \{1 \leq i \leq 2^{\lfloor \log_2 N \rfloor} : \forall \tau < t, l_{i,\tau} = 0\}$ (initially, S_1 includes all the experts $1 \leq i \leq 2^{\lfloor \log_2 N \rfloor}$). The set S_t is then divided randomly into two equal-sized sets, where the first set is given loss 1 at time t and the other loss 0. All other experts are given loss 1. Even if the learner has the benefit of knowing S_t before deciding on $\mathbf{p}_t$, its expected loss at time t is 1/2. (The learner only stands to lose if it puts weight outside S_t.) Therefore, in any case, $\mathbb{E}[L_{A,T}] \geq T/2$, while $L_T^* = 0$, implying that $\mathbb{E}[R_{A,T}] \geq T/2$. There are exactly $T+1$ distinct experts in the above construction.

(ii) Consider the following construction of loss sequences by an adversary. Let $\tau = \lfloor T/(K-1) \rfloor$ and divide the time range $1, \ldots, \tau(K-1)$ into $K-1$ time slices of size τ. Times $(\tau(K-1), T]$ may be ignored, since the adversary may assign $l_{i,t} = 0$ for every $1 \leq i \leq N$ and $t \in (\tau(K-1), T]$, so the regret of any algorithm is unaffected. The adversary defines sets $S_1 \supset S_2 \supset \cdots \supset S_K$ of experts, where initially $S_1 = \{1, \ldots, N\}$. For time slice j, all experts not in S_j incur τ times the loss 1. Denote S_j^1 for the first $\lfloor |S_j|/2 \rfloor$ experts in S_j and $S_j^2 = S_j \setminus S_j^1$. The adversary generates two $\{0,1\}$-valued loss sequences of size τ by making 2τ i.i.d. draws from the uniform distribution on $\{0,1\}$. Experts in S_j^1 incur the losses in the first sequence, and experts in S_j^2 incur the losses in the second sequence. If the sequences are identical, the adversary modifies its choice by picking instead the sequences $\{0, \ldots, 0, 1\}$ and $\{0, \ldots, 0, 0\}$ and assigning one of them randomly to the experts in S_j^1 and the other to experts in S_j^2. The set S_{j+1} is then defined as the set among S_j^1 and S_j^2 with the smallest cumulative loss, or S_j^1 in case of a tie. Note that we end up with exactly K distinct loss sequences, and S_j always contains an expert with the smallest cumulative loss at any time.

We may assume w.l.o.g. that in time slice j, the algorithm puts weight only on S_j, and we denote R^j for its regret on the j-th time slice. We also denote $\widehat{R}^j$ for the regret if we had not made the modification for identical sequences. By Lemma 4.14 (see Section 4.8), we have $\mathbb{E}[\widehat{R}^j] \geq \sqrt{\tau}/4$, since the expected loss of the algorithm on the time slice is $\tau/2$. Denote B_j for the event that the sequences in slice j are identical and $\bar{B}_j$ for its complement. We have $\sqrt{\tau}/4 \leq \mathbb{E}[\widehat{R}^j] = \mathbb{E}[\widehat{R}^j \mid \bar{B}_j]\mathbb{P}(\bar{B}_j) + \mathbb{E}[\widehat{R}^j \mid B_j]\mathbb{P}(B_j) = \mathbb{E}[\widehat{R}^j \mid \bar{B}_j]\mathbb{P}(\bar{B}_j)$, since the regret is 0 if the sequences are identical. In addition, we have

$$\begin{aligned}
\mathbb{E}[R^j] &= \mathbb{E}[R^j \mid \bar{B}_j]\mathbb{P}(\bar{B}_j) + \mathbb{E}[R^j \mid B_j]\mathbb{P}(B_j) \\
&= \mathbb{E}[\widehat{R}^j \mid \bar{B}_j]\mathbb{P}(\bar{B}_j) + \mathbb{E}[R^j \mid B_j]\mathbb{P}(B_j) \\
&\geq \mathbb{E}[\widehat{R}^j \mid \bar{B}_j]\mathbb{P}(\bar{B}_j) \geq \sqrt{\tau}/4 \ ,
\end{aligned}$$

where the first inequality is true because when B_j occurs, $R^j \geq 0$ (one sequence is all zeros). The set S_j always contains an expert with minimal cumulative loss, so

$$\mathbb{E}[R_{A,T}] = \sum_{j=1}^{K-1} \mathbb{E}[R^j] \geq \frac{1}{4}\sqrt{\tau}(K-1) = \frac{1}{4}\sqrt{\lfloor T/(K-1) \rfloor}(K-1) = \Omega\big(\sqrt{T(K-1)}\big) \ .$$

□

With the above lemma handy, we start by considering the branching setup. Given T and $K < T$, the adversary may generate $N = 2^{T-1}$ loss sequences of length T according to part (ii) of Lemma 4.10. At time t it maintains sets of experts with identical histories up until time t, according to the stochastic construction. These sets are the leaves of its branching tree, which is a comb-shaped tree with $K-1$ binary splits along a single splitting branch, so we have $\Pi = 2^{K-1}$. The adversary will be extra helpful and reveal N to the learner in advance, and also reveal at each time t the current composition of the sets in the leaves. Even so, by Lemma 4.10, the regret of any algorithm A satisfies $\mathbb{E}[R_{A,T}] = \Omega\big(\sqrt{T(K-1)}\big) = \Omega\big(\sqrt{L_T^* \ln \Pi}\big)$, so the key term $\sqrt{L_T^* \ln \Pi}$ in the regret bound for Hed_C (Corollary 4.4) may not be improved in general.

For α-similar subsets, Lemma 4.10 clearly gives an $\Omega\big(\sqrt{L_T^*\left(\mathcal{N}_{0,T}-1\right)}\big)$ expected regret bound in the $\alpha = 0$ case, if $\mathcal{N}_{0,T} \le 1 + \lfloor \log_2 N \rfloor \le T$. Since the adversary is oblivious, we also have $\mathcal{N}_{0,T} \ge \mathcal{N}_{\alpha,T}$ and therefore an $\Omega\big(\sqrt{L_T^*\left(\mathcal{N}_{\alpha,T}-1\right)}\big)$ bound.

Finally, we may show that for the few leaders scenario, if $\Lambda_{0,T} \le 1+\lfloor \log_2 N \rfloor \le cT$, for some $c > 0$, then the expected regret is $\Omega\big(\sqrt{L_T^*\left(\Lambda_{0,T}-1\right)}\big)$. The application of Lemma 4.10 for this case requires an additional technical step. A closer examination of our stochastic construction reveals that if a run with $K \le \log_2 N$ is stopped at time $T/2$, then the expected regret is still $\Omega\big(\sqrt{L_T^*\left(K-1\right)}\big)$, the number of leaders is in $\{1,\ldots,K\}$, and the set of best experts is of size $\Theta\big(\sqrt{N}\big)$. We may then artificially raise the number of leaders to $K \le \log_2 N = O\big(\sqrt{N}\big)$ by sequentially giving ϵ loss to one member of the best expert set, and a round later to all the others. The remaining rounds may be filled with zero losses for all experts. Thus, for any $\Lambda_{0,T} \le \log_2 N$, if we run this modified procedure with $K = \Lambda_{0,T}$, we achieve expected regret of $\Omega\big(\sqrt{L_T^*\left(\Lambda_{0,T}-1\right)}\big)$ and the number of leaders is exactly $\Lambda_{0,T}$. As before, this implies an $\Omega\big(\sqrt{L_T^*\left(\Lambda_{\alpha,T}-1\right)}\big)$ as well. The following theorem summarizes our lower bounds.

Theorem 4.11. *Let T be a known time horizon.*

(i) For the branching setting and for any $d < T$, there is a random tree generation with $d_T = d$ such that the expected regret of any algorithm A satisfies $\mathbb{E}[R_{A,T}] = \Omega\big(\sqrt{T\ln\Pi}\big)$.

(ii) For any $\mathcal{N}_{0,T} \le 1+\lfloor \log_2 N \rfloor \le T$, there is a random construction of N experts of which $\mathcal{N}_{0,T}$ are distinct, such that the expected regret of any algorithm A satisfies $\mathbb{E}[R_{A,T}] = \Omega\big(\sqrt{T(\mathcal{N}_{0,T}-1)}\big)$.

(iii) For any $\Lambda_{0,T} \le 1+\lfloor \log_2 N \rfloor \le cT$, for some $c > 0$, there is a random construction of N experts of which $\Lambda_{0,T}$ are leaders, such that the expected regret of any algorithm A satisfies $\mathbb{E}[R_{A,T}] = \Omega\big(\sqrt{T(\Lambda_{0,T}-1)}\big)$.

4.7 Branching Experts for the Multi-Armed Bandit Setting

In this section we introduce and analyze a variant of the randomized multi-armed bandit algorithm Exp3 of [8] for the branching setting. For the sake of simplicity, we focus on the case of perfect cloning. This means that new actions $j \in C(i, t+1)$ all start off with the same cumulative loss $L_{i,t}$ as their parent i. This variant, called Exp3.C, is described below here.

Branching Exp3 (Exp3.C)

Parameters: A sequence $\eta_1, \eta_2, \dots$ of real-valued functions satisfying the assumptions of Theorem 4.12.

For each round $t = 1, 2, \dots$

1. For each action $i = 1, \dots, N_{t-1}$, after the adversary reveals the set $C(i,t)$:
   If $t = 1$, then let $\widetilde{L}_{j,0} = 0$ for every $j = 1, \dots, N_1$;
   else, if $t > 1$, then $\widetilde{L}_{j,t-1} = \widetilde{L}_{i,t-2} + \widetilde{l}_{i,t-1} + \frac{1}{\eta_{t-1}} \ln |C(i,t)|$ for every $j \in C(i,t)$, including i.
2. Compute the new distribution over actions $\mathbf{p}_t = \left(p_{1,t}, \dots, p_{N_t,t}\right)$, where
$$p_{i,t} = \frac{\exp\left(-\eta_t \widetilde{L}_{i,t-1}\right)}{\sum_{k=1}^{N_t} \exp\left(-\eta_t \widetilde{L}_{k,t-1}\right)} .$$
3. Draw an action I_t from the probability distribution $\mathbf{p}_t$ and observe loss $l_{I_t,t}$.
4. For each action $i = 1, \dots, N_t$ compute the estimated loss $\widetilde{l}_{i,t} = \dfrac{l_{i,t}}{p_{i,t}} \mathbb{I}\{I_t = i\}$.

The main modification with respect to Exp3 is in the way cumulative loss estimates $\widetilde{L}_{i,t}$ are computed (step 1 in the pseudo-code). The additional term $\frac{1}{\eta_{t-1}} \ln |C(i,t)|$ in these estimates serves a role similar to that of the partial restart in the full information case. There, we divided the weight of a parent expert i among children $C(i,t)$. Here, we increase the loss estimate of $j \in C(i,t)$ to achieve the same effect.

The next theorem bounds the expected regret of Exp3.C against an oblivious adversary.[5] This is defined as $\mathbb{E}[R_{Exp3.C,T}] = \mathbb{E}\big[L_{Exp3.C,T}\big] - L_T^*$, where $L_{Exp3.C,T} = l_{I_1,1} + \cdots + l_{I_T,T}$ is the random variable denoting Exp3.C's total loss with respect to the sequence $I_1, \dots, I_T$ of random draws.

Theorem 4.12. *Let $\eta_1, \eta_2, \dots$ be a sequence of functions $\eta_t : \mathbb{N} \to \mathbb{R}^+$ such that for every $k_1 \le k_2 \le \dots$, it holds that $\eta_1(k_1) \ge \eta_2(k_2) \ge \dots$ (in what follows, we write $\eta_t = \eta_t(N_t)$ for short). If $i_0, \dots, i_T = m(T)$ are the actions on the path from the root*

[5] Extensions to non-oblivious adversaries are possible, with some assumptions on the adversary's control on the quantities $C(i,t)$.

to the best action $m(T)$, then

$$\mathbb{E}[R_{Exp3.C,T}] \leq \frac{1}{2}\sum_{t=1}^{T} N_t\eta_t + \sum_{t=1}^{T}\frac{1}{\eta_t}\ln\frac{N_t|C(i_t,t+1)|}{N_{t+1}} + \frac{\ln N_{T+1}}{\eta_T}\,. \tag{4.2}$$

If Exp3.C is run with $\eta_t(k) = \sqrt{\frac{\ln ek}{tk}}$, then

$$\mathbb{E}[R_{Exp3.C,T}] \leq 2\sqrt{TN_T\ln eN_T}\left(1 + \frac{\ln\prod_{t=1}^{T}|C(i_t,t+1)|}{2\ln eN_T}\right)\,. \tag{4.3}$$

Proof. Note first that if splits always occur uniformly for all actions i (i.e, for all t we have that $|C(i,t+1)|$ is the same for all $i = 1,\ldots,N_t$), then $N_{t+1} = N_t|C(i,t+1)|$ implying $\ln\bigl(N_t|C(i,t+1)|/N_{t+1}\bigr) = 0$. Hence we would get

$$\mathbb{E}[R_{Exp3.C,T}] \leq \frac{1}{2}\sum_{t=1}^{T} N_t\eta_t + \frac{\ln N_{T+1}}{\eta_T}\,.$$

In particular, for the standard bandit setting, where $N_1 = \cdots = N_{T+1} = N$, we get

$$\mathbb{E}[R_{Exp3.C,T}] \leq \frac{N}{2}\sum_{t=1}^{T}\eta_t + \frac{\ln N}{\eta_T}$$

and recover the original result.

To prove (4.3) from (4.2), we first note that $\ln\frac{N_t}{N_{t+1}} \leq 0$ for all t. In addition, without loss of generality, $N_T = N_{T+1}$ (otherwise we add an artificial round). We obtain

$$\begin{aligned}
\mathbb{E}[R_{Exp3.C,T}] &\leq \frac{1}{2}\sum_{t=1}^{T}\sqrt{\frac{N_t\ln eN_t}{t}} + \frac{1}{\eta_T}\sum_{t=1}^{T}\ln|C(i_t,t+1)| + \sqrt{\frac{TN_T(\ln N_T)^2}{\ln eN_T}} \\
&\leq \frac{1}{2}\sqrt{N_T\ln eN_T}\sum_{t=1}^{T}\frac{1}{\sqrt{t}} + \sqrt{\frac{TN_T}{\ln eN_T}}\ln\prod_{t=1}^{T}|C(i_t,t+1)| + \sqrt{TN_T\ln eN_T} \\
&\leq 2\sqrt{TN_T\ln eN_T} + \sqrt{\frac{TN_T}{\ln eN_T}}\ln\prod_{t=1}^{T}|C(i_t,t+1)| \\
&= 2\sqrt{TN_T\ln eN_T}\left(1 + \frac{\ln\prod_{t=1}^{T}|C(i_t,t+1)|}{2\ln eN_T}\right)\,,
\end{aligned}$$

where we used the fact that $\sum_{t=1}^{T}\frac{1}{\sqrt{t}} \leq \int_0^T\frac{1}{\sqrt{t}}dt = 2\sqrt{T}$.

The proof of (4.2) is an adaptation of the proof of [21, Theorem 3.1], which is divided into five steps. Here we just focus on the main differences. In the following, we write $\mathbb{E}_{i\sim p_t}$ to denote the expectation w.r.t. the random draw of i from the distribution p_t specified by the probability vector $\mathbf{p}_t = \big(p_{1,t},\dots,p_{N_t,t}\big)$. Moreover, given any action $k \in N_T$, we use k also to index any action i on the path from the root to k. This is OK because, since we have perfect cloning, we have that $L_{k,T} = l_{i_1,1} + \cdots + l_{i_T,T}$, where $i_1,\dots,i_T$ are the actions on the path from the root i_0 to $i_T = k$.

The first two steps of the proof are identical to the proof in [21]:

$$\sum_{t=1}^{T} l_{I_t,t} - \sum_{t=1}^{T} l_{k,t} = \sum_{t=1}^{T} \mathbb{E}_{i\sim p_t}\widetilde{l}_{i,t} - \sum_{t=1}^{T} \mathbb{E}_{I_t\sim p_t}\widetilde{l}_{k,t}\ . \tag{4.4}$$

Now we rewrite $\mathbb{E}_{i\sim p_t}\widetilde{l}_{i,t}$ as follows

$$\mathbb{E}_{i\sim p_t}\widetilde{l}_{i,t} = \frac{1}{\eta_t}\ln \mathbb{E}_{i\sim p_t}\exp\Big(-\eta_t\big(\widetilde{l}_{i,t} - \mathbb{E}_{k\sim p_t}\widetilde{l}_{k,t}\big)\Big) - \frac{1}{\eta_t}\ln \mathbb{E}_{i\sim p_t}\exp\Big(-\eta_t\widetilde{l}_{i,t}\Big)\ . \tag{4.5}$$

Following the second step in the proof in [21] we obtain

$$\ln \mathbb{E}_{i\sim p_t}\exp\Big(-\eta_t(\widetilde{l}_{i,t} - \mathbb{E}_{k\sim p_t}\widetilde{l}_{k,t})\Big) \le \frac{\eta_t^2}{2p_{I_t,t}}\ . \tag{4.6}$$

Next, we study the second term in (4.5). This relies on the specific properties of Exp3.C. Let $\Phi_0(\eta) = 0$ and $\Phi_t(\eta) = \frac{1}{\eta}\ln\frac{1}{N_{t+1}}\sum_{i=1}^{N_{t+1}}\exp\Big(-\eta\widetilde{L}_{i,t}\Big)$. By definition of p_t, and recalling that $\widetilde{L}_{j,t} = \widetilde{L}_{i,t}$ for every $j \in C(i,t+1)$ and every $i = 1,\dots,N_t$, we have

$$\begin{aligned}
-\frac{1}{\eta_t}\ln \mathbb{E}_{i\sim p_t}\exp\Big(-\eta_t\widetilde{l}_{i,t}\Big) &= -\frac{1}{\eta_t}\ln\frac{\sum_{i=1}^{N_t}\exp\Big(-\eta_t\widetilde{L}_{i,t-1}\Big)\exp\Big(-\eta_t\widetilde{l}_{i,t}\Big)}{\sum_{i=1}^{N_t}\exp\Big(-\eta_t\widetilde{L}_{i,t-1}\Big)} \\
&= -\frac{1}{\eta_t}\ln\frac{\sum_{i=1}^{N_t}|C(i,t+1)|\exp\Big(-\eta_t\widetilde{L}_{i,t}\Big)}{\sum_{i=1}^{N_t}\exp\Big(-\eta_t\widetilde{L}_{i,t-1}\Big)} \\
&= -\frac{1}{\eta_t}\ln\frac{\sum_{i=1}^{N_{t+1}}\exp\Big(-\eta_t\widetilde{L}_{i,t}\Big)}{\sum_{i=1}^{N_t}\exp\Big(-\eta_t\widetilde{L}_{i,t-1}\Big)} \\
&= \Phi_{t-1}(\eta_t) - \Phi_t(\eta_t) + \frac{1}{\eta_t}\ln\frac{N_t}{N_{t+1}}\ .
\end{aligned} \tag{4.7}$$

Putting together (4.4), (4.5), (4.6) and (4.7) we obtain

$$\sum_{t=1}^{T} l_{I_t,t} - \sum_{t=1}^{T} l_{k,t} \le \sum_{t=1}^{T}\frac{\eta_t}{2p_{I_t,t}} + \sum_{t=1}^{T}\Big(\Phi_{t-1}(\eta_t) - \Phi_t(\eta_t)\Big) + \sum_{t=1}^{T}\frac{1}{\eta_t}\ln\frac{N_t}{N_{t+1}} - \sum_{t=1}^{T}\mathbb{E}_{I_t\sim p_t}\widetilde{l}_{k,t}\ .$$

The first term is easy to bound in expectation since by the rule of conditional expectations we have

$$\mathbb{E}\sum_{t=1}^{T}\frac{\eta_t}{2p_{I_t,t}}=\mathbb{E}\sum_{t=1}^{T}\mathbb{E}_{I_t\sim p_t}\frac{\eta_t}{2p_{I_t,t}}=\frac{1}{2}\sum_{t=1}^{T}N_t\eta_t\,.$$

For the second term we again proceed similarly to [21],

$$\sum_{t=1}^{T}\big(\Phi_{t-1}(\eta_t)-\Phi_t(\eta_t)\big)=\sum_{t=1}^{T-1}\big(\Phi_t(\eta_{t+1})-\Phi_t(\eta_t)\big)-\Phi_T(\eta_T)$$

since $\Phi_0(\eta_1)=0$. Note that

$$\begin{aligned}
-\Phi_T(\eta_T)&=\frac{\ln N_{T+1}}{\eta_T}-\frac{1}{\eta_T}\ln\left(\sum_{i=1}^{N_{T+1}}\exp\left(-\eta_T\widetilde{L}_{i,T}\right)\right)\\
&\le\frac{\ln N_{T+1}}{\eta_T}-\frac{1}{\eta_T}\ln\left(\exp\left(-\eta_T\widetilde{L}_{k,T}\right)\right)\\
&=\frac{\ln N_{T+1}}{\eta_T}+\sum_{t=1}^{T}\left(\widetilde{l}_{k,t}+\frac{1}{\eta_t}\ln|C(k,t+1)|\right)\,,
\end{aligned}$$

and thus we have

$$\begin{aligned}
\mathbb{E}\left[\sum_{t=1}^{T}l_{I_t,t}-\sum_{t=1}^{T}l_{k,t}\right]&\le\frac{1}{2}\sum_{t=1}^{T}N_t\eta_t+\mathbb{E}\sum_{t=1}^{T-1}\big(\Phi_t(\eta_{t+1})-\Phi_t(\eta_t)\big)\\
&\quad+\sum_{t=1}^{T}\frac{1}{\eta_t}\ln\frac{N_t|C(k,t+1)|}{N_{t+1}}+\frac{\ln N_{T+1}}{\eta_T}\,.
\end{aligned}$$

The proof is concluded by showing that $\Phi_t'(\eta)\ge 0$. Since $\eta_{t+1}\le\eta_t$, this would give $\Phi_t(\eta_{t+1})-\Phi_t(\eta_t)\le 0$. In fact, the proof of this claim goes along the same lines as the proof in [21], and is therefore omitted. □

4.8 Appendix: Additional Claims

Claim 4.13. *There exists a scenario with $N+1$ experts s.t. for a large enough N,* Hed$_C$ *has regret $\Omega\big(\sqrt{L_T^*N}\big)$ while Hedge has regret $O\big(\ln N+\sqrt{L_T^*\ln N}\big)$.*

Proof. Consider the following special scenario. The adversary initially reveals $N+1$ experts, one in every round, without any losses, in a comb-shaped branching tree. The adversary next gives losses 0 to the last revealed expert and 1 to the others for t rounds, and afterwards gives loss 1 to all experts for τ rounds. Suppose further that N, t, and τ are known in advance.

We now consider the regret of Hed_C. We have $L_T^* = \tau$, so $\eta = \ln\bigl(1+\sqrt{(2\ln 2)N/\tau}\bigr)$, and it is easily seen that the regret is lower bounded by $\bigl(1 - p_{N+1,t+N}\bigr)t$, where

$$p_{N+1,t+N} = \frac{2^{-N}}{2^{-N} + (1-2^{-N})e^{-\eta t}} = \frac{1}{1+(2^N-1)e^{-\eta t}} \leq \frac{e^{\eta t}}{2^N} = \exp(\eta t - N\ln 2) \ .$$

Taking $t = \bigl\lfloor (N\ln 2 - 1)/\eta \bigr\rfloor$ ensures the regret is lower bounded by $t/2$. Since

$$\frac{N\ln 2 - 1}{\eta} \geq \frac{N\ln 2 - 1}{\sqrt{(2\ln 2)N/\tau}} \geq \frac{(N/2)\ln 2}{\sqrt{(2\ln 2)N/\tau}} \geq \sqrt{N\tau(\ln 2)/8} \ ,$$

the result follows. □

Lemma 4.14. *If $X = Z_1 + \ldots + Z_n$ and $Y = Z_{n+1} + \ldots + Z_{2n}$, where Z_i are independent Bernoulli variables with $p = \frac{1}{2}$, then $\mathbb{E}[\min\{X,Y\}] \leq \frac{1}{2}n - \frac{1}{4}\sqrt{n}$.*

Proof. We have that

$$\begin{aligned}
\mathbb{E}[\min\{X,Y\}] &= \mathbb{E}\left[\frac{1}{2}(X + Y - |X-Y|)\right] \\
&= \frac{1}{2}\mathbb{E}[X] + \frac{1}{2}\mathbb{E}[Y] - \frac{1}{2}\mathbb{E}[|X-Y|] \\
&= \frac{n}{2} - \frac{1}{2}\mathbb{E}[|X-Y|] \ ,
\end{aligned}$$

so we need only determine the value of $\mathbb{E}[|X-Y|]$. Now, $\sigma_i = 2Z_i - 1$, for $1 \leq i \leq 2n$, are independent Rademacher variables, and $\frac{1}{2}(\sigma_j - \sigma_{j+n}) = Z_j - Z_{j+n}$, for $1 \leq j \leq n$. Thus, $|X - Y| = |\sum_{j=1}^{n}(Z_j - Z_{j+n})| = \frac{1}{2}|\sum_{i=1}^{2n} a_i\sigma_i|$, where $|a_i| = 1$ for every i. By Khinchine's inequality (see, e.g., [23], Lemma A.9),

$$\mathbb{E}\left[\left|\sum_{i=1}^{2n} a_i\sigma_i\right|\right] \geq \frac{1}{\sqrt{2}}\sqrt{\sum_{i=1}^{2n} a_i^2} = \sqrt{n} \ .$$

Thus,

$$\mathbb{E}[\min\{X,Y\}] = \frac{n}{2} - \frac{1}{2}\mathbb{E}[|X-Y|] = \frac{n}{2} - \frac{1}{4}\mathbb{E}\left[\left|\sum_{i=1}^{2n} a_i\sigma_i\right|\right] \leq \frac{1}{2}n - \frac{1}{4}\sqrt{n} \ .$$

□

Part II

Derivative Pricing

Chapter 5

Background and Model

5.1 Introduction

In the second part of this work, we apply the robust methodology of regret minimization to one of the most fundamental questions in finance: derivative pricing.

A *derivative* is a security whose price is dependent on the price of one or more underlying assets, for example, stocks. An important type of derivative is an *option*, which is a financial instrument that allows its holder to buy or sell a certain asset for a given price at a given time. For example, a *European call option*, at its expiration time T, allows its holder to buy the asset for a price K, called the *strike price*. In other words, the option pays $\max\{S_T - K, 0\}$ at time T, where S_T is the asset (stock) price at time T. Another standard option is the *European put option*, which allows its holder to *sell* the asset for a price K, or, equivalently, receive a payoff of $\max\{K - S_T, 0\}$ at time T. Buying such options enables stock holders to hedge their investments against future rises or falls in the asset price. These options are also widely traded as securities in their own right.

Apart from the standard call and put options, there are numerous other option types that are traded or developed to meet particular financial needs. Such options are collectively known as *exotic options*. One example is the *European fixed-strike lookback call option*, which, at its expiration time T, allows its holder to choose the best time in hindsight to buy the asset for a price K. Namely, the lookback option pays $\max\{M_T - K, 0\}$ at time T, where M_T is the maximal asset price over the lifetime of the option.

Derivative pricing has been greatly influenced, both theoretically and practically, by the pricing formula and model of Black and Scholes [11] and Merton [68]. In their Nobel Prize-winning works, they modeled the price of a stock as a geometric Brownian motion stochastic process, and assumed an arbitrage-free market, namely, that market prices provide no opportunities for riskless profit.

The idea behind derivative pricing in the Black-Scholes-Merton (BSM) model is as follows. Consider, for example, the pricing of a European call option. At every time, there exists an amount of stock that together with a short position in one option creates a risk-free portfolio. In an arbitrage-free market, the instantaneous return of this portfolio is determined by the prevailing risk-free interest rate. These observations may be used to derive a partial differential equation (the BSM equation) that together with boundary conditions suited for the specific payoff of the European call option may be solved to obtain an exact price. (For an exposition of pricing in the BSM model see, e.g., [57].)

The assumptions of the BSM model, however, have several known drawbacks. First, the model is only an abstraction of price changes, while in reality prices are discrete and experience sharp jumps, and the daily returns are neither independent nor identically distributed. Second, the main parameter that is required, the stock volatility, is not observable, and has to be estimated. In fact, calculating an *implied volatility* for the BSM model based on actual market prices of European call options leads to different values for different strike prices, even for a given expiration time.

These empirical discrepancies motivated the introduction of an adversarial online learning model for derivative pricing in the work of [35]. In their model, which we use in this work as well, trading takes place in discrete time, and the market is assumed to be adversarial. The only restrictions on a stock price consist of a bound on the sum of the squared single-period returns of the stock (quadratic variation) and a bound on its absolute single-period returns. An optional bound considered in [35], and which we adopt, is an upper bound on the stock price. This model, which will be described in Section 5.2, clearly accommodates the price jumps and dependence observed in real markets.

The authors of [35] showed that in an arbitrage-free market, regret minimization algorithms may imply an upper bound on the price of the European call option. For this purpose they specially modified an algorithm for the best expert setting, Polynomial Weights [25], to work in a trading setting. Their price bounds feature the quadratic variation of the stock price, and do not depend on the number of trading periods. This property is highly desirable, since it allows the trading frequency to be increased without an explicit effect on the price.

In the following chapters, we present results that significantly widen the scope of regret minimization-based pricing. In Chapter 6, we employ the regret minimization component used in [35] to obtain price bounds in terms of the quadratic variation for a variety of exotic derivatives. In Chapter 7, we combine this component with one-way trading algorithmic components to obtain price bounds for fixed-strike lookback options in particular. In Chapter 8, we obtain price bounds in terms of variation using other regret minimization algorithmic components, by directly translating regret bounds obtained for the best expert setting into a trading setting. The results of Chapter 3, in particular, are shown to imply lower bounds on the price of certain call options in terms of variation. These are the first such adversarial regret-based lower bounds that are asymptotically optimal.

5.2 The Model

We consider a discrete-time finite-horizon model, with a risky asset (stock) and a risk-free asset (bond or cash). The price of the stock at time $t \in \{0, 1, \ldots, T\}$ is S_t and the price of the risk-free asset is B_t. Initially, $B_0 = S_0 = 1$. We assume that the price of cash does not change, i.e., $B_t = 1$, which is equivalent to assuming a zero risk-free interest rate. We further assume that we can buy or sell any real quantity of stocks with no transaction costs. For the stock we denote by r_t the single period return between $t-1$ and t, so $S_t = S_{t-1}(1 + r_t)$.

A realization of the prices is a *price path*, which is the vector $\mathbf{r}_t = (r_1, \ldots, r_t)$. We define a few parameters of a price path. We denote by M an upper bound on stock prices S_t, by R an upper bound on absolute single-period returns $|r_t|$, by Q an upper bound on the quadratic variation $\sum_{t=1}^{T} r_t^2$, and by M_t the maximum price up to time t, $M_t = \max_{0 \le u \le t}\{S_u\}$.[1] We will assume that the bounds R, Q, and M are given, and $\Pi_{M,R,Q}$, or simply Π, will denote the set of all price paths satisfying these bounds.

[1]The quadratic variation defined here is different from the quadratic variation used in Part I in two ways. First, r_t is not an additive concept of loss (or gain), but rather a multiplicative one. Second, the quadratic variation here is defined for a sequence of scalars and is a property of an individual asset, whereas in Part I it is defined for a sequence of loss vectors and is the property of a collection. In Chapter 8 we revisit the additive setting of Part I and show that it is the *relative* quadratic variation that is more relevant in a financial setting.

We note that stock prices are always upper bounded by $\exp(\sqrt{QT})$, as can be easily verified (see Lemma 7.16 in Section 7.6), and thus we may assume $M \leq \exp(\sqrt{QT})$. Since $\max_{1 \leq t \leq T}\{r_t^2\} \leq \sum_{t=1}^{T} r_t^2 < R^2 T$, we may assume that $R^2 \leq Q \leq R^2 T$ (note that if $Q < R^2$ we may set $R = \sqrt{Q}$). The number of time steps, T, is influenced both by the frequency of trading and the absolute time duration. For this reason it is instructive to consider M and Q as fixed, rather than as increasing in T.

A *trading algorithm* A starts with a total asset value V_0. At every time period $t \geq 1$, A sets weights $w_{s,t} \geq 0$ for stock and $w_{c,t} \geq 0$ for cash, and we define the fractions $p_{s,t} = w_{s,t}/W_t$ and $p_{c,t} = w_{c,t}/W_t$, where $W_t = w_{s,t} + w_{c,t}$. A fraction $p_{s,t}$ of the total asset value, V_{t-1}, is placed in stock, and likewise, $p_{c,t}V_{t-1}$ is placed in cash. Following that, the stock price S_t becomes known, the asset value is updated to $V_t = V_{t-1}(1 + p_{s,t}r_t)$, and time period $t+1$ begins.

We comment that since we assume that both weights are non-negative, the algorithms we consider use neither short selling of the stock, nor buying on margin (negative positions in cash). However, as part of the arbitrage-free assumption, we assume that short selling is, in general, allowed in the market.

An algorithm is referred to as *one-way trading* from stock to cash if the amount of cash never decreases over time, i.e., for every $t \geq 1$ we have $p_{c,t}V_{t-1} \leq p_{c,t+1}V_t$, otherwise it is referred to as *two-way trading*.

It is important to note that the above asset and trading model extends naturally to several risky assets, and this extension will be applied in what follows. However, the risky "assets" we will consider will be algorithms that trade in the same single stock and cash (and might also be thought of as stock derivatives). Thus, such a multiple-asset picture will, in fact, involve only one risky asset.

We next define specific types of options referred to in this work:

C A *European call option* $\mathbf{C}(K,T)$ is a security paying $\max\{S_T - K, 0\}$ at time T, where K is the *strike price* and T is the *expiration time*.[2]

LC A *fixed-strike lookback call option* $\mathbf{LC}(K,T)$ is a security paying $\max\{M_T - K, 0\}$ at time T.

EX An *exchange option* $\mathbf{EX}(\mathbf{X}_1, \mathbf{X}_2, T)$ allows the holder to exchange asset $\mathbf{X}_2$ for asset $\mathbf{X}_1$ at time T, making its payoff $\max\{X_{1,T} - X_{2,T}, 0\}$. A European call

[2] We will sometimes refer to $\mathbf{C}(K,T)$ simply as a *call option*.

option $\mathbf{C}(K,T)$ is a special case of an exchange option, where $\mathbf{X}_1$ is the stock and $\mathbf{X}_2$ is a sum of K in cash.

SH A *shout option* $\mathbf{SH}(K,T)$ allows its holder to "shout" and lock in a minimum value for the payoff at one time $0 \leq \tau \leq T$ during the lifetime of the option. Its payoff at time T is, therefore, $\max\{S_T - K, S_\tau - K, 0\}$. (If the holder does not shout, the payoff is $\max\{S_T - K, 0\}$, identical to the payoff of $\mathbf{C}(K,T)$.)

AP An *average price call option* $\mathbf{AP}(K,T)$ is a type of Asian[3] option that pays $\max\{\bar{S}_T - K, 0\}$ at time T, where $\bar{S}_T$ may be either the arithmetic or the geometric mean of the stock's prices. To distinguish between these two possibilities, we will denote $\mathbf{AP}_A(K,T)$ for the option whose payoff is $\max\{\bar{S}_T^A - K, 0\}$, where $\bar{S}_T^A = \frac{1}{T+1}\sum_{t=0}^{T} S_t$ is the arithmetic mean, and $\mathbf{AP}_G(K,T)$ for the option whose payoff is $\max\{\bar{S}_T^G - K, 0\}$, where $\bar{S}_T^G = (\prod_{t=0}^{T} S_t)^{\frac{1}{T+1}}$ is the geometric mean.

AS An *average strike call option* $\mathbf{AS}(T)$ is a type of Asian option that allows its holder to get the difference between the final stock price and the average stock price, namely, a payoff of $\max\{S_T - \bar{S}_T^A, 0\}$.

We will use bold text for denoting securities and plain text for denoting their values at time 0. For example, we write $SH(K,T)$ for the value of the option $\mathbf{SH}(K,T)$ at time 0.

The above options are a special case of stock derivatives whose value at time T is some function of $\mathbf{r}_T$, the price path of the underlying stock.[4] A European call option is an example of a *path-independent* derivative, since its payoff of $\max\{S_T - K, 0\}$ depends only on the price at time T. Lookback and Asian options are *path-dependent* derivatives, since their payoff depends on the entire price path of the stock.

Most of the options listed above have a payoff of the form $\max\{X_T - K, 0\}$ where $X_T > 0$ is the value of some derivative at time T. We will use the term *call option on a derivative* to refer to all such options, and the notation $\mathbf{C}(\mathbf{X}, K, T)$ for an option that pays $\max\{X_T - K, 0\}$ at time T. Note that if X_T is restricted to some interval $[m_X, M_X]$, then the interesting range for K is $[m_X, M_X)$. The reason is that $K \geq M_X$ implies that the payoff is always 0, so $C(\mathbf{X}, K, T) = 0$, and for $K < m_X$, $C(\mathbf{X}, K, T) = C(\mathbf{X}, m_X, T) + m_X - K$. We may thus assume that for European call options, average

[3]Involving averaging over asset values.
[4]Except for the shout option.

price call options, and shout options, $K \in [0, M)$, and for lookback options, $K \in [S_0, M)$.

5.2.1 Arbitrage-Free Bounds

We assume that the pricing is *arbitrage-free*, which is defined as follows. Trading algorithm A_1 *dominates* trading algorithm A_2 w.r.t. the set of price paths Π, if for every price path in Π, the final value of A_1 is at least the final value of A_2. The *arbitrage-free assumption* says that if A_1 dominates A_2, then the initial value of A_1 is at least the initial value of A_2. This assumption is natural, because if it is violated, it becomes possible to make a riskless profit by buying into A_1 and selling A_2. The resulting flow of funds from A_2 to A_1 affects the stock price in a way that causes even a small arbitrage opportunity to quickly disappear.

For example, define trading algorithm A_{LC}, which simply buys the fixed-strike lookback option and holds it. Its initial value is $LC(K, T)$ and its final value is $\max\{M_T - K, 0\}$. (We would sometimes refer to A_{LC} as simply the option $\mathbf{LC}(K, T)$.) Assume we design a trading strategy A_1 whose initial value is V_0, and its value at time T always satisfies $V_T \geq \max\{M_T - K, 0\}$. This means that A_1 dominates A_{LC}. Therefore, by the arbitrage-free assumption, we have that $LC(K, T) \leq V_0$.

Chapter 6

Pricing Exotic Derivatives

6.1 Pricing Based on Multiplicative Regret

There is a fundamental connection between pricing derivatives and trading algorithms, implied by the arbitrage-free assumption. Given a trading algorithm that dominates the payoff of a derivative, its setup cost is an upper bound on the initial price of the derivative. Similarly, if a trading algorithm is always dominated by the payoff of a derivative, its setup cost is a lower bound on the initial price of the derivative. That said, in this chapter and the next we will consider only upper bounds. We will return to lower bounds in Chapter 8.

Our bounds will be based on the connection, established in [35], between arbitrage-free pricing and the notion of *multiplicative regret*. We now proceed to describe a useful extension of their result.

Consider n derivatives, $\mathbf{X}_1, \ldots, \mathbf{X}_n$. Let $X_{1,t}, \ldots, X_{n,t}$ be their values at time t, and assume $X_{i,T} > 0$ for every $1 \leq i \leq n$. We now consider an option that pays the maximal value of a set of given derivatives.[1] More specifically, we will denote $\mathbf{\Psi}(\mathbf{X}_1, \ldots, \mathbf{X}_n, T)$ for an option that pays $\max\{X_{1,T}, \ldots, X_{n,T}\}$ at time T, and $\Psi(\mathbf{X}_1, \ldots, \mathbf{X}_n, T)$ for its value at time 0.

Definition 6.1. *Let A be a trading algorithm with initial value $V_0 = 1$, and let $\beta_1, \ldots, \beta_n > 0$. The algorithm A is said to have a $(\beta_1, \ldots, \beta_n)$ multiplicative regret w.r.t. derivatives $\mathbf{X}_1, \ldots, \mathbf{X}_n$ if for every price path and every $1 \leq i \leq n$, $V_T \geq \beta_i X_{i,T}$.*

[1] Equivalently, this option is a call on the maximum with a zero strike price.

Lemma 6.2. *If there exists a trading algorithm with a* $(\beta_1, \dots, \beta_n)$ *multiplicative regret w.r.t. derivatives* $\mathbf{X}_1, \dots, \mathbf{X}_n$*, then* $\Psi(\mathbf{X}_1, \dots, \mathbf{X}_n, T) \le 1/\beta$*, where* $\beta = \min_{1\le i\le n}\{\beta_i\}$.

Proof. We have that $V_T \ge \beta \max_{1\le i\le n}\{X_{i,T}\}$, therefore, the payoff of the algorithm dominates β units of the option $\mathbf{\Psi}(\mathbf{X}_1, \dots, \mathbf{X}_n, T)$. By the arbitrage-free assumption, $\Psi(\mathbf{X}_1, \dots, \mathbf{X}_n, T) \le 1/\beta$. □

Moreover, the lemma indicates exactly how improved regret bounds for a trading algorithm relate to tighter upper bounds on $\Psi(\mathbf{X}_1, \dots, \mathbf{X}_n, T)$.

6.2 Price Bounds for a Variety of Options

In order to obtain concrete price bounds, we require a specific trading algorithm whose multiplicative regret bounds we can plug into Lemma 6.2. We use an adaptation of the Polynomial Weights algorithm [25] called *Generic*, which was introduced in [35] and is defined as follows.

The Generic Algorithm

Parameters: A learning rate $\eta > 0$ and initial weights $w_{i,1} > 0$, $1 \le i \le n$.

For each round $t = 1, \dots, T$

1. Define fractions $p_{i,t} = w_{i,t}/W_t$, where $W_t = \sum_{i=1}^n w_{i,t}$.

2. For each asset $i = 1, \dots, n$, let $w_{i,t+1} = w_{i,t}(1 + \eta r_t)$.

It is important to note that in this section we consider derivatives that are *tradable*. Crucially, such derivatives may represent the wealth resulting from investing with some trading algorithm. We will define specific tradable derivatives later for pricing specific options. Let $\mathbf{X}_1, \dots, \mathbf{X}_n$ be such derivatives, where $\mathbf{r}_{i,T} = (r_{i,1}, \dots, r_{i,T})$ is the price path of $\mathbf{X}_i$, for $1 \le i \le n$. We will require that $|r_{i,t}| < R < 1 - 1/\sqrt{2} \approx 0.3$, and $\sum_{t=1}^T r_{i,t}^2 \le Q_i$, for every $1 \le i \le n$. We will assume $X_{i,0} > 0$ for every $1 \le i \le n$, which implies that the derivatives have positive values at all times.

Theorem 6.3. ([35]) *Assume* $X_{1,0} = \dots = X_{n,0} = 1$*. Let* V_T *be the final value of the Generic algorithm investing one unit of cash in* $\mathbf{X}_1, \dots, \mathbf{X}_n$ *with initial fractions* $p_{1,1}, \dots, p_{n,1}$*, and* $\eta \in [1, \frac{1-2R}{2R(1-R)}]$*. Then for every* $1 \le i \le n$*,*

$$V_T \;\ge\; p_{i,1}^{\frac{1}{\eta}} e^{-(\eta-1)Q_i} X_{i,T} \,.$$

In what follows, we will write $\eta_{max} = \frac{1-2R}{2R(1-R)}$ for short. We can now derive a bound on $\Psi(\mathbf{X}_1, \ldots, \mathbf{X}_n, T)$.

Theorem 6.4. *For every $\eta \in [1, \eta_{max}]$,*

$$\Psi(\mathbf{X}_1, \ldots, \mathbf{X}_n, T) \quad \leq \quad \left(\sum_{i=1}^{n} e^{\eta(\eta-1)Q_i} X_{i,0}^{\eta} \right)^{\frac{1}{\eta}} .$$

Proof. For every $1 \leq i \leq n$, define $\mathbf{X}_i' = X_{i,0}^{-1} \mathbf{X}_i$, namely, a fraction of $\mathbf{X}_i$ with value 1 at time 0. Applying Theorem 6.3 to these new assets, we have that for every $1 \leq i \leq n$,

$$V_T \quad \geq \quad p_{i,1}^{\frac{1}{\eta}} e^{-(\eta-1)Q_i} X_{i,T}' = p_{i,1}^{\frac{1}{\eta}} e^{-(\eta-1)Q_i} X_{i,0}^{-1} X_{i,T} \ .$$

Denoting $\beta_i = p_{i,1}^{1/\eta} e^{-(\eta-1)Q_i} X_{i,0}^{-1}$ and $\beta = \min_{1 \leq i \leq n}\{\beta_i\}$, we have by Lemma 6.2 that $\Psi(\mathbf{X}_1, \ldots, \mathbf{X}_n, T) \leq 1/\beta$. For any fixed η, we may optimize this bound by picking a probability vector $(p_{1,1}, \ldots, p_{n,1})$ that maximizes β. Clearly, β is maximized if $\beta_1 = \ldots = \beta_n = c$ for some constant $c > 0$. This is equivalent to having $p_{i,1} = c^{\eta} e^{\eta(\eta-1)Q_i} X_{i,0}^{\eta}$ for every $1 \leq i \leq n$. To ensure that $(p_{1,1}, \ldots, p_{n,1})$ is a probability vector, we must set $c = (\sum_{i=1}^{n} e^{\eta(\eta-1)Q_i} X_{i,0}^{\eta})^{-1/\eta}$. We thus have that $\Psi(\mathbf{X}_1, \ldots, \mathbf{X}_n, T) \leq 1/\beta = 1/c = (\sum_{i=1}^{n} e^{\eta(\eta-1)Q_i} X_{i,0}^{\eta})^{1/\eta}$. □

We next utilize the bound on $\Psi(\mathbf{X}_1, \ldots, \mathbf{X}_n, T)$ to bound the price of various exotic options, as well as the ordinary call option.

Theorem 6.5. *For every $\eta \in [1, \eta_{max}]$, the following bounds hold:*

- $EX(\mathbf{X}_1, \mathbf{X}_2, T) \leq (e^{\eta(\eta-1)Q_1} X_{1,0}^{\eta} + e^{\eta(\eta-1)Q_2} X_{2,0}^{\eta})^{1/\eta} - X_{2,0}$
- $SH(K, T) \leq (K^{\eta} + 2e^{\eta(\eta-1)Q} S_0^{\eta})^{1/\eta} - K$
- $LC(K, T) \leq (K^{\eta} + Te^{\eta(\eta-1)Q} S_0^{\eta})^{1/\eta} - K$
- $AS(T) \leq S_0(e^{(\eta-1)Q+(\ln 2)/\eta} - 1)$

Proof. Throughout this proof, we use the notation $\mathbf{X}_3 = \mathbf{X}_1 + \mathbf{X}_2$ to indicate that the payoff of the derivative $\mathbf{X}_3$ is always equal to the combined payoffs of the derivatives $\mathbf{X}_1$ and $\mathbf{X}_2$. Equivalently, we will write $\mathbf{X}_2 = \mathbf{X}_3 - \mathbf{X}_1$. We point out that by the arbitrage-free assumption, equal payoffs imply equal values at time 0. Therefore, we have that $X_{3,0} = X_{1,0} + X_{2,0}$.

- Since $\mathbf{EX}(\mathbf{X}_1, \mathbf{X}_2, T) = \mathbf{\Psi}(\mathbf{X}_1, \mathbf{X}_2, T) - \mathbf{X}_2$, we have that $EX(\mathbf{X}_1, \mathbf{X}_2, T) = \Psi(\mathbf{X}_1, \mathbf{X}_2, T) - X_{2,0} \leq (\sum_{i=1}^{2} e^{\eta(\eta-1)Q_i} X_{i,0}^{\eta})^{1/\eta} - X_{2,0}$, where the inequality is by Theorem 6.4.

- Let $\mathbf{X}_1$ be the stock. Let $\mathbf{X}_2$ be an algorithm that buys a single stock at time 0, and if the option holder shouts, sells it immediately. In addition, let $\mathbf{X}_3$ be K in cash (implying $Q_3 = 0$). Note that the quadratic variations of both $\mathbf{X}_1$ and $\mathbf{X}_2$ are upper bounded by Q. Since $\mathbf{SH}(K, T) = \mathbf{\Psi}(\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, T) - \mathbf{X}_3$, we have by Theorem 6.4 that $SH(K, T) \leq (K^{\eta} + \sum_{i=1}^{2} e^{\eta(\eta-1)Q_i} X_{i,0}^{\eta})^{1/\eta} - K \leq (K^{\eta} + 2e^{\eta(\eta-1)Q} S_0^{\eta})^{1/\eta} - K$.

- Let $\mathbf{X}_t$ be an algorithm that buys a single stock at time 0 and sells it at time t, $1 \leq t \leq T$. Thus, the quadratic variation of $\mathbf{X}_t$ is upper bounded by Q for every t. In addition, let $\mathbf{X}'$ be K in cash. We have that $\mathbf{LC}(K, T) = \mathbf{\Psi}(\mathbf{X}', \mathbf{X}_1, \ldots, \mathbf{X}_T, T) - \mathbf{X}'$, where we used the fact that $S_0 \leq K$. By Theorem 6.4, $LC(K, T) = \Psi(\mathbf{X}', \mathbf{X}_1, \ldots, \mathbf{X}_T, T) - K \leq (K^{\eta} + Te^{\eta(\eta-1)Q} S_0^{\eta})^{1/\eta} - K$.

- For the bound on $AS(T)$, let $\mathbf{X}_1$ be the stock and let $\mathbf{X}_2$ be an algorithm that buys a single stock at time 0 and sells a fraction $\frac{1}{T+1}$ of the stock at each time $0 \leq t \leq T$. We thus have that $X_{2,T} = \frac{1}{T+1} \sum_{t=0}^{T} S_t$. Denote by Q_2 an upper bound on the quadratic variation of $\mathbf{X}_2$. Since $\mathbf{AS}(T) = \mathbf{EX}(\mathbf{X}_1, \mathbf{X}_2, T)$, then by our bound on $EX(\mathbf{X}_1, \mathbf{X}_2, T)$, we have that $AS(T) \leq (e^{\eta(\eta-1)Q} S_0^{\eta} + e^{\eta(\eta-1)Q_2} S_0^{\eta})^{1/\eta} - S_0 = S_0((e^{\eta(\eta-1)Q} + e^{\eta(\eta-1)Q_2})^{1/\eta} - 1)$. For every t, $X_{2,t} = \frac{1}{T+1} \sum_{\tau=0}^{t} S_{\tau} + \frac{T-t}{T+1} S_t$, therefore,

$$
\begin{aligned}
|r_{2,t}| &= \left| \frac{\frac{1}{T+1} \sum_{\tau=0}^{t} S_{\tau} + \frac{T-t}{T+1} S_t}{\frac{1}{T+1} \sum_{\tau=0}^{t-1} S_{\tau} + \frac{T+1-t}{T+1} S_{t-1}} - 1 \right| = \left| \frac{\frac{1}{T+1} S_t + \frac{T-t}{T+1} S_t - \frac{T+1-t}{T+1} S_{t-1}}{\frac{1}{T+1} \sum_{\tau=0}^{t-1} S_{\tau} + \frac{T+1-t}{T+1} S_{t-1}} \right| \\
&= \frac{\frac{T+1-t}{T+1} |S_t - S_{t-1}|}{\frac{1}{T+1} \sum_{\tau=0}^{t-1} S_{\tau} + \frac{T+1-t}{T+1} S_{t-1}} \leq \frac{\frac{T+1-t}{T+1} |S_t - S_{t-1}|}{\frac{T+1-t}{T+1} S_{t-1}} = |r_t| \,.
\end{aligned}
$$

  Thus, $\sum_{t=1}^{T} r_{2,t}^2 \leq \sum_{t=1}^{T} r_t^2 \leq Q$, and we may assume $Q_2 = Q$. We therefore have that $AS(T) \leq S_0[(2e^{\eta(\eta-1)Q})^{1/\eta} - 1]$, and the result follows.

□

Since an ordinary call is actually $\mathbf{EX}(\mathbf{X}_1, \mathbf{X}_2, T)$, where $\mathbf{X}_1$ is the stock and $\mathbf{X}_2$ is K in cash, we can derive the following bound from [35]:

Corollary 6.6. ([35]) *The price of a European call option satisfies* $C(K,T) \leq \min_{1\leq\eta\leq\eta_{max}} \left(K^\eta + S_0^\eta e^{\eta(\eta-1)Q}\right)^{1/\eta} - K$.

For the average strike option we can optimize for η explicitly:

Corollary 6.7. *The price of an average strike call option satisfies* $AS(T) \leq S_0(e^{(\eta_{opt}-1)Q+(\ln 2)/\eta_{opt}} - 1)$, *where* $\eta_{opt} = \max\{1, \min\{\sqrt{(\ln 2)/Q}, \eta_{max}\}\}$.

We note that the above bound has different behaviors depending on the value of Q. The bound has a value of $S_0(e^{(\eta_{max}-1)Q+(\ln 2)/\eta_{max}} - 1)$ for $Q < (\ln 2)/\eta_{max}^2$, $S_0(e^{\sqrt{4Q\ln 2}-Q} - 1)$ for $(\ln 2)/\eta_{max}^2 \leq Q < \ln 2$, and (a trivial) S_0 for $Q \geq \ln 2$.

Average Price Call Options

Average price call options provide a smoothed version of European call options by averaging over the whole price path of the stock, and they are less expensive than European call options. To allow a counterpart of this phenomenon in our model, we will allow the quadratic variation parameter Q to depend on time. More specifically, we will assume that Q_t is an upper bound on $\sum_{\tau=1}^t r_\tau^2$, so $Q_1 \leq \ldots \leq Q_T$.

We start with a simple bound that relates the price of average price options to the price of ordinary call options and lookback options.

Theorem 6.8. *The prices of average price call options satisfy*

$$AP_G(K,T) \leq AP_A(K,T) \leq \frac{1}{T+1}\sum_{t=0}^{T} C(K,t) \leq LC(K,T)\ .$$

Proof. We have that

$$\prod_{t=0}^{T} S_t^{\frac{1}{T+1}} - K \leq \frac{1}{T+1}\sum_{t=0}^{T} S_t - K \leq \frac{1}{T+1}\sum_{t=0}^{T}\max\{S_t - K, 0\}\ ,$$

where the first inequality is by the inequality of the arithmetic and geometric means. We thus have that

$$\max\left\{\prod_{t=0}^{T} S_t^{\frac{1}{T+1}} - K, 0\right\} \leq \max\left\{\frac{1}{T+1}\sum_{t=0}^{T} S_t - K, 0\right\} \leq \frac{1}{T+1}\sum_{t=0}^{T}\max\left\{S_t - K, 0\right\}\ ,$$

and by the arbitrage-free assumption, $AP_G(K,T) \leq AP_A(K,T) \leq \frac{1}{T+1}\sum_{t=0}^{T} C(K,t)$.

Since $\max\{S_t - K, 0\} \le \max\{M_T - K, 0\}$ for every t, we also have that

$$\frac{1}{T+1}\sum_{t=0}^{T}\max\{S_t - K, 0\} \le \max\{M_T - K, 0\} ,$$

and, therefore, $\frac{1}{T+1}\sum_{t=0}^{T} C(K,t) \le LC(K,T)$. □

Using the bound of Corollary 6.6 on the price of call options in conjunction with the above theorem, we may obtain a concrete bound for both types of average price calls. However, its dependence on the sequence $Q_1, \ldots, Q_T$ is complicated. We therefore proceed with a more involved derivation that, in the case of $AP_G(K,T)$, will yield a simpler and more meaningful bound. This bound is obtained in a more general context, which is described next.

Consider n assets, $\mathbf{X}_1, \ldots, \mathbf{X}_n$, with returns $r_{i,t}$, for $1 \le i \le n$, $1 \le t \le T$, satisfying $|r_{i,t}| \le R_i$ and $\sum_{\tau=1}^{t} r_{i,\tau}^2 \le Q_{i,t}$. We assume w.l.o.g. that the initial values of all assets are 1. Given a probability vector $\mathbf{a} = (a_1, \ldots, a_n)$, we will consider a trading algorithm that maintains a constant fraction a_i of its wealth invested in asset $\mathbf{X}_i$, for each i, by rebalancing its holdings on every round. Such an algorithm is known as a *constantly rebalanced portfolio* (CRP), and we denote $\mathbf{P} = \mathbf{P}(\mathbf{a}, \mathbf{X}_1, \ldots, \mathbf{X}_n)$ for the algorithm and $\hat{r}_t$ for its return at time t. Since

$$P_t = \sum_{i=1}^{n} a_i P_{t-1}(1 + r_{i,t}) = P_{t-1}\left(1 + \sum_{i=1}^{n} a_i r_{i,t}\right) ,$$

we have that $\hat{r}_t = \sum_{i=1}^{n} a_i r_{i,t}$, and by the Cauchy-Schwarz inequality,

$$|\hat{r}_t| \le \|\mathbf{a}\|_2 \|(r_{1,t}, \ldots, r_{n,t})\|_2 \le \|\mathbf{a}\|_2 \|(R_1, \ldots, R_n)\|_2 .$$

By the first inequality, we also have $\sum_{\tau=1}^{t} \hat{r}_\tau^2 \le \|\mathbf{a}\|_2^2 \sum_{i=1}^{n} Q_{i,t}$. Thus, we may treat $\mathbf{P}$ as a single risky asset with bounds $\hat{R} = \|\mathbf{a}\|_2 \|(R_1, \ldots, R_n)\|_2$ and $\hat{Q} = \|\mathbf{a}\|_2^2 \sum_{i=1}^{n} Q_{i,T}$ on its absolute returns and quadratic variation, respectively.

One last property that will be required in bounding $AP_G(K,T)$ relates the geometric average of asset wealths to the wealth of a uniformly distributed CRP:

$$\begin{aligned}\prod_{i=1}^{n} X_{i,T}^{1/n} &= \prod_{i=1}^{n}\prod_{t=1}^{T}(1+r_{i,t})^{1/n} = \prod_{t=1}^{T}\prod_{i=1}^{n}(1+r_{i,t})^{1/n} \le \prod_{t=1}^{T}\frac{1}{n}\sum_{i=1}^{n}(1+r_{i,t}) \\ &= \prod_{t=1}^{T}\left(1+\frac{1}{n}\sum_{i=1}^{n} r_{i,t}\right) = P_T(\mathbf{a},\mathbf{X}_1,\ldots,\mathbf{X}_n)\ ,\end{aligned}$$

for $\mathbf{a} = (1/n,\ldots,1/n)$, where the inequality is due to the inequality of the arithmetic and geometric means, and the last equality holds since $\hat{r}_t = \sum_{i=1}^{n} a_i r_{i,t}$. With all these elements handy, we can now derive the desired bound.

Theorem 6.9. *Let* $\hat{Q}_T = \frac{1}{T+1}\sum_{t=1}^{T} Q_t$*,* $\hat{R} = \sqrt{\frac{T}{T+1}}R$*, and* $\hat{\eta}_{max} = \frac{1-2\hat{R}}{2\hat{R}(1-\hat{R})}$*. It holds that*

$$AP_G(K,T) \ \le\ \min_{1\le\eta\le\hat{\eta}_{max}} \left(K^{\eta} + S_0^{\eta} e^{\eta(\eta-1)\hat{Q}_T}\right)^{\frac{1}{\eta}} - K\ .$$

Proof. For $i = 0,\ldots,T$, define asset $\mathbf{X}_i$ as holding the stock until time i and then selling it. The option $\mathbf{AP}_G(K,T)$ is equivalent to a call option on the geometric mean of these assets. As already observed, this geometric mean is upper bounded by $P_T(\mathbf{a},\mathbf{X}_0,\ldots,\mathbf{X}_T)$, where $\mathbf{a} = (\frac{1}{T+1},\ldots,\frac{1}{T+1})$, and therefore it suffices to upper bound the initial price of the option $\mathbf{C}(\mathbf{P}(\mathbf{a},\mathbf{X}_0,\ldots,\mathbf{X}_T),K,T)$. By definition of the assets, $R_i = R$ and $Q_{i,T} = Q_i$ for $i \ne 0$, and $R_0 = Q_0 = 0$. Thus, denoting the single-period returns of the CRP by $\hat{r}_t$, we have that

$$|\hat{r}_t| \le \|\mathbf{a}\|_2 \|(R_0,\ldots,R_T)\|_2 \le (T+1)^{-\frac{1}{2}} T^{\frac{1}{2}} R = \sqrt{\frac{T}{T+1}}R\ ,$$

and

$$\sum_{\tau=1}^{T}\hat{r}_\tau^2 \le \|\mathbf{a}\|_2^2 \sum_{i=0}^{T} Q_{i,T} = \frac{1}{T+1}\sum_{t=1}^{T} Q_t\ .$$

Plugging these values into the bound of Corollary 6.6 for the price of a call option completes the proof. □

Since $\hat{\eta}_{max} \ge \eta_{max}$, the above expression does not exceed the bound of Corollary 6.6 with $\hat{Q}_T$ as the value of the quadratic variation. In other words, $AP_G(K,T)$ is upper bounded by the bound for a regular European call option with the *averaged* quadratic variation, which, depending on $Q_1,\ldots,Q_T$, may be significantly smaller than Q_T.

We conclude by comparing the bound of Theorem 6.9 with the bound based on averaging call option prices (Theorem 6.8), if we price the call options using the bound of Corollary 6.6. The fact that each call option bound may be minimized by a different value of η makes this comparison difficult. However, if the same value of η is used for all the call options, we may show that the bound of Theorem 6.8 cannot be superior.

Lemma 6.10. *Let $\hat{Q}_T = \frac{1}{T+1}\sum_{t=1}^{T} Q_t$, $\hat{R} = \sqrt{\frac{T}{T+1}}R$, and $\hat{\eta}_{max} = \frac{1-2\hat{R}}{2\hat{R}(1-\hat{R})}$. It holds that*

$$\min_{1\leq\eta\leq\hat{\eta}_{max}} (K^\eta + e^{\eta(\eta-1)\hat{Q}_T})^{\frac{1}{\eta}} - K \leq \min_{1\leq\eta\leq\eta_{max}} \frac{1}{T+1}\sum_{t=0}^{T}\left((K^\eta + e^{\eta(\eta-1)Q_t})^{\frac{1}{\eta}} - K\right) .$$

Proof. Let $a \geq 0$, $b > 0$, and $\eta \geq 1$, and denote $f(x) = \ln(a + b^x)$. Then

$$f''(x) = \left(\frac{b^x \ln b}{a + b^x}\right)' = \left(\ln b - \frac{a \ln b}{a + b^x}\right)' = \frac{ab^x(\ln b)^2}{(a + b^x)^2} \geq 0 ,$$

so f is convex and so is $(1/\eta)\cdot f = \ln(a + b^x)^{1/\eta}$. Thus, $(a + b^x)^{1/\eta}$ is log-convex and therefore convex. Taking $a = K^\eta$, $b = e^{\eta(\eta-1)}$ and subtracting K we have that $(K^\eta + e^{\eta(\eta-1)x})^{1/\eta} - K$ is convex in x, and as a result,

$$\begin{aligned}(K^\eta + e^{\eta(\eta-1)\hat{Q}_T})^{\frac{1}{\eta}} - K &= \left(K^\eta + e^{\eta(\eta-1)\frac{1}{T+1}\sum_{t=0}^{T} Q_t}\right)^{\frac{1}{\eta}} - K \\ &\leq \frac{1}{T+1}\sum_{t=0}^{T}\left((K^\eta + e^{\eta(\eta-1)Q_t})^{\frac{1}{\eta}} - K\right) ,\end{aligned}$$

where we define $Q_0 = 0$. It is easily verified that the expression $\frac{1-2R}{2R(1-R)}$ is decreasing in R, and therefore $\hat{R} < R$ implies $[1, \eta_{max}] \subseteq [1, \hat{\eta}_{max}]$, and the result follows. □

6.3 Convex Path-Independent Derivatives

In this section we move beyond specific options, and provide adversarial price bounds for a general class of derivatives. For this purpose, we add a new assumption to our model, namely, that the final stock price, S_T, has finite resolution. We thus assume that $S_T \in \mathcal{P} = \{j\Delta p\colon 0 \leq j \leq N\}$, where $\Delta p = M/N$, for some $N > 2$. This assumption mirrors the situation in reality.

It is a well-known result from finance, that exact pricing of European call options yields exact pricing for every path-independent derivative of a stock [20]. This result relies on the fact that every derivative is equivalent to a portfolio of call options with various strike prices.

Formally, for any $f\colon \mathcal{P} \to \mathbb{R}$, we define $f^{(1)}_{\Delta p}\colon \mathcal{P} \setminus \{M\} \to \mathbb{R}$ as $f^{(1)}_{\Delta p}(p) = \frac{f(p+\Delta p)-f(p)}{\Delta p}$ and $f^{(2)}_{\Delta p}\colon \mathcal{P} \setminus \{0, M\} \to \mathbb{R}$ as $f^{(2)}_{\Delta p}(p) = \frac{f(p+\Delta p)-2f(p)+f(p-\Delta p)}{\Delta p^2}$. The next lemma and theorem follow immediately from a result in [20].

Lemma 6.11. ([20]) *Define $g_K\colon \mathcal{P} \to \mathbb{R}$ as $g_K(p) = \max\{p - K, 0\}$. For every $p \in \mathcal{P}$,*

$$f(p) \quad = \quad f(0) + f^{(1)}_{\Delta p}(0) \cdot g_0(p) + \sum_{K=\Delta p}^{M-\Delta p} f^{(2)}_{\Delta p}(K) \cdot g_K(p) \cdot \Delta p \,.$$

Note that $g_K(p)$ is the payoff of $\mathbf{C}(K, T)$ and that, in addition, $C(0, T) = S_0$, because the payoff of $\mathbf{C}(0, T)$ is equivalent to a single stock. We thus get the following theorem:

Theorem 6.12. ([20]) *Let $\mathbf{X}$ be a path-independent derivative with payoff $f\colon \mathcal{P} \to \mathbb{R}$. The initial value of $\mathbf{X}$ is given by*

$$X \quad = \quad f(0) + f^{(1)}_{\Delta p}(0) \cdot S_0 + \sum_{K=\Delta p}^{M-\Delta p} f^{(2)}_{\Delta p}(K) \cdot C(K, T) \cdot \Delta p \,.$$

In the BSM model, where an exact pricing of call options is known, this amounts to pricing all path-independent derivatives exactly. In our model, however, where only *upper bounds* on call option prices are available, we cannot utilize this relation in every case. We must have that $f^{(2)}_{\Delta p}(K) \geq 0$ for every K in order to substitute those upper bounds for the terms $C(K, T)$. This requirement is fulfilled by convex derivatives.

Theorem 6.13. *Let $\mathbf{X}$ be a path-independent derivative with payoff $f\colon \mathcal{P} \to \mathbb{R}$, where f is the restriction to $\mathcal{P}$ of some convex function $\bar{f}\colon [0, M] \to \mathbb{R}$. Let $C(K, T) \leq U(K)$ for every $K \in \mathcal{P}$, where $U\colon \mathcal{P} \to \mathbb{R}$. Then*

$$X \quad \leq \quad f(0) + f^{(1)}_{\Delta p}(0) \cdot S_0 + \sum_{K=\Delta p}^{M-\Delta p} f^{(2)}_{\Delta p}(K) \cdot U(K) \cdot \Delta p \,.$$

Proof. By Theorem 6.12, it is enough to show that $f^{(2)}_{\Delta p}(K) \geq 0$ for every $K \in [\Delta p, M - \Delta p]$. By the convexity of $\bar{f}$, $f(K) = f(\frac{1}{2}(K - \Delta p) + \frac{1}{2}(K + \Delta p)) \leq \frac{1}{2} f(K - \Delta p) + \frac{1}{2} f(K + \Delta p)$, and thus $\frac{f(K+\Delta p)-2f(K)+f(K-\Delta p)}{\Delta p^2} \geq 0$. ∎

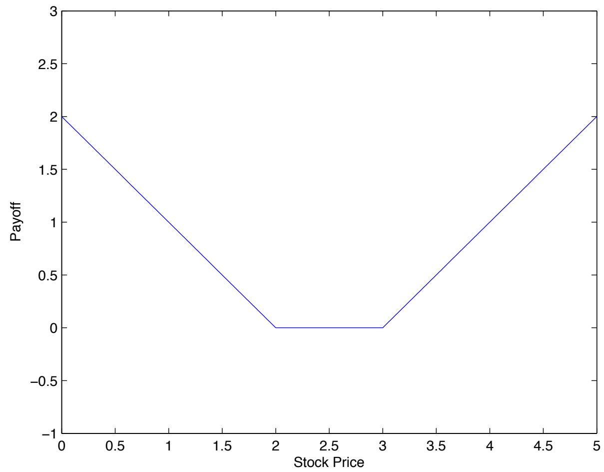

Figure 6.1: The payoff of a long strangle investment strategy as a function of the stock price, for $K_1 = 2$, $K_2 = 3$.

An Example: a Long Strangle Strategy

A long strangle investment strategy involves the purchase of a put option with a strike price of K_1 and a call option with a strike price of K_2, where $K_1 < K_2$. The payoff of this strategy is $\max\{K_1 - S_T, 0\} + \max\{S_T - K_2, 0\}$, which is a convex function of S_T (see Figure 6.1). By Theorem 6.12, the value of a long strangle is $K_1 - S_0 + C(K_1, T) + C(K_2, T)$. Denoting $C_u(K, T)$ for the bound of Corollary 6.6, we can upper bound the price of a long strangle by $K_1 - S_0 + C_u(K_1, T) + C_u(K_2, T)$.

6.4 Discussion of the Bounds

Comparison to standard portfolio selection. The problem of pricing $\mathbf{\Psi}(\mathbf{X}_1, \ldots, \mathbf{X}_n, T)$ may be viewed as a portfolio selection problem with a reference class of n alternative strategies. For example, in the case of lookback options, $n = T + 1$, and the strategies are holding cash or selling the stock at some time t, for $1 \le t \le T$. Working with this simple reference class, we can achieve better regret bounds than those given w.r.t. superior benchmarks such as the best CRP. Consider a naïve *buy and hold* algorithm that initially divides capital equally between the alternative strategies, performing no further action. This algorithm has a $(1/n, \ldots, 1/n)$ multiplicative regret w.r.t. the n strategies. In standard portfolio selection terminology, where regret is defined as the logarithm of the ratio between the final wealth of the best strategy in the reference class and the final wealth of the algorithm, the naïve algorithm has a regret upper bound of $\ln n$. In comparison, the universal portfolio's (optimal) bound w.r.t. the best CRP is $\frac{n-1}{2}\ln 2T + \ln\frac{\Gamma(1/2)^n}{\Gamma(n/2)} + o(1)$ (see, e.g., [23]). In the case of look-back options, we get an $\Omega(T)$ regret bound, which grows arbitrarily large as the trading frequency increases. The problematic dependence on T persists even if n is small, for example, in the case of call options, where $n = 2$ (the alternatives are holding cash or holding the stock). More recent algorithms [28, 48] replaced the dependence on T with different notions of variation of the single-period returns (different from quadratic variation). Their bounds are a great improvement over the previous bound, under realistic conditions where the variation is much smaller than T. Nevertheless, the bound of [48] is still lower bounded by the naïve bound of $\ln n$, for every n, while the bound given in [28] is $\Omega(n \ln n)$ (although we note that the constant is not clear from their result). Therefore, the regret bounds w.r.t. the best CRP cannot be used directly to achieve interesting upper bounds on the price of $\mathbf{\Psi}$ options.

Bounds for average price call options. Average price call options are less expensive than European call options in practice due to the smoothing effect of the price averaging. In the BSM model it is possible to derive an analytic pricing formula only in the case of geometric averaging. Interestingly, the BSM price of such an option is equivalent to the price of a European call option on an asset with a volatility parameter that is smaller by a multiplicative factor of $\sqrt{3}$ (see, e.g., [57]).[2]

The quadratic variation plays a similar role to the volatility in our model. The result of Theorem 6.9 upper bounds $AP_G(K, T)$ in terms of the call option price where the quadratic variation is replaced by its averaged value. That in itself does not necessarily provide a meaningful improvement over the price of a European call option, since all the quadratic variation may be concentrated in the first few rounds, causing the averaged value and the final value of the quadratic variation to be roughly the same.

We may therefore consider a reasonable model of gradual growth in the quadratic variation, by assuming a bound on Q_t that grows linearly, as $a + \frac{Q_T - a}{T-1}\cdot(t-1)$, where a is a parameter. (The assumption becomes more robust as a grows, until for $a = Q_T$ we recover our original model.) It holds that

$$\frac{1}{T+1}\sum_{t=1}^{T} Q_t \leq \frac{1}{T+1}\cdot\frac{T}{2}\cdot(a + Q_T) \leq \frac{1}{2}(a + Q_T) ,$$

[2] This asset also pays a dividend.

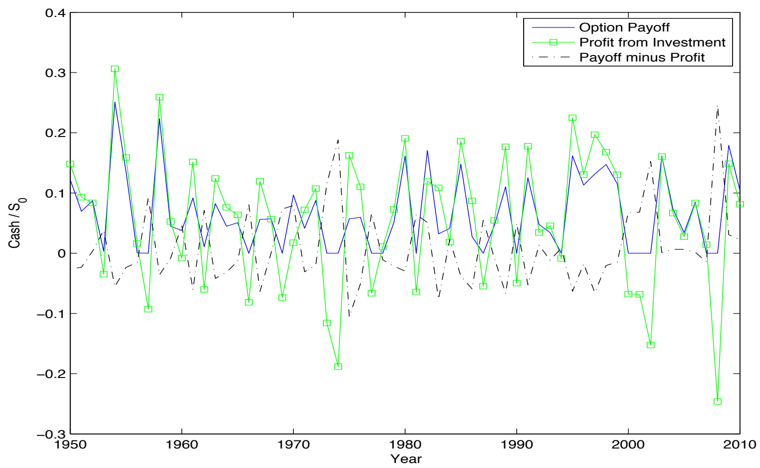

Figure 6.2: A breakdown of the option writer's cash flow, for average strike call options, in terms of option payoff, option price, and total profit from the algorithm's trading. Data is calculated for 1-year average strike call options on the S&P 500 index for the years 1950-2010, with $R = 0.15$ and $Q = 0.1$. For the option writer to make a profit, the payoff minus profit line must be below the option price, which is 0.53 in this setting. Note that the "hindsight" empirical price is 0.25.

where we simply summed an arithmetic sequence. Thus, the smoothed quadratic variation featured in Theorem 6.9 is bounded by $(a + Q_T)/2$ rather than Q_T. If $a \ll Q_T$, the value is about half the original quadratic variation of the stock.

6.5 Empirical Results

In order to examine our results empirically, we consider the S&P 500 index data for the years 1950-2010. (The results are plotted in Figure 6.2.) We computed a price for a 1-year average strike call option using our bound, with $R = 0.15$ and $Q = 0.1$. These R and Q values hold for all years but two in the test. In addition, for each year we computed the payoff of an average strike call option and also ran the Generic algorithm and computed the total profit. We used a single value of η, namely, the optimal value for the price bound. The stock prices for each year were normalized so that $S_0 = 1$ at the beginning of the year.

It is instructive to compare our upper bound on the option price, which was calculated to be 0.53, to the net payoff. The net payoff is the difference between the payoff to the option holder (always non-negative) and the profit (or loss) the algorithm made in trading. It can be seen that our option price dominates the net payoff for every year, with the maximal net payoff at 0.25.

We point out that our empirical results are influenced by the fact that we assume a zero risk-free interest rate, while the interest rates pertaining to the data used were positive and variable.

Chapter 7

A Closer Look at Lookback Options

7.1 Revisiting Multiplicative Regret

Recall the payoff of a fixed-strike lookback option, which is $\max\{M_T - K, 0\}$, or equivalently, $\max\{K, S_1, \ldots, S_T\} - K$, since $K \geq S_0$. Thus, if we define $\mathbf{X}_t$ to be an algorithm that buys a single stock at time 0 and sells it at time t, $1 \leq t \leq T$, and $\mathbf{X}$ to be holding K in cash, then $LC(K,T) = \Psi(\mathbf{X}, \mathbf{X}_1, \ldots, \mathbf{X}_T, T) - K$. If we have a trading algorithm that starts with $V_0 = 1$ and guarantees $V_T \geq \alpha$ and $V_T \geq \beta_t X_{t,T}$, $1 \leq t \leq T$, for every price path (which we defined as having an $(\alpha/K, \beta_1, \ldots, \beta_T)$ multiplicative regret w.r.t. $\mathbf{X}, \mathbf{X}_1, \ldots, \mathbf{X}_T$), then we obtain by Lemma 6.2 that

$$LC(K,T) \leq \frac{1}{\min\{\alpha/K, \beta_1, \ldots, \beta_T\}} - K \ .$$

Observe that if we substitute a single value $\beta = \min\{\alpha, \beta_1, \ldots, \beta_T\}$ for all values β_t, we only weaken our requirements without weakening the bound. The guarantees now become $V_T \geq \max\{\alpha, \beta M_T\}$, and we have $LC(K,T) \leq 1/\min\{\alpha/K, \beta\} - K$. It is therefore convenient to use an adjusted definition of multiplicative regret in the context of lookback options, as follows:

Definition 7.1. *Let A be a trading algorithm and let V_t be its asset value at time $0 \leq t \leq T$, assuming $V_0 = 1$. The algorithm A is said to have an (α, β) multiplicative regret, for some $\alpha, \beta > 0$, if for every price path, $V_T \geq \max\{\alpha, \beta M_T\}$.*

We thus have the following bound on the price of lookback options:

Lemma 7.2. *If an algorithm with (α, β) multiplicative regret exists, then*

$$LC(K,T) \leq \frac{1}{\min\{\alpha/K, \beta\}} - K \ .$$

7.2 Simple Arbitrage-Free Bounds

We start with a few simple arbitrage-free bounds. Let A_{TS} be a trading algorithm that starts with T stocks (initial value S_0T) and sells one stock at each time $t \in [1,T]$. Since the final value of A_{TS} is $\sum_{t=1}^{T} S_t \geq M_T - S_0 \geq M_T - K$, we have that

Theorem 7.3. *The algorithm A_{TS} dominates* $\mathbf{LC}(K,T)$*, implying that*

$$LC(K,T) \leq TS_0 = T \ .$$

A similar strategy A_{TC} uses ordinary call options rather than stocks. It buys one call option for each expiration time $t \in [1,T]$ (initial value $\sum_{t=1}^{T} C(K,t)$) and simply collects the payoffs. Since the final value of A_{TC} is $\sum_{t=1}^{T} \max\{S_t - K, 0\} \geq \max\{M_T - K, 0\}$, we have that

Theorem 7.4. *The algorithm A_{TC} dominates* $\mathbf{LC}(K,T)$*, therefore,*

$$LC(K,T) \leq \sum_{t=1}^{T} C(K,t) \ .$$

While the previous strategies were time-oriented, the next strategy is price-oriented. The algorithm A_{PS} starts with $N = 1 + \lfloor \log_2(M/K) \rfloor$ stocks, and, therefore, has an initial value of NS_0. It sells stock i, $i \in [0, N-1]$, at the first time $S_t \geq 2^i K$. (This strategy is very similar to one discussed in [36].)

Theorem 7.5. *The algorithm A_{PS} dominates* $\mathbf{LC}(K,T)$*, and*

$$LC(K,T) \leq NS_0 = 1 + \lfloor \log_2(M/K) \rfloor \ .$$

Proof. Let c be the total wealth obtained from selling stocks. If no stock is sold, then $c = 0 \geq \max\{M_T - K, 0\} = 0$. Otherwise, let $i_0 = \lfloor \log_2(M_T/K) \rfloor$ be the highest index of a stock that was sold. We will prove that $c \geq 2^{i_0+1}K - K$, and then, since $2^{i_0+1}K - K \geq M_T - K = \max\{M_T - K, 0\}$, the proof would be complete. We proceed by induction on i_0. If $i_0 = 0$, then $c \geq K = 2^{i_0+1}K - K$. We assume correctness for values smaller than i_0 and prove for i_0. If another stock is sold at the same time as stock i_0, then $c \geq 2 \cdot 2^{i_0}K \geq 2^{i_0+1}K - K$. If stock i_0 is sold alone, then applying the induction assumption to stock $i_0 - 1$, and adding the wealth at the time stock $i_0 - 1$ was sold, we have $c \geq (2^{i_0}K - K) + 2^{i_0}K = 2^{i_0+1}K - K$, and the induction is complete. □

7.3 Combining Regret Minimization and One-Way Trading

The simple algorithms reveal what is required of an algorithm to dominate $\mathbf{LC}(K, T)$. Such an algorithm must always set aside enough cash to cover the payoff of the option. In the vicinity of record highs, cash reserves may need to be increased by selling stocks. At the same time, enough stocks must be retained in case even higher price levels are reached.

The problem may also be seen as that of predicting the point where the stock price reaches a maximum. We may think of a set of experts, each with its own selling trigger as a function of market conditions, where the job of the algorithm is to choose the best expert. This casts the problem in the best expert and regret minimization frameworks. Such a formulation turns out to be equivalent to a conventional regret minimization algorithm working with stock and cash, combined with a one-way trading mechanism.

Our algorithm thus contains both a regret minimization element and a one-way trading element. This combination will enable us to bound the regret w.r.t. stock prices *over the whole price path of the stock.* This is in contrast to performing conventional regret minimization, which would only bound the regret w.r.t. the final stock price. At the same time, this approach is also a generalization of regret minimization with stock and cash, because by picking an inactive one-way trading element, we can obtain an ordinary regret minimization algorithm.

We define our trading algorithm using two parameters, a *one-way trading rule* and a *regret minimization rule.* Intuitively, the one-way trading rule will move funds gradually from stock to cash, while the regret minimization rule will try to follow whichever asset is performing better, stock or cash (using a regret minimization algorithm).

The *one-way trading rule* is defined by a function h mapping histories to $[0, 1]$. We require that h be monotonically non-increasing along a price path, i.e., if $\mathbf{r}_t$ is a prefix of $\mathbf{r}_{t+1}$ then $h(\mathbf{r}_t) \geq h(\mathbf{r}_{t+1})$. In addition, h will start at 1, i.e., $h(\mathbf{r}_0) = 1$. In the algorithm we will use the notation $H_t = h(\mathbf{r}_{t-1})$, so that $1 = H_1 \geq H_2 \geq \ldots \geq 0$, and $H = (H_1, \ldots, H_{T+1})$. If we set $H_t = 1$ for every t, we will essentially not transfer funds from stock to cash.

The *regret minimization rule* performs regret minimization between stock and cash. It is defined by the update equations $w_{s,t+1} = w_{s,t} f(\mathbf{r}_t)$, and $w_{c,t+1} = w_{c,t}$, where $f : \cup_{t\geq 1}\mathbb{R}^t \to \mathbb{R}^+$. In what follows, we use $f(\mathbf{r}_t) = 1 + \eta r_t$, which is the regret minimization rule of the Polynomial Weights algorithm [25], as adapted in [35]. We will require that $\eta \in [1, \eta_{max}]$, where $\eta_{max} = \frac{1-2R}{2R(1-R)}$, and that $R < 1 - 1/\sqrt{2} \approx 0.3$. In Section 7.4 we will show that $\eta = 1$ has the interpretation of a *buy and hold* strategy, namely, the regret minimization rule maintains its initial allocation of stock and cash. For convenience, we denote $g(\mathbf{r}_t) = \prod_{u=1}^{t} f(\mathbf{r}_u)$, for the accumulated update of f.

We now give the exact definition of our family of trading algorithms. Initially, $w_{c,1}, w_{s,1} > 0$, and for $t \geq 1$,

$$
\begin{aligned}
w_{s,t+1} &= w_{s,t}\frac{H_{t+1}}{H_t} f(\mathbf{r}_t) = w_{s,1} H_{t+1} g(\mathbf{r}_t) \ , \\
w_{c,t+1} &= w_{c,t} + w_{s,1}(H_t - H_{t+1}) g(\mathbf{r}_t) = w_{c,t} + w_{s,t} f(\mathbf{r}_t) - w_{s,t+1} \ .
\end{aligned}
$$

Recall that a trading algorithm invests a fraction $w_{s,t}/(w_{s,t} + w_{c,t})$ of assets in stock, and the rest in cash. We assume that initially, $V_0 = 1$. The following lemma relates the current weights to the initial weights, using the stock price, the variation parameter Q and the one-way trading rule.

Lemma 7.6. *For every $\eta \in [1, \eta_{max}]$, it holds that*

$$
\begin{aligned}
w_{s,t+1} &\geq H_{t+1} w_{s,1} S_t^{\eta} e^{-\eta(\eta-1)Q} \ , \\
w_{c,t+1} &\geq w_{c,1} + w_{s,1} e^{-\eta(\eta-1)Q} \sum_{u=1}^{t} (H_u - H_{u+1}) S_u^{\eta} \ .
\end{aligned}
$$

Proof. We first observe that for every $1 \leq \tau \leq T$, since $g(\mathbf{r}_\tau) = \prod_{u=1}^{\tau} f(\mathbf{r}_u) = \prod_{u=1}^{\tau}(1+\eta r_u)$, we have that

$$
\begin{aligned}
\ln g(\mathbf{r}_\tau) &= \sum_{u=1}^{\tau} \ln(1 + \eta r_u) \geq \eta \sum_{u=1}^{\tau} \ln(1 + r_u) - \eta(\eta-1) \sum_{u=1}^{\tau} r_u^2 \\
&= \eta \ln S_\tau - \eta(\eta-1) \sum_{u=1}^{\tau} r_u^2 \geq \ln S_\tau^{\eta} - \eta(\eta-1)Q = \ln S_\tau^{\eta} e^{-\eta(\eta-1)Q} \ ,
\end{aligned}
$$

where the first inequality uses the fact that for every $\eta \in [1, \eta_{max}]$ and $r > -R$, $\ln(1 + \eta r) \geq \eta \ln(1 + r) - \eta(\eta - 1)r^2$ (see [35]). Thus, for every $1 \leq \tau \leq T$, $g(\mathbf{r}_\tau) \geq S_\tau^\eta e^{-\eta(\eta-1)Q}$. We therefore have that

$$\frac{w_{s,t+1}}{w_{s,1}} = \prod_{u=1}^{t} \frac{w_{s,u+1}}{w_{s,u}} = \prod_{u=1}^{t} \frac{H_{u+1}}{H_u} f(\mathbf{r}_u) = \frac{H_{t+1}}{H_1} g(\mathbf{r}_t) \geq H_{t+1} S_t^\eta e^{-\eta(\eta-1)Q} ,$$

for the weight of stock, and

$$\begin{aligned} w_{c,t+1} &= w_{c,1} + \sum_{u=1}^{t} (w_{c,u+1} - w_{c,u}) = w_{c,1} + w_{s,1} \sum_{u=1}^{t} (H_u - H_{u+1}) g(\mathbf{r}_u) \\ &\geq w_{c,1} + w_{s,1} \sum_{u=1}^{t} (H_u - H_{u+1}) S_u^\eta e^{-\eta(\eta-1)Q} \\ &= w_{c,1} + w_{s,1} e^{-\eta(\eta-1)Q} \sum_{u=1}^{t} (H_u - H_{u+1}) S_u^\eta , \end{aligned}$$

for the weight of cash. Both inequalities used our lower bound on $g(\mathbf{r}_\tau)$. □

The following theorem lower bounds the profit of the algorithm in terms of H, and facilitates the proof of (α, β) multiplicative regret results for given one-way trading rules.

Theorem 7.7. *For every $\eta \in [1, \eta_{max}]$,*

$$V_T \geq \left(p_{c,1} + p_{s,1} e^{-\eta(\eta-1)Q} \left(\sum_{t=1}^{T} (H_t - H_{t+1}) S_t^\eta + H_{T+1} S_T^\eta \right) \right)^{\frac{1}{\eta}} .$$

Proof. We have that

$$\begin{aligned} \ln \frac{W_{T+1}}{W_1} &= \sum_{t=1}^{T} \ln \frac{W_{t+1}}{W_t} = \sum_{t=1}^{T} \ln \frac{w_{c,t+1} + w_{s,t+1}}{W_t} = \sum_{t=1}^{T} \ln \frac{w_{c,t} + w_{s,t} f(\mathbf{r}_t)}{W_t} \\ &= \sum_{t=1}^{T} \ln \frac{w_{c,t} + w_{s,t}(1 + \eta r_t)}{W_t} = \sum_{t=1}^{T} \ln(1 + \eta p_{s,t} r_t) \\ &= \sum_{t=1}^{T} \ln \left(1 + \eta \left(\frac{V_t}{V_{t-1}} - 1 \right) \right) \leq \sum_{t=1}^{T} \eta \ln \frac{V_t}{V_{t-1}} = \eta \ln V_T , \end{aligned}$$

where the inequality uses the fact that for every $\eta \in [1, \eta_{max}]$ and $r > -R$, $\ln(1 + \eta r) \leq$

$\eta \ln(1+r)$ (see [35]). On the other hand, we have that

$$\begin{aligned}
\ln \frac{W_{T+1}}{W_1} &= \ln \frac{w_{c,T+1} + w_{s,T+1}}{W_1} \\
&\geq \ln \left(\frac{1}{W_1} \left(w_{c,1} + w_{s,1} e^{-\eta(\eta-1)Q} \left(\sum_{t=1}^{T} (H_t - H_{t+1}) S_t^{\eta} + H_{T+1} S_T^{\eta} \right) \right) \right) \\
&= \ln \left(p_{c,1} + p_{s,1} e^{-\eta(\eta-1)Q} \left(\sum_{t=1}^{T} (H_t - H_{t+1}) S_t^{\eta} + H_{T+1} S_T^{\eta} \right) \right) ,
\end{aligned}$$

where the inequality is by Lemma 7.6. Together, we have that

$$\eta \ln V_T \geq \ln \left(p_{c,1} + p_{s,1} e^{-\eta(\eta-1)Q} \left(\sum_{t=1}^{T} (H_t - H_{t+1}) S_t^{\eta} + H_{T+1} S_T^{\eta} \right) \right) ,$$

and, rearranging, we get the desired result. □

7.3.1 A Price-Oriented Rule and Bound

The one-way trading rule presented in this subsection is reminiscent of the simple algorithm A_{PS} in that it depends only on the maximal price so far, M_t. Like A_{PS}, unless the price is at an all-time high, this rule does not trigger a sale of stocks, i.e., $H_{t+1} = H_t$. However, the combined trading algorithm may still sell stocks at any price level due to the regret minimization rule.

Lemma 7.8. *Using $H_t = \ln(M/\mu_{t-1})/\ln(M/K)$, where $\mu_t = \max\{K, M_t\}$, the algorithm achieves an (α, β) multiplicative regret with $\alpha = p_{c,1}^{1/\eta}$, and*

$$\beta = \min \left\{ \frac{p_{c,1}^{\frac{1}{\eta}}}{K}, \left(\frac{p_{s,1} e^{-\eta(\eta-1)Q}}{\eta \ln(M/K)} \right)^{\frac{1}{\eta}} \right\} .$$

Proof. From Theorem 7.7 we have that

$$V_T \geq \left[p_{c,1} + p_{s,1} e^{-\eta(\eta-1)Q} \left(\sum_{t=1}^{T} (H_t - H_{t+1}) S_t^{\eta} + H_{T+1} S_T^{\eta} \right) \right]^{\frac{1}{\eta}} . \tag{7.1}$$

Since $H_1 \geq \ldots \geq H_{T+1} \geq 0$, we have that $V_T \geq p_{c,1}^{1/\eta} = \alpha$, which proves the α part of the (α, β) multiplicative regret. For the β part, we consider two cases. If $M_T \leq K$, then $V_T \geq p_{c,1}^{1/\eta} \geq p_{c,1}^{1/\eta} K^{-1} M_T \geq \beta M_T$, as required. We now assume $M_T > K$. From

Equation 7.1 we get

$$V_T \geq \left(p_{c,1} + p_{s,1}e^{-\eta(\eta-1)Q}\frac{1}{\ln(M/K)}\sum_{t=1}^{T} S_t^\eta \ln\frac{\mu_t}{\mu_{t-1}}\right)^{\frac{1}{\eta}} , \tag{7.2}$$

by definition of H_t and using the fact that $H_{T+1} \geq 0$. The values μ_t form a non-decreasing sequence, and we have that $\mu_T = M_T > K = \mu_0$. Let $\mu_T^\eta = x_l > x_{l-1} > \ldots > x_0 = \mu_0^\eta$ be the distinct values of μ_t^η, $0 \leq t \leq T$, for some $l \geq 1$. Note that $S_t^\eta \ln\frac{\mu_t}{\mu_{t-1}} = 0$ if $\mu_t = \mu_{t-1}$, and $S_t^\eta \ln\frac{\mu_t}{\mu_{t-1}} = \mu_t^\eta \ln\frac{\mu_t}{\mu_{t-1}} = \frac{1}{\eta}\cdot\mu_t^\eta \ln\frac{\mu_t^\eta}{\mu_{t-1}^\eta}$ otherwise. Therefore,

$$\eta\sum_{t=1}^{T} S_t^\eta \ln\frac{\mu_t}{\mu_{t-1}} = \sum_{i=1}^{l} x_i \ln\frac{x_i}{x_{i-1}} \geq x_l - x_0 = \mu_T^\eta - \mu_0^\eta = M_T^\eta - K^\eta ,$$

where the inequality is true since for every $z_0, \ldots, z_n > 0$, we have that $\sum_{i=1}^{n} z_i \ln(z_i/z_{i-1}) \geq z_n - z_0$ (see Lemma 7.17 in Section 7.6). Plugging this into Equation 7.2 gives

$$V_T \quad \geq \quad \left(p_{c,1} + \frac{p_{s,1}e^{-\eta(\eta-1)Q}}{\eta\ln(M/K)}\left(M_T^\eta - K^\eta\right)\right)^{\frac{1}{\eta}} .$$

Denoting $\gamma = \frac{p_{s,1}e^{-\eta(\eta-1)Q}}{\eta\ln(M/K)}$, we have

$$V_T \quad \geq \quad \left((p_{c,1} - \gamma K^\eta) + \gamma M_T^\eta\right)^{\frac{1}{\eta}} = \left(\frac{p_{c,1} - \gamma K^\eta}{M_T^\eta} + \gamma\right)^{\frac{1}{\eta}} \cdot M_T .$$

If $p_{c,1} - \gamma K^\eta \geq 0$, then $V_T \geq \gamma^{1/\eta} M_T \geq \beta M_T$, as required. Otherwise, since $M_T > K$, we have that

$$\left(\frac{p_{c,1} - \gamma K^\eta}{M_T^\eta} + \gamma\right)^{\frac{1}{\eta}} \geq \left(\frac{p_{c,1} - \gamma K^\eta}{K^\eta} + \gamma\right)^{\frac{1}{\eta}} = \frac{p_{c,1}^{\frac{1}{\eta}}}{K} ,$$

and, therefore, $V_T \geq (p_{c,1}^{1/\eta}/K) \cdot M_T \geq \beta M_T$, and the proof is complete. □

This multiplicative regret result implies a new bound on $LC(K, T)$.

Theorem 7.9. *For $\eta \in [1, \eta_{max}]$, $LC(K, T) \leq (K^\eta + S_0^\eta \eta e^{\eta(\eta-1)Q} \ln\frac{M}{K})^{1/\eta} - K$.*

Proof. Recall that $S_0 = 1$. Combining Lemmas 7.2 and 7.8, we have that

$$LC(K,T) \leq \frac{1}{\min\{\alpha/K, \beta\}} - K \ ,$$

where $\alpha = p_{c,1}^{1/\eta}$ and $\beta = \min\{p_{c,1}^{1/\eta}/K, \gamma^{1/\eta}\}$, denoting $\gamma = \frac{p_{s,1}e^{-\eta(\eta-1)Q}}{\eta \ln(M/K)}$. Since $\alpha/K \geq \beta$, we have $LC(K,T) \leq 1/\beta - K = \max\{Kp_{c,1}^{-1/\eta}, \gamma^{-1/\eta}\} - K$. We now optimize for $p_{c,1} = 1 - p_{s,1}$. The term $Kp_{c,1}^{-1/\eta}$ is decreasing in $p_{c,1}$, while the term $\gamma^{-1/\eta}$ is increasing in $p_{c,1}$. Thus, to minimize the bound, we would like to pick $p_{c,1}$ such that $Kp_{c,1}^{-1/\eta} = \gamma^{-1/\eta}$. It is easy to verify that $p_{c,1} = (1 + K^{-\eta}\eta e^{\eta(\eta-1)Q}\ln(M/K))^{-1}$ satisfies this requirement, and is also in $[0,1]$, and thus a valid choice. Therefore,

$$\begin{aligned} LC(K,T) &\leq Kp_{c,1}^{-\frac{1}{\eta}} - K = K \cdot \left(1 + K^{-\eta}\eta e^{\eta(\eta-1)Q}\ln(M/K)\right)^{\frac{1}{\eta}} - K \\ &= \left(K^{\eta} + \eta e^{\eta(\eta-1)Q}\ln(M/K)\right)^{\frac{1}{\eta}} - K \ . \end{aligned}$$

□

7.3.2 Bounds Based on Competitive Ratio

For every one-way trading rule h, and every $\eta \in [1, \eta_{max}]$, we may define

$$\rho_h(\eta) = \sup_{\Pi} \frac{M_T^{\eta}}{\sum_{t=1}^{T}(H_t - H_{t+1})S_t^{\eta} + H_{T+1}S_T^{\eta}} \ .$$

Namely, $\rho_h(\eta)$ is the competitive ratio of h w.r.t. the sequence $\{S_t^{\eta}\}_{t=1}^{T}$. Assume $\bar{\rho}_h$ is a known upper bound on ρ_h, so $\bar{\rho}_h(\eta) \geq \rho_h(\eta)$ for every $\eta \in [1, \eta_{max}]$ (if ρ_h is known explicitly, then $\bar{\rho}_h = \rho_h$). We next bound $LC(K,T)$ in terms of $\bar{\rho}_h(\eta)$.

Theorem 7.10. *Let h be the one-way trading rule used in the trading algorithm. For every $\eta \in [1, \eta_{max}]$, there is a choice of $p_{c,1}$ for which the trading algorithm has a $(\beta K, \beta)$ multiplicative regret, and, therefore, $LC(K,T) \leq 1/\beta - K$, where*

$$\beta = \left(\frac{K^{\eta} - S_0^{\eta} + S_0^{\eta}e^{\eta(\eta-1)Q}\bar{\rho}_h(\eta)\left(1 - (K/M)^{\eta}\right)}{1 - (S_0/M)^{\eta}}\right)^{-\frac{1}{\eta}} \ .$$

Proof. Recall that $S_0 = 1$. Denote $b = e^{-\eta(\eta-1)Q}\bar{\rho}_h(\eta)^{-1}$ and $p = p_{c,1}$ for convenience.

By definition of $\bar{\rho}_h(\eta)$ and by Theorem 7.7,

$$V_T \geq (p + (1-p)e^{-\eta(\eta-1)Q}\bar{\rho}_h(\eta)^{-1}M_T^\eta)^{\frac{1}{\eta}} = (p + (1-p)bM_T^\eta)^{\frac{1}{\eta}} \ .$$

Thus, we have that $V_T \geq (p + (1-p)b)^{1/\eta} = (p(1-b)+b)^{1/\eta}$, and, in addition,

$$V_T \geq M_T \left(pM^{-\eta} + (1-p)b\right)^{\frac{1}{\eta}} = M_T \left(p\left(M^{-\eta} - b\right) + b\right)^{\frac{1}{\eta}} \ .$$

We therefore have an (α, β) multiplicative regret with $\alpha = (p(1-b)+b)^{1/\eta}$ and $\beta = (p(M^{-\eta} - b) + b)^{1/\eta}$. By Lemma 7.2, $LC(K,T) \leq \frac{1}{\min\{\alpha/K, \beta\}} - K$. Define $p_0 = \frac{(K^\eta - 1)b}{(K^\eta-1)b+1-(K/M)^\eta}$. Since $1 \leq K < M$, we have that $p_0 \in [0,1]$, so we may pick $p = p_0$. It can be easily verified that, given this choice, $\alpha/K = \beta$, and, therefore, $LC(K,T) \leq 1/\beta - K$, and

$$\begin{aligned}
\beta &= \left(p_0(M^{-\eta} - b) + b\right)^{\frac{1}{\eta}} = \left(\frac{K^\eta - 1 + \frac{1}{b}(1-(K/M)^\eta)}{1-M^{-\eta}}\right)^{-\frac{1}{\eta}} \\
&= \left(\frac{K^\eta - 1 + e^{\eta(\eta-1)Q}\bar{\rho}_h(\eta)(1-(K/M)^\eta)}{1-M^{-\eta}}\right)^{-\frac{1}{\eta}} \ .
\end{aligned}$$

□

For $\eta = 1$, the above bound becomes simpler:

Corollary 7.11. *If $\eta = 1$, then $LC(K,T) \leq S_0(\bar{\rho}_h(1) - 1)\frac{M-K}{M-S_0}$.*

Setting $H_t \equiv 1$, the principle used in the proof of Theorem 7.10 can be utilized to improve the upper bound for $C(K,T)$ stated in Corollary 6.6.

Theorem 7.12. *For every $\eta \in [1, \eta_{max}]$,*

$$C(K,T) \leq \left(K^\eta + S_0^\eta e^{\eta(\eta-1)Q}(1-(K/M)^\eta)\right)^{\frac{1}{\eta}} - K \ .$$

Proof. The theorem holds trivially for $K = 0$, so assume $K > 0$. Setting $H_t \equiv 1$ and denoting $b = e^{-\eta(\eta-1)Q}$, we have by Theorem 7.7 that $V_T \geq (p_{c,1} + p_{s,1}bS_T^\eta)^{1/\eta}$. This implies that $V_T \geq p_{c,1}^{1/\eta}$ and also that

$$V_T \geq S_T(p_{c,1}S_T^{-\eta} + p_{s,1}b)^{1/\eta} \geq S_T(p_{c,1}M^{-\eta} + p_{s,1}b)^{1/\eta} \ .$$

Denoting $\alpha = p_{c,1}^{1/\eta}$ and $\beta = (p_{c,1}M^{-\eta} + (1-p_{c,1})b)^{1/\eta}$, we thus have $V_T \geq \max\{\alpha, \beta S_T\}$. We will use the value $p = \frac{K^\eta b}{K^\eta b + 1 - (K/M)^\eta}$ as our choice for $p_{c,1}$. This choice is valid, since $0 < K < M$, and therefore $p \in [0,1]$. We now have that

$$\frac{K^\eta \beta^\eta}{\alpha^\eta} = \frac{K^\eta p(M^{-\eta} - b) + K^\eta b}{p} = K^\eta(M^{-\eta} - b) + K^\eta b \cdot \frac{K^\eta b + 1 - (K/M)^\eta}{K^\eta b} = 1 \ ,$$

or $\alpha = \beta K$, yielding that $V_T \geq \beta \max\{S_T, K\}$, or equivalently, that $(1/\beta)V_T - K \geq \max\{S_T - K, 0\}$. Since the r.h.s. of the last inequality is the payoff of $\mathbf{C}(K,T)$, we may obtain directly by the arbitrage-free assumption that $C(K,T) \leq 1/\beta - K$. Since

$$\begin{aligned} \frac{1}{\beta} &= \frac{K}{\alpha} = \left(\frac{K^\eta b + 1 - (K/M)^\eta}{K^\eta b}\right)^{\frac{1}{\eta}} \cdot K = \left(K^\eta + (1 - (K/M)^\eta)b^{-1}\right)^{\frac{1}{\eta}} \\ &= \left(K^\eta + (1 - (K/M)^\eta)e^{\eta(\eta-1)Q}\right)^{\frac{1}{\eta}} \ , \end{aligned}$$

the result follows (recalling that $S_0 = 1$). □

7.4 Discussion of the Bounds

Direction of trading: one-way versus two-way. The algorithm family of Section 7.3 combines regret minimization and one-way trading components. We would like to show that in general, this results in two-way trading algorithms, although we may set the parameter η so as to obtain a one-way trading algorithm. The analysis applies also to the trading algorithm of [35], which corresponds to the special case where the one-way trading component is inactive, i.e., $H_t \equiv 1$.

The dynamics of our algorithms are an interplay between the dynamics of their two components. However, whenever $H_{t+1} = H_t$, the only active element is the regret minimization component $f(\mathbf{r}_t)$, making the dynamics easier to characterize. Note that for any h, simply cashing in less frequently can guarantee that $H_{t+1} = H_t$ often, without changing the essential workings of the algorithm. For $H_t = \frac{\ln(M/\mu_{t-1})}{\ln(M/K)}$, the price-oriented rule used in Theorem 7.9, we have that $H_{t+1} = H_t$ if the stock price is not at a new all-time high. The following claim gives a characterization of the direction of trading.

Lemma 7.13. *For any $f(\mathbf{r}_t)$, if $H_{t+1} = H_t$, then the sign of $f(\mathbf{r}_t) - 1 - r_t$ determines the direction of trading at time $t+1$. If $f(\mathbf{r}_t) = 1 + \eta r_t$, then for $\eta = 1$ the algorithm* *is one-way trading; for $\eta > 1$, if $H_{t+1} = H_t$, the algorithm buys stocks when prices rise and sells stocks when prices drop.*

Proof. Denote by $N_{s,t}$ the number of stocks at time t. Since $N_{s,t} = p_{s,t}V_{t-1}/S_{t-1}$ and

$$V_t = V_{t-1}(p_{s,t}(1+r_t) + p_{c,t}) = V_{t-1}(1 + p_{s,t}r_t) \ ,$$

we have

$$\begin{aligned} \frac{N_{s,t+1}}{N_{s,t}} &= \frac{p_{s,t+1}}{p_{s,t}} \cdot \frac{V_t}{V_{t-1}} \cdot \frac{S_{t-1}}{S_t} = \frac{w_{s,t+1}}{w_{s,t}} \cdot \frac{W_t}{W_{t+1}} \cdot (1 + p_{s,t}r_t) \cdot \frac{1}{1+r_t} \\ &= \frac{H_{t+1}}{H_t} \cdot f(\mathbf{r}_t) \cdot \frac{W_t}{W_{t+1}} \cdot \frac{1 + p_{s,t}r_t}{1+r_t} \ . \end{aligned}$$

Noting that $w_{c,t+1} = w_{c,t} + w_{s,t}f(\mathbf{r}_t) - w_{s,t+1}$, we have

$$W_{t+1} = w_{c,t} + w_{s,t}f(\mathbf{r}_t) = W_t(p_{c,t} + p_{s,t}f(\mathbf{r}_t)) \ ,$$

and therefore,

$$\begin{aligned} \frac{N_{s,t+1}}{N_{s,t}} &= \frac{H_{t+1}}{H_t} \cdot f(\mathbf{r}_t) \cdot \frac{1}{p_{c,t} + p_{s,t}f(\mathbf{r}_t)} \cdot \frac{1 + p_{s,t}r_t}{1+r_t} \\ &= \frac{H_{t+1}}{H_t} \cdot \frac{1 + (f(\mathbf{r}_t) - 1)}{1 + p_{s,t}(f(\mathbf{r}_t) - 1)} \cdot \left(\frac{1+r_t}{1+p_{s,t}r_t}\right)^{-1} \\ &= \frac{H_{t+1}}{H_t} \cdot \frac{\xi(f(\mathbf{r}_t) - 1)}{\xi(r_t)} \ , \end{aligned}$$

where $\xi(x) = \frac{1+x}{1+p_{s,t}x}$. We always have $w_{c,t} > 0$, implying that $p_{s,t} < 1$, so it can be easily verified that $\xi(x)$ is increasing in x. Thus, if $H_{t+1} = H_t$, then $f(\mathbf{r}_t) > 1 + r_t$ implies $N_{s,t+1} > N_{s,t}$, $f(\mathbf{r}_t) < 1 + r_t$ implies $N_{s,t+1} < N_{s,t}$, and $f(\mathbf{r}_t) = 1 + r_t$ implies $N_{s,t+1} = N_{s,t}$. This means that for $f(\mathbf{r}_t) = 1 + \eta r_t$ with $\eta > 1$, given that $H_{t+1} = H_t$, the algorithm buys stocks iff $r_t > 0$ and sells stocks iff $r_t < 0$.

Finally, for $f(\mathbf{r}_t) = 1 + r_t$, $N_{s,t+1}/N_{s,t} = H_{t+1}/H_t \leq 1$, so the algorithm is one-way trading. □

Relation to the search and one-way trading problems. Any one-way trading algorithm h may be used in our configuration. In particular, Theorem 7.10 shows how the competitive ratio characteristics of h determine the multiplicative regret of our algorithm. In particular, it shows that improving $\rho_h(\eta)$, the competitive ratio of h

w.r.t. the sequences $\{S_t^\eta\}_{t=1}^T$, improves the multiplicative regret of the two-way trading algorithm.

A one-way trading algorithm with optimal competitive ratio is derived in [36], in a scenario where only an upper bound M and a lower bound M/φ on the stock price are known, and T trades are allowed. The optimal competitive ratio $\rho^* = \rho^*(T, \varphi)$ is defined by $(\varphi-1)(1-\rho^*/T)^T+1 = \rho^*$. Equivalently, this implies an (α, β) multiplicative regret with $\beta = 1/\rho^*$.

In our model we deviate in two important ways from their model. First, we introduce an additional parameter that bounds the quadratic variation, Q, and second, we allow for two-way trading. Therefore, it is not surprising that for certain settings we may improve on their optimal competitive ratio.

More specifically, we apply Theorem 7.10, taking h to be the optimal one-way trading algorithm, and setting $K = 1$. The theorem states that for any $\eta \in [1, \eta_{max}]$ we may attain a $(\beta(\eta), \beta(\eta))$ multiplicative regret, where $\beta(\eta) = e^{-(\eta-1)Q}\rho_h(\eta)^{-1/\eta}$. For $\eta = 1$ we have $\beta(1) = 1/\rho_h(1)$, which means that our algorithm attains the same competitive ratio as h. Thus, showing that $\beta'(1) > 0$ would imply that $\beta(\eta) > \beta(1) = 1/\rho_h(1)$ for η in some interval $(1, \eta_0)$, namely, an improvement over the competitive ratio of h. Since

$$\frac{\partial \ln \beta(\eta)}{\partial \eta} = -Q + \frac{1}{\eta^2}\ln \rho_h(\eta) - \frac{1}{\eta}\frac{\rho_h'(\eta)}{\rho_h(\eta)} ,$$

and $\ln x$ is increasing in x, we have that $\beta'(1) > 0$ if

$$\ln \rho_h(1) - \frac{\rho_h'(1)}{\rho_h(1)} > Q .$$

This last condition may be seen to hold for specific values of φ and Q (by means of numeric computation, due to the implicit definition of $\rho_h(\eta)$). In general, bigger values of φ and smaller values of Q make for bigger improvements.

Comparison to the simpler bounds. We now show that the price-oriented bound of Theorem 7.9 is superior to the bounds of the simple algorithms of Section 7.2 and also to the bound given in Theorem 6.5. For algorithms A_{TS} and A_{PS}, which are clearly one-way trading, we will show that the price-oriented bound is superior even when we set $\eta = 1$ (and thus have a one-way trading algorithm).

The bound for A_{PS} is $1 + \lfloor \log_2(M/K) \rfloor$, while, for $\eta = 1$, Theorem 7.9 gives a bound of $\ln(M/K)$. We have that

$$\ln(M/K) = \log_2(M/K) \cdot \ln 2 \approx 0.69 \log_2(M/K) < 1 + \lfloor \log_2(M/K) \rfloor \,.$$

If we make no extra assumptions on the growth of Q over time, then the bound of Theorem 7.9 is superior to the call option-based bound of A_{TC}, where we use the bound of Theorem 7.12 on the price of call options. This is summarized in the following lemma.

Lemma 7.14. *If $T > 1$, then for any $\eta \in [1, \eta_{max}]$,*

$$\left(K^\eta + \eta e^{\eta(\eta-1)Q} \ln \frac{M}{K}\right)^{\frac{1}{\eta}} - K < T\left[\left(K^\eta + e^{\eta(\eta-1)Q}\left(1 - \left(\frac{K}{M}\right)^\eta\right)\right)^{\frac{1}{\eta}} - K\right].$$

Proof. By the concavity of $\zeta(y) = y^{1/\eta}$ for $\eta \ge 1$, we have that for any $y > 0$,

$$\frac{1}{T}(K^\eta + y)^{\frac{1}{\eta}} + \frac{T-1}{T}(K^\eta)^{\frac{1}{\eta}} \le \left(\frac{1}{T}(K^\eta + y) + \frac{T-1}{T}K^\eta\right)^{\frac{1}{\eta}} = \left(K^\eta + \frac{y}{T}\right)^{\frac{1}{\eta}} ,$$

and rearranging, we have

$$(K^\eta + y)^{\frac{1}{\eta}} - K \le T \cdot \left(\left(K^\eta + \frac{y}{T}\right)^{\frac{1}{\eta}} - K\right) .$$

Picking $y = Te^{\eta(\eta-1)Q}(1 - (K/M)^\eta)$ yields

$$\left(K^\eta + Te^{\eta(\eta-1)Q}(1 - (K/M)^\eta)\right)^{\frac{1}{\eta}} - K \le T \cdot \left(\left(K^\eta + e^{\eta(\eta-1)Q}(1 - (K/M)^\eta)\right)^{\frac{1}{\eta}} - K\right) .$$

It is now enough to show that for $T > 1$ and for any valid η,

$$\left(K^\eta + \eta e^{\eta(\eta-1)Q} \ln(M/K)\right)^{\frac{1}{\eta}} - K < \left(K^\eta + Te^{\eta(\eta-1)Q}(1 - (K/M)^\eta)\right)^{\frac{1}{\eta}} - K \,, \quad (7.3)$$

or equivalently, that $\eta \ln(M/K) < T(1 - (K/M)^\eta)$. Denoting $x = (K/M)^\eta$, we have that $M^{-\eta} \le x < 1$, and defining $\xi(x) = T(1-x) + \ln x$, it is left to show that $\xi(x) > 0$ for $x \in [M^{-\eta}, 1)$. It holds that $\xi'(x) < 0$ iff $x > 1/T$, so since $\xi(1) = 0$, we have that $\xi(x) > 0$ for $x \in [1/T, 1)$. Assume then that $x \in [M^{-\eta}, 1/T)$. If $M^{-\eta} \ge 1/T$ we are done, so assume that $M^{-\eta} < 1/T$. It is left to show that $\xi(M^{-\eta}) > 0$. Since $\eta < 1/(2R)$ and by our model assumptions $Q \le R^2 T$ and $M \le \exp(\sqrt{QT})$, we have that

$$-\ln(M^{-\eta}) = \eta \ln M \leq \eta\sqrt{QT} \leq \eta RT < T/2 \; .$$

Therefore,

$$\xi(M^{-\eta}) = T(1 - M^{-\eta}) + \ln(M^{-\eta}) > T(1 - 1/T) - \frac{T}{2} \geq \frac{T}{2} - \frac{T}{2} = 0 \; ,$$

completing the proof. □

The above proof in fact yields the stronger result given in Equation 7.3, for $T > 1$ and any $\eta \in [1, \eta_{max}]$. The r.h.s. of that equation is clearly smaller than the bound given in Theorem 6.5, namely, $\left(K^\eta + Te^{\eta(\eta-1)Q}\right)^{1/\eta} - K$. The strict inequality is easily seen to hold for $T = 1$ as well, since

$$\eta \ln(M/K) \leq \eta \ln M \leq \sqrt{QT}/(2R) \leq RT/(2R) = T/2 < T \; ,$$

and thus we have:

Lemma 7.15. *For any $\eta \in [1, \eta_{max}]$, the price-oriented bound is strictly better than the bound of Theorem 6.5.*

Finally, for the special case $\eta = 1$, Lemma 7.15 implies that the price-oriented bound is strictly better than T, the bound of A_{TS}.

Comparison to the work of Dawid et al. The work of [34] priced options whose payoff is $F(M_T)$, for a class of functions F. The function $F(x) = \max\{x - K, 0\}$, which describes a fixed-strike lookback option, does not belong in that class, and therefore their results cannot be applied here. However, their methodology still suggests a one-way trading algorithm for dominating this option, and it coincides with our price-oriented rule. It is easy to verify that our $\ln(M/K)$ bound may not be improved by this approach, as follows. For the price path $r_t = (\ln M)/T$, $1 \leq t \leq T$, we have that $M_T = (1 + (\ln M)/T)^T \leq M$, and as $T \to \infty$, $|r_t| \to 0$, $\sum_{t=1}^T r_t^2 \to 0$, and $M_T \to M$, satisfying arbitrary parameters R, Q, and M. As T grows, the wealth obtained by investing a sum of $\ln(M/K)$ with the price-oriented rule alone converges to $\int_K^M x d\ln x$, which equals $M - K$. Thus, at least $\ln(M/K)$ in cash is required to dominate $\mathbf{LC}(K, T)$ with this strategy.

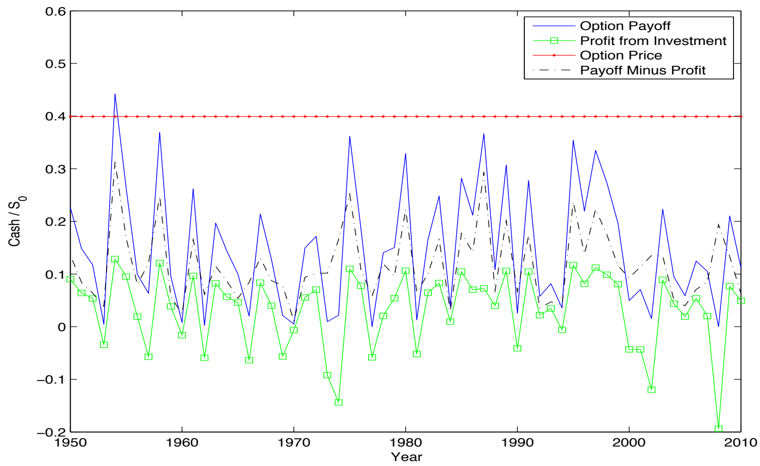

Figure 7.1: The plot gives a breakdown of the option writer's cash flow in terms of option payoff, option price, and profit from the algorithm's trading. Data is calculated for 1-year lookbacks on the S&P 500 index for the years 1950-2010, with $K = 1$, $R = 0.15$, $Q = 0.1$, and $M = 1.5$. The option writer makes a profit if the payoff minus profit line is below the option price line. (Note that the "hindsight" empirical price is 0.314 while our bound gives 0.399.) The calculated BSM price is 0.129.

7.5 Empirical Results

In order to examine our bounds empirically, we again consider the S&P 500 index data for the years 1950-2010, which was used in the empirical examination of average strike call options (Section 6.5). The results are plotted in Figure 7.1. We computed a price for a 1-year lookback using the price-oriented rule and bound (see Subsection 7.3.1), with $K = 1$, $R = 0.15$, $Q = 0.1$, and $M = 1.5$. These R, Q, and M values hold for all years but two in our test. In addition, for each year we computed the payoff of the lookback option, and ran our algorithm with the price-oriented rule and computed its profit. In our calculations we used a single value of η, which is the value that minimized our price bound. The BSM pricing [30] was computed for comparison, using the average volatility for the whole data, and assuming zero risk-free interest. Note that the stock prices for each year were normalized so that $S_0 = 1$ at the beginning of the year.

We compare the net payoff to the lookback option price. Recall that the net payoff is the difference between the payoff to the option holder (always non-negative) and the profit (or loss) the algorithm made in trading. We observe that our lookback option price dominates the net payoff for every year, with our option price at about 0.4 and the maximal net payoff at about 0.3. Note that the option payoff itself occasionally

K	1.0	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8
Opt. One-Way Trading Rule	0.602	0.541	0.481	0.421	0.361	0.301	0.241	0.180	0.120
Price-Oriented Rule	0.687	0.586	0.493	0.408	0.330	0.259	0.195	0.137	0.086
Price-Oriented Rule, $\eta = 1$	0.693	0.598	0.511	0.431	0.357	0.288	0.223	0.163	0.105

Table 7.1: A comparison of different price bounds for different values of the strike price. The first row features the competitive ratio-based bound using the optimal one-way trading rule of El-Yaniv et al. and the optimal η (always $\eta = 1$). The other rows feature the price-oriented bound, either with the optimal η or with $\eta = 1$. The settings used were $R = 0.2$, $Q = 0.2$, and $M = 2$, as well as $T = 252$ trading days and a lower bound of 0.5 on the stock price for the optimal one-way trading bound.

exceeds this maximal net payoff, highlighting the influence of the trading algorithm.

We conclude with a comparison of our two different bound derivations for different values of the strike price K (see Table 7.1). One derivation is the price-oriented bound of Theorem 7.9, and the other is the competitive ratio-based bound of Theorem 7.10. For the latter bound we use the optimal one-way trading rule of [36]. We set $R = 0.2$, $Q = 0.2$, and $M = 2$. In addition, we set $T = 252$ trading days and a lower bound of 0.5 on the stock price for the optimal one-way trading bound. For the price-oriented rule, we give both the bound with optimal η and the bound with $\eta = 1$. For the optimal one-way trading rule, $\eta = 1$ was always the best choice for this setting.

It can be seen that for the price-oriented rule, working with the optimal η is better than merely one-way trading ($\eta = 1$). For lower values of K, the optimal one-way trading rule does better, whereas for higher values of K, the price-oriented rule does better.

7.6 Appendix: Additional Claims

Lemma 7.16. *It holds that $M_T \leq (1 + \sqrt{Q/T})^T \leq \exp(\sqrt{QT})$.*

Proof. Clearly, for M_T to reach its highest possible value given R and Q we must have $r_t \geq 0$ for every $1 \leq t \leq T$, in which case $M_T = S_T$. By the inequality of the geometric, arithmetic, and quadratic means, we have:

$$\begin{aligned} S_T^{\frac{1}{T}} &= \left(\prod_{t=1}^{T}(1+r_t)\right)^{\frac{1}{T}} \leq \frac{1}{T}\sum_{t=1}^{T}(1+r_t) = 1 + \frac{1}{T}\sum_{t=1}^{T} r_t \\ &\leq 1 + \left(\frac{1}{T}\sum_{t=1}^{T} r_t^2\right)^{\frac{1}{2}} \leq 1 + \sqrt{Q/T}\ , \end{aligned}$$

hence

$$S_T \leq (1+\sqrt{Q/T})^T \leq \exp(T\sqrt{Q/T}) = \exp(\sqrt{QT}) \ ,$$

where we used the fact that $1+x \leq \exp(x)$. $\square$

Lemma 7.17. *If $z_0, \ldots, z_n > 0$, then $\sum_{i=1}^n z_i \ln(z_i/z_{i-1}) \geq z_n - z_0$.*

Proof. For every $1 \leq i \leq n$ we have that

$$z_i \ln \frac{z_i}{z_{i-1}} \geq z_i \cdot \frac{z_i/z_{i-1} - 1}{z_i/z_{i-1}} = z_i - z_{i-1} \ ,$$

where the inequality uses the fact that $\ln(1+x) \geq \frac{x}{x+1}$ for $x \in (-1, \infty)$. The result follows by simple summation. $\square$

Chapter 8

Pricing Based on Additive Regret

8.1 Relating Multiplicative Regret to Standard Regret

In Chapters 6 and 7 we priced various derivatives based on the performance guarantees of trading algorithms. These results used the intimate connection between the price of the option $\mathbf{\Psi}(\mathbf{X}_1, \ldots, \mathbf{X}_N, T)$ and the multiplicative regret guarantees of algorithms trading in the derivatives $\mathbf{X}_1, \ldots, \mathbf{X}_N$. In our analysis we depended on the multiplicative regret guarantees of the Generic algorithm as a basic regret minimization component.

The Generic algorithm is an adaptation of the Polynomial Weights algorithm [25], which was originally developed for the standard best expert setting. However, the (additive) regret bounds proved in [25] do not readily translate into multiplicative regret bounds, and required an elaborate new analysis by the authors of Generic in [35].

The Polynomial Weights Algorithm

Parameters: A learning rate $\eta > 0$ and initial weights $w_{i,1} > 0$, $1 \le i \le N$.

For each round $t = 1, \ldots, T$

1. Define probabilities $p_{i,t} = w_{i,t}/W_t$, where $W_t = \sum_{i=1}^{N} w_{i,t}$.

2. For each expert $i = 1, \ldots, N$, let $w_{i,t+1} = w_{i,t}(1 - \eta l_{i,t})$.

The case of the Polynomial Weights and Generic algorithms highlights two important points. First, algorithms for the best expert setting may be naturally interpreted as trading algorithms, by taking the probabilities of picking experts to be fractions of wealth to invest in assets. Second, results in the best expert setting do not carry over directly to the trading setting. These facts clearly motivate finding a scheme for translating additive regret bounds into multiplicative regret bounds, and thereby harnessing the power of results in the best expert setting.

For this purpose, we consider a natural representation of the trading scenario in terms of a best expert scenario. Let $\mathbf{X}_1, \ldots, \mathbf{X}_N$ be derivatives (experts), where $X_{i,t} > 0$ for every $1 \leq i \leq N$ and $1 \leq t \leq T$, and let $r_{i,t}$ be their single-period returns. For every $1 \leq i \leq N$, the initial value $X_{i,0}$ is allowed to be an arbitrary positive number, and the algorithm will trade in the fractional assets $X_{i,0}^{-1}\mathbf{X}_i$, whose initial value is 1. We will assume w.l.o.g. that the algorithm starts with one unit of cash, that is, $V_0 = 1$. We next define the single-period losses of the experts as $l_{i,t} = -\ln(1 + r_{i,t})$. This way, the cumulative losses of the experts translate directly to minus the logarithms of their fractional asset values, that is,

$$L_{i,t} = \sum_{\tau=1}^{t} l_{i,\tau} = -\sum_{\tau=1}^{t} \ln(X_{i,\tau}/X_{i,\tau-1}) = -\ln(X_{i,t}/X_{i,0}) \ .$$

However, the cumulative loss of an algorithm A for the best expert setting does not simply translate to minus the logarithm of its final wealth as a trading algorithm. We have that

$$L_{A,t} = \sum_{\tau=1}^{t} l_{A,\tau} = -\sum_{\tau=1}^{t}\sum_{i=1}^{N} p_{i,\tau} \ln(1 + r_{i,\tau}) \ ,$$

whereas

$$\begin{aligned} -\ln V_t &= -\sum_{\tau=1}^{t} \ln(V_\tau/V_{\tau-1}) = -\sum_{\tau=1}^{t} \ln \sum_{i=1}^{N} p_{i,\tau}(1 + r_{i,\tau}) \\ &= -\sum_{\tau=1}^{t} \ln \left(1 + \sum_{i=1}^{N} p_{i,\tau} r_{i,\tau}\right) \ . \end{aligned}$$

In order to relate these two different quantities, we will need the next lemma.

Lemma 8.1. *Let* $\sum_{i=1}^{N} p_i = 1$*, where* $0 \leq p_i \leq 1$ *for every* $1 \leq i \leq N$*, and let* $z_i \in (-1, \infty)$ *for every* $1 \leq i \leq N$*. Then*

$$\ln\left(1 + \sum_{i=1}^{N} p_i z_i\right) - (1/8)\ln^2\left(\frac{1 + \max_i\{z_i\}}{1 + \min_i\{z_i\}}\right) \leq \sum_{i=1}^{N} p_i \ln(1 + z_i) \ .$$

Proof. Note first that it is sufficient to prove the statement for the case where the values

z_i are distinct, and then invoke the continuity of the above expressions in $z_1, \ldots, z_N$. Consider a random variable Y, which has the value $\ln(1+z_i)$ with probability p_i, for every i. By Hoeffding's lemma (Lemma A.1), we have that

$$\begin{aligned}\ln(\mathbb{E}[e^Y]) &\leq \mathbb{E}[Y] + (1/8)(\max_i\{\ln(1+z_i)\} - \min_i\{\ln(1+z_i)\})^2 \\ &= \mathbb{E}[Y] + (1/8)\ln^2\left(\frac{1+\max_i\{z_i\}}{1+\min_i\{z_i\}}\right) .\end{aligned}$$

Since

$$\ln(\mathbb{E}[e^Y]) = \ln\left(\sum_{i=1}^N p_i e^{\ln(1+z_i)}\right) = \ln\left(1 + \sum_{i=1}^N p_i z_i\right)$$

and $\mathbb{E}[Y] = \sum_{i=1}^N p_i \ln(1+z_i)$, we have that

$$\ln\left(1+\sum_{i=1}^N p_i z_i\right) \leq \sum_{i=1}^N p_i \ln(1+z_i) + (1/8)\ln^2\left(\frac{1+\max_i\{z_i\}}{1+\min_i\{z_i\}}\right) ,$$

and, rearranging, we get the desired result. □

Before proceeding, we recall the definition of the *relative quadratic variation* q_T, given in Chapter 2, and note that for the current scenario we have that

$$q_T = \sum_{t=1}^T (\max_i\{l_{i,t}\} - \min_i\{l_{i,t}\})^2 = \sum_{t=1}^T \ln^2\left(\frac{1+\max_i\{r_{i,t}\}}{1+\min_i\{r_{i,t}\}}\right) .$$

We can now prove the following relation.

Theorem 8.2. *It holds that* $0 \leq L_{A,T} + \ln V_T \leq q_T/8$*, and as a result,*

$$0 \leq R_{A,T} + \ln\frac{V_T}{\max_i\{X_{i,0}^{-1} X_{i,T}\}} \leq q_T/8 .$$

Proof. We have that

$$\ln V_T = \sum_{t=1}^T \ln(V_t/V_{t-1}) = \sum_{t=1}^T \ln(1 + \sum_{i=1}^N p_{i,t} r_{i,t}) .$$

By the concavity of $\ln(1+z)$,

$$\sum_{t=1}^{T}\ln(1+\sum_{i=1}^{N}p_{i,t}r_{i,t}) \geq \sum_{t=1}^{T}\sum_{i=1}^{N}p_{i,t}\ln(1+r_{i,t}) = -\sum_{t=1}^{T}\sum_{i=1}^{N}p_{i,t}l_{i,t} = -L_{A,T} \ ,$$

and thus, $\ln V_T \geq -L_{A,T}$, or $0 \leq L_{A,T} + \ln V_T$, as required. For the other side, we have by Lemma 8.1 that

$$\begin{aligned} -L_{A,T} &= \sum_{t=1}^{T}\sum_{i=1}^{N}p_{i,t}\ln(1+r_{i,t}) \\ &\geq \sum_{t=1}^{T}\left[\ln(1+\sum_{i=1}^{N}p_{i,t}r_{i,t}) - (1/8)\ln^2\left(\frac{1+\max_i\{r_{i,t}\}}{1+\min_i\{r_{i,t}\}}\right)\right] \\ &= \ln V_T - q_T/8 \ , \end{aligned}$$

as needed. Since $\min_i\{L_{i,T}\} = -\ln\max_i\{X_{i,0}^{-1}X_{i,T}\}$, we have that

$$0 \leq L_{A,T} - \min_i\{L_{i,T}\} - \ln\max_i\{X_{i,0}^{-1}X_{i,T}\} + \ln V_T \leq q_T/8 \ ,$$

or equivalently,

$$0 \leq R_{A,T} + \ln\frac{V_T}{\max_i\{X_{i,0}^{-1}X_{i,T}\}} \leq q_T/8 \ ,$$

completing the proof. □

8.2 Upper Bounds on Option Prices

The result of Theorem 8.2 may be used in conjunction with Lemma 6.2 to upper bound $\Psi(\mathbf{X}_1,\ldots,\mathbf{X}_N,T)$. The only missing element now is an algorithm for the best expert setting, which provides performance guarantees. We have the following:

Theorem 8.3. *Let A be an algorithm for the best expert setting, which guarantees $L_{A,T} - L_{i,T} + \ln X_{i,0} \leq \gamma$ for some γ, for every expert i and any loss sequence. Then*

$$\Psi(\mathbf{X}_1,\ldots,\mathbf{X}_N,T) \leq e^{\gamma} \ .$$

Proof. Since $L_{i,T} = -\ln(X_{i,T}/X_{i,0})$, we have by Theorem 8.2 that

$$0 \leq L_{A,T} - L_{i,T} - \ln(X_{i,T}/X_{i,0}) + \ln V_T \ ,$$

and therefore

$$-\gamma \leq -(L_{A,T} - L_{i,T} + \ln X_{i,0}) \leq \ln V_T - \ln X_{i,T} .$$

Thus, we have that $V_T/X_{i,T} \geq e^{-\gamma}$ for every i, and Lemma 6.2 yields the result. □

Corollary 8.4. *If U is an upper bound on $R_{A,T}$, then*

$$\Psi(\mathbf{X}_1, \ldots, \mathbf{X}_N, T) \leq e^U \cdot \max_i\{X_{i,0}\} .$$

We next apply the above theorem to specific algorithms.

8.2.1 Application to the Polynomial Weights Algorithm

We begin by considering the Polynomial Weights algorithm, which was the basis for Generic. Denoting A for the Polynomial Weights algorithm run with parameters η and $w_{1,1}, \ldots, w_{N,1}$, we have the following guarantees:

Theorem 8.5. ([25]) *Assume that $l_{i,t} \leq B$ for every $t = 1, \ldots, T$ and $i = 1, \ldots, N$ for some $B > 0$. Then for any sequence of losses and any expert i, $0 < \eta \leq 1/(2B)$, and $T \geq 1$, it holds that $L_{A,T} \leq L_{i,T} - (1/\eta)\ln p_{i,1} + \eta \sum_{t=1}^T l_{i,t}^2$.*

Note that the requirement $l_{i,t} \leq B$ translates into $r_{i,t} \geq e^{-B} - 1$ for every i and t. This is a more minimal one-sided requirement compared with the one in our model, which bounds $|r_{i,t}|$.

We will now assume a known upper bound $\mathcal{Q}_{i,T}$ on $\sum_{t=1}^T l_{i,t}^2$, namely, $\mathcal{Q}_{i,T} \geq \sum_{t=1}^T \ln^2(1 + r_{i,t})$ for every i. In the context of Theorem 8.3, the guarantees of Theorem 8.5 then imply a bound γ defined by $\gamma = \max_i\{\gamma_i\}$, where

$$\gamma_i = -(1/\eta)\ln p_{i,1} + \eta \mathcal{Q}_{i,T} + \ln X_{i,0} .$$

We now proceed to choose parameters that minimize γ and thereby the upper bound on $\Psi(\mathbf{X}_1, \ldots, \mathbf{X}_N, T)$. Since $p_{1,1} + \ldots + p_{N,1} = 1$, optimizing for the values $p_{i,1}$ necessarily means equalizing the quantities γ_i. It is easy to see that in order to achieve this, one should choose, for every i, the values

$$p_{i,1} = \frac{X_{i,0}^\eta e^{\eta^2 \mathcal{Q}_{i,T}}}{\sum_{k=1}^N X_{k,0}^\eta e^{\eta^2 \mathcal{Q}_{k,T}}} ,$$

and that for this choice

$$\gamma = \gamma_1 = \ldots = \gamma_N = \ln\left(\sum_{k=1}^{N} X_{k,0}^{\eta} e^{\eta^2 Q_{k,T}}\right)^{1/\eta} .$$

We may further optimize for η, and in conjunction with Theorem 8.3 obtain the following:

Theorem 8.6. *Assuming $r_{i,t} \geq e^{-B} - 1$ for every i and t, for some $B > 0$, it holds that*

$$\Psi(\mathbf{X}_1, \ldots, \mathbf{X}_N, T) \leq \min_{0<\eta\leq 1/(2B)} \left(\sum_{k=1}^{N} X_{k,0}^{\eta} e^{\eta^2 Q_{k,T}}\right)^{1/\eta} .$$

8.2.2 Application to the Hedge Algorithm

In this subsection we prove a new performance guarantee for Hedge, which is based on the relative quadratic variation, and then apply it to upper bound option prices.[1]

Theorem 8.7. *Let A be the Hedge algorithm run with parameters $\eta > 0$ and $p_{1,0}, \ldots, p_{N,0} > 0$, where $\sum_{i=1}^{N} p_{i,0} = 1$. Then for every expert i,*

$$L_{A,T} - L_{i,T} \leq \frac{1}{\eta} \ln \frac{1}{p_{i,0}} + \frac{\eta}{8} \cdot q_T .$$

If q' is a known upper bound on q_T, then setting $\eta = \sqrt{(8/q')\ln N}$ and $p_{i,0} = 1/N$ for every i implies

$$R_{A,T} \leq \sqrt{(q'/2)\ln N} .$$

Proof. As discussed in Chapter 3, Hedge may be defined as the gradient of the concave potential $\Phi_\eta(\mathbf{L}) = -(1/\eta)\ln(\sum_{j=1}^{N} p_{j,0} e^{-\eta L_j})$. Applying Theorem 3.3 to Φ_η and setting $\mathbf{L}_0 = \mathbf{0}$ yields

$$R_{A,T} = \Phi_\eta(\mathbf{L}_T) - \min_j\{L_{j,T}\} - \frac{1}{2}\sum_{t=1}^{T} \mathbf{l}_t^\top \nabla^2\Phi(\mathbf{z}_t)\mathbf{l}_t ,$$

where $\mathbf{z}_t \in [\mathbf{L}_{t-1}, \mathbf{L}_t]$, $t = 1, \ldots, T$, or equivalently,

$$L_{A,T} = \Phi_\eta(\mathbf{L}_T) - \frac{1}{2}\sum_{t=1}^{T} \mathbf{l}_t^\top \nabla^2\Phi(\mathbf{z}_t)\mathbf{l}_t . \tag{8.1}$$

[1] We comment that a bound in the same spirit on the regret of Hedge was given in [25], but it has a worse constant than the bound presented here for the case we consider.

By definition of Φ_η it holds for every i that

$$\begin{aligned}\Phi_\eta(\mathbf{L}_T) - L_{i,T} &= -\frac{1}{\eta}\ln\left(\sum_{j=1}^{N} p_{j,0}e^{-\eta L_{j,T}}\right) + \frac{1}{\eta}\ln e^{-\eta L_{i,T}} \\ &= \frac{1}{\eta}\ln\left(\frac{\exp(-\eta L_{i,T})}{\sum_{j=1}^{N} p_{j,0}e^{-\eta L_{j,T}}}\right) = \frac{1}{\eta}\ln\frac{p_{i,T+1}}{p_{i,0}} \\ &\leq \frac{1}{\eta}\ln\frac{1}{p_{i,0}}\ .\end{aligned}$$

Together with Equation 8.1 we have that for every i,

$$L_{A,T} - L_{i,T} \leq \frac{1}{\eta}\ln\frac{1}{p_{i,0}} - \frac{1}{2}\sum_{t=1}^{T}\mathbf{l}_t^\top\nabla^2\Phi(\mathbf{z}_t)\mathbf{l}_t\ . \tag{8.2}$$

Now we bound the sum on the right hand side. By Lemma 3.14, for every t, $-\mathbf{l}_t^\top\nabla^2\Phi(\mathbf{z}_t)\mathbf{l}_t = \eta Var(Y_t)$, where Y_t is some discrete random variable that may attain only the values $l_{1,t},\ldots,l_{N,t}$. Thus, by Popoviciu's inequality (Corollary A.3),

$$-\mathbf{l}_t^\top\nabla^2\Phi(\mathbf{z}_t)\mathbf{l}_t \leq \frac{\eta}{4}\cdot(\max_j\{l_{j,t}\} - \min_j\{l_{j,t}\})^2$$

for every t, and Equation 8.2 yields

$$L_{A,T} - L_{i,T} \leq \frac{1}{\eta}\ln\frac{1}{p_{i,0}} + \frac{\eta}{8}\cdot q_T\ ,$$

as needed. Setting $p_{i,0} = 1/N$ for every i and $\eta = \sqrt{(8/q')\ln N}$ then yields

$$L_{A,T} - L_{i,T} \leq \frac{1}{\eta}\ln N + \frac{\eta}{8}\cdot q' = \sqrt{(q'/2)\ln N}\ ,$$

completing the proof. □

The above guarantees can now be used to upper bound $\Psi(\mathbf{X}_1,\ldots,\mathbf{X}_N,T)$ in a similar way to the derivation in the case of Polynomial Weights.

Theorem 8.8. *If q' is a known upper bound on q_T, then*

$$\Psi(\mathbf{X}_1,\ldots,\mathbf{X}_N,T) \leq \min_{\eta>0}\left\{\exp\left(\eta q'/8\right)\cdot\left(\sum_{k=1}^{N} X_{k,0}^{\eta}\right)^{1/\eta}\right\}\ .$$

Proof. Theorem 8.7 yields for every i that $L_{A,T} - L_{i,T} + \ln X_{i,0} \leq \gamma_i$, where

$$\gamma_i = \frac{1}{\eta} \ln \frac{1}{p_{i,0}} + \frac{\eta}{8} \cdot q' + \ln X_{i,0} \,.$$

Choosing the values

$$p_{i,0} = \frac{X_{i,0}^{\eta}}{\sum_{k=1}^{N} X_{k,0}^{\eta}}$$

obtains

$$\gamma_1 = \ldots = \gamma_N = \ln \left(\exp(\eta q'/8) \cdot \left(\sum_{k=1}^{N} X_{k,0}^{\eta} \right)^{1/\eta} \right) \,,$$

and the result follows by Theorem 8.3. □

8.3 Lower Bounds on Option Prices

Establishing lower bounds on option prices is conceptually similar to establishing upper bounds, given the arbitrage-free assumption. If a trading algorithm is guaranteed to satisfy $V_T \leq \beta \max_i\{X_{i,T}\}$ for some $\beta > 0$, then $1/\beta$ units of cash invested with it will always be dominated by the payoff of $\mathbf{\Psi}(\mathbf{X}_1, \ldots, \mathbf{X}_N, T)$, implying a lower bound of $1/\beta$ on the price of the option. If this bound did not hold, then the strategy of buying the option and investing $-1/\beta$ units of cash with the algorithm (shorting it) could guarantee a riskless profit. We thus have

Lemma 8.9. *If there exists a trading algorithm that guarantees $V_T \leq \beta \max_i\{X_{i,T}\}$, then $\Psi(\mathbf{X}_1, \ldots, \mathbf{X}_N, T) \geq 1/\beta$.*

In other words, to show lower bounds we are required to devise a regret *guaranteeing* algorithm, rather than a regret minimizing algorithm, as was the case with upper bounds. We stress that such guarantees (and lower bounds) will hold for *any* price sequence that has certain properties. This is in contrast to BSM exact pricing, where such guarantees hold only for some of those sequences.

In the context of additive regret, this problem may be solved using the results of Chapter 3. First, a lower bound on the *anytime regret* of a specific algorithm is shown. Then, we note that such an algorithm may be modified so as to "lock in" that regret from the moment the regret bound is exceeded until time T. (This method of locking in regret is used in [35].) This idea is formalized below.

Lemma 8.10. *Let λ be a lower bound on the anytime regret of some algorithm A for the best expert setting. Then there exists an algorithm for which*

$$V_T \le \exp(q_T/8 - \lambda) \cdot (\min_i\{X_{i,0}\})^{-1} \cdot \max_i\{X_{i,T}\} .$$

In addition, if all price paths are guaranteed to satisfy $q_T \le q'$, then

$$\Psi(\mathbf{X}_1, \ldots, \mathbf{X}_N, T) \ge \exp(\lambda - q'/8) \cdot \min_i\{X_{i,0}\} .$$

Proof. We will trade with a modified version of A, denoted A', defined as follows. Up until the first time t for which the regret of A is greater or equal to λ, A' is identical to A. Then, A' places all its weight on (transfers all wealth to) the best expert at time t, $m(t)$, from time $t+1$ to T. Thus,

$$R_{A',T} = L_{A',T} - L_{m(T),T} \ge L_{A',T} - L_{m(t),T} = L_{A',t} - L_{m(t),t} \ge \lambda .$$

By Theorem 8.2, we have for A' that

$$R_{A',T} + \ln \frac{V_T}{\max_i\{X_{i,0}^{-1} X_{i,T}\}} \le q_T/8$$

and therefore,

$$\begin{aligned}
V_T &\le \exp(q_T/8 - R_{A',T}) \cdot \max_i\{X_{i,0}^{-1} X_{i,T}\} \\
&\le \exp(q_T/8 - R_{A',T}) \cdot (\min_i\{X_{i,0}\})^{-1} \cdot \max_i\{X_{i,T}\} \\
&\le \exp(q_T/8 - \lambda) \cdot (\min_i\{X_{i,0}\})^{-1} \cdot \max_i\{X_{i,T}\} .
\end{aligned}$$

The second claim of the lemma now follows immediately from Lemma 8.9. □

Lower bounds on the anytime regret are provided in Chapter 3 for Hedge, assuming there are known lower and upper bound guarantees for q_T. Those results may now be used to prove the following theorem:

Theorem 8.11. *If $q_T \in [q, \gamma q]$ holds for every price path for some $\gamma \ge 1$, then*

$$\Psi(\mathbf{X}_1, \ldots, \mathbf{X}_N, T) \ge \exp(\sqrt{q}/(4N) - \gamma q/8) \cdot \min_i\{X_{i,0}\} .$$

Furthermore, for $N = 2$,

$$\Psi(\mathbf{X}_1, \mathbf{X}_2, T) \geq \exp(0.195\sqrt{q} - \gamma q/8) \cdot \min\{X_{1,0}, X_{2,0}\} .$$

Proof. By Theorem 3.19, there are specific parameter settings for Hedge for which its anytime regret is lower bounded by $\lambda = \sqrt{q}/(4N)$. Therefore, by Lemma 8.10,

$$\Psi(\mathbf{X}_1, \ldots, \mathbf{X}_N, T) \geq \exp(\lambda - \gamma q/8) \cdot \min_i\{X_{i,0}\} .$$

For $N = 2$, Theorem 3.16 provides the stronger guarantee $\lambda = 0.195\sqrt{q}$, and thus

$$\Psi(\mathbf{X}_1, \mathbf{X}_2, T) \geq \exp(0.195\sqrt{q} - \gamma q/8) \cdot \min_i\{X_{i,0}\} ,$$

as desired. □

Note that for the simple case $X_{1,0} = \ldots = X_{N,0} = 1$, it is easy to obtain a trivial lower bound of $\Psi(\mathbf{X}_1, \ldots, \mathbf{X}_N, T) \geq 1$ by considering an algorithm that initially places some fractions of its wealth on each asset and does not trade. This algorithm clearly guarantees $V_T \leq \max_i\{X_{i,T}\}$, yielding the trivial price bound by Lemma 8.9. Thus, for this case, the bounds of the above theorem are non-trivial only if the expressions inside the exponents are positive, which requires that $q < 4/(\gamma N)^2$ for a general N and $q < 2.43/\gamma^2$ for $N = 2$.

8.3.1 A Lower Bound on the Price of "at the Money" Call Options

The price of a call option satisfies $C(K, T) = \Psi(\mathbf{X}_1, \mathbf{X}_2, T) - K$, where $\mathbf{X}_1$ is the stock and $\mathbf{X}_2$ is K in cash. For these assets,

$$q_T = \sum_{t=1}^{T} \ln^2\left(\frac{1 + \max_i\{r_{i,t}\}}{1 + \min_i\{r_{i,t}\}}\right) = \sum_{t=1}^{T} \ln^2(1 + r_{1,t}) .$$

We will apply Theorem 8.11 to the pricing of a call option that is "at the money", namely, with $K = S_0$ (recall that $S_0 = 1$).[2] For that case, the theorem implies that

$$C(1, T) \geq \exp(0.195\sqrt{q} - \gamma q/8) - 1 .$$

[2]We comment that the bounds implied by Theorem 8.11 for $K \neq 1$ are less interesting, since they do not improve on bounds that are derived from the bound on $C(1, T)$ through the trivial relations $C(K, T) \geq C(1, T) + 1 - K$ for $K > 1$, and $C(K, T) \geq C(1, T)$ for $K < 1$.

To get a clearer picture, consider the specific scenario where $\gamma = 1$ (namely, $q_T = q$), and assume $q < 0.5$, implying that $q < \sqrt{q/2}$. Theorem 8.11 then yields the following:

Corollary 8.12. *If* $\gamma = 1$ *and* $q < 0.5$*, then* $C(1, T) \geq \exp(0.1\sqrt{q}) - 1 \geq 0.1\sqrt{q}$.

Note that by combining Theorem 8.7 and Corollary 8.4 we may also obtain an *upper* bound of $C(1, T) \leq \exp(\sqrt{(q/2)\ln 2}) - 1$, which for asymptotically small values of q behaves like $\sqrt{(q/2)\ln 2} \sim 0.59\sqrt{q}$. In comparison, the BSM pricing has an asymptotic value that corresponds to $\sqrt{q}/\sqrt{2\pi} \sim 0.4\sqrt{q}$ for small values of q (see Appendix B). All three expressions therefore have the same type of asymptotic behavior. The difference in the constants is expected, since our model is adversarial while the BSM model assumes a stochastic setting.

Chapter 9

Conclusion

The results presented in this thesis include improved regret minimization algorithms for specific scenarios as well as regret lower bounds that apply to widely used existing algorithms. In addition, our results comprise robust arbitrage-free price bounds for various derivatives and show how to directly translate regret bounds into price bounds.

These results demonstrate a solid link between regret minimization and arbitrage-free pricing. Namely, we show that regret bounds imply price bounds more directly and for more general derivatives and scenarios than considered before. Conversely, the consideration of pricing problems has motivated some of the "pure" regret minimization results presented, highlighting a fruitful interaction between these two different types of problems.

9.1 Summary of Results and Future Work

The main results presented in this work are detailed below, by subject. For each subject we also mention interesting directions for further research.

Lower bounds on individual sequence regret. We characterize a class of algorithms that guarantee non-negative regret for any sequence of losses in the online linear optimization setting. This class is shown to include all linear Regularized Follow the Leader regret minimization algorithms.

Furthermore, we present a lower bound on the anytime regret that depends on the quadratic variation of the loss sequence, Q_T, and the learning rate. We give a trade-off result whereby any learning rate that guarantees a regret upper bound of $O(\sqrt{Q_T})$

necessarily implies an $\Omega(\sqrt{Q_T})$ anytime regret on any sequence with quadratic variation Q_T. This result is proved for two sub-classes of algorithms that include Hedge and linear cost Online Gradient Descent.

It is not clear how tight our lower bounds on the anytime regret are. In particular, the dependence of our bound for Hedge on the number of experts could possibly be improved. This would immediately improve the lower bounds on option prices that we derive using this regret bound.

Another open question concerns the assumption we make regarding a fixed learning rate, which somewhat limits the scope of our results. It would be interesting to see how our methods and lower bounds may be altered to handle some flexibility in the choice of the learning rate.

Regret minimization for branching experts. We present algorithms and analysis for a novel regret minimization setting of branching experts, in which the set of experts may grow over time according to a tree-like structure, determined by an adversary. Our results cover both the full information and the bandit scenarios.

We show that these results may be applied to two scenarios of the standard best expert setting with full information, in which the expert set is of limited realized complexity. In one scenario the number of leaders, namely, different experts that were best at some point in time, is small. In the other, all experts remain grouped in just a few clusters in terms of their realized cumulative losses. In both cases we give optimal regret bounds that depend only on the effective number of experts, that is, the number of leaders or clusters. These bounds improve on the bounds of general-purpose algorithms such as Hedge for scenarios where the number of clusters or leaders is logarithmic in the number of experts.

However, our bound for the bandit setting is not independent of the number of experts and whether or not this dependence is avoidable remains unclear. In addition, we do not have algorithms for the few leaders and few clusters scenarios in the bandit setting. Designing such algorithms appears difficult in this setting, but there are no definitive results.

Derivative pricing. We apply a unified regret minimization framework to obtain variation-based price upper bounds for a variety of options, based on the method of [35]. These options include the exchange option, the shout option, and several types of

Asian options. These bounds are derived by considering a security whose payoff is the maximum of several derivatives.

In addition, we give robust variation-based upper bounds on the prices of convex path-independent derivatives. These price upper bounds are derived given a robust upper bound on the price of European call options.

For the particular problem of pricing lookback options, we present a new family of regret minimization algorithms that combines two algorithmic components: a regret minimization component and a one-way trading component. We translate the performance of these algorithms into explicit variation-based upper bounds on the price of fixed-strike lookback options. These algorithms, which are in general two-way trading, offer a new two-way trading solution to the problem of searching for the maximum of an online sequence. Furthermore, our methods may achieve better competitive ratios than the optimal one-way trading algorithm of [36].

We develop a general new formula that allows regret bounds to be directly converted to a financial setting. As a result, existing regret bounds may be applied without any need for algorithmic modifications or additional analysis. This method is applied to obtain new variation-based price upper bounds that are applicable in broader settings than the one considered in [35].

Finally, we give the first robust lower bound on the price of "at the money" call options with the same asymptotic behavior as the robust upper bound given in [35]. This bound is obtained by combining our lower bounds on the individual sequence anytime regret of Hedge with our method for converting additive regret bounds to the financial setting. This price bound has the same asymptotic behavior as the Black-Scholes-Merton price, even though our assumptions are minimal and adversarial, rather than stochastic.

Our pricing results are obtained in an idealized model of the market. Perhaps the most important aspect of this idealization is the absence of transaction costs when buying or selling securities. Transaction costs lower the returns of online trading algorithms, increasing the regret of such algorithms w.r.t. any fixed security, and thereby increasing upper bounds on option prices. Incorporating transaction costs in our pricing results in a non-trivial way would be an important robust pricing result, and is likely to have "pure" regret minimization implications as well.

Another aspect that is absent in our results is the use of loans and short selling by the trading algorithms. Short selling is inherent in the arbitrage-free assumption, which is part of the model we use. However, our application of online algorithms to pricing does not utilize short selling. It is possible that algorithms that do employ short selling may be used to provide improved price bounds.

Appendix A

Additional Claims

Lemma A.1. (Hoeffding's lemma, [56]) *Let $s \in \mathbb{R}$, and let X be a random variable such that $a \leq X \leq b$, for some $a, b \in \mathbb{R}$. Then*

$$\ln(\mathbb{E}[e^{sX}]) \leq s\mathbb{E}[X] + s^2(b-a)^2/8 \ .$$

Theorem A.2. (The Bhatia-Davis inequality, [10]) *If X is a bounded random variable with values in $[m, M]$, then $Var(X) \leq (M - \mathbb{E}[X])(\mathbb{E}[X] - m)$, with equality iff $\mathbb{P}(X \in \{m, M\}) = 1$.*

Proof. We have that

$$\begin{aligned}
(M - \mathbb{E}[X])(\mathbb{E}[X] - m) &= M\mathbb{E}[X] - mM - \mathbb{E}[X]^2 + m\mathbb{E}[X] \\
&= \mathbb{E}[MX + mX - mM] - \mathbb{E}[X^2] + \mathbb{E}[X^2] - \mathbb{E}[X]^2 \\
&= \mathbb{E}[MX + mX - mM - X^2] + Var(X) \\
&= \mathbb{E}[(M - X)(X - m)] + Var(X) \\
&\geq Var(X) \ ,
\end{aligned}$$

where the inequality is true because $(M - X)(X - m) \geq 0$. We get an equality iff $\mathbb{P}((M - X)(X - m) = 0) = 1$, which happens iff $\mathbb{P}(X \in \{m, M\}) = 1$. $\square$

Corollary A.3. (Popoviciu's inequality, [74]) *If X is a bounded random variable with values in $[m, M]$, then $Var(X) \leq (M - m)^2/4$, with equality iff $\mathbb{P}(X = M) = \mathbb{P}(X = m) = 1/2$.*

Proof. The expression $(M - x)(x - m)$ reaches a maximal value of $(M - m)^2/4$ iff

$x = (M+m)/2$. Therefore, by the Bhatia-Davis inequality,

$$Var(X) \leq (M - \mathbb{E}[X])(\mathbb{E}[X] - m) \leq (M-m)^2/4 \ .$$

We get an equality iff $\mathbb{P}(X \in \{m, M\}) = 1$ and also $\mathbb{E}[X] = (M+m)/2$, which is equivalent to the condition $\mathbb{P}(X = M) = \mathbb{P}(X = m) = 1/2$. $\square$

Appendix B

The Black-Scholes Formula

Denote S_0 for the initial stock price, T for the expiration time, K for the strike price, r for the risk-free interest rate, and σ for the volatility. Denote also

$$d_1 = \frac{\ln(S_0/K) + (r + \sigma^2/2) \cdot T}{\sigma\sqrt{T}} \ ,$$

and $d_2 = d_1 - \sigma\sqrt{T}$. Then the Black-Scholes value for $C(K,T)$ is

$$C(K,T) = \Phi(d_1) \cdot S_0 - \Phi(d_2) K e^{-rT} \ ,$$

where $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{z^2}{2}} dz$ (not to be confused with potential functions Φ). In our case, $r = 0$ and $S_0 = 1$, therefore

$$\begin{aligned} C(K,T) &= \Phi(d_1) - \Phi(d_1 - \sigma\sqrt{T})K \\ &= \Phi\left(\frac{-\ln(K) + \sigma^2 T/2}{\sigma\sqrt{T}}\right) - \Phi\left(\frac{-\ln(K) - \sigma^2 T/2}{\sigma\sqrt{T}}\right) K \ . \end{aligned}$$

For $K = 1$, we get the simpler expression

$$C(1,T) = \Phi\left(\frac{\sigma\sqrt{T}}{2}\right) - \Phi\left(-\frac{\sigma\sqrt{T}}{2}\right) \ .$$

For $x = \frac{\sigma\sqrt{T}}{2} \to 0$, we may use the first order approximation $\Phi(x) \approx \frac{1}{2} + \Phi'(x)x$. Thus,

$$\begin{aligned}
C(1,T) &= \Phi(x) - \Phi(-x) \approx \frac{1}{2} + \Phi'(x)x - \frac{1}{2} - \Phi'(-x)(-x) \\
&= (\Phi'(x) + \Phi'(-x))x = 2\Phi'(x)x \\
&= \frac{2}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} x \approx \frac{2}{\sqrt{2\pi}} \left(1 - \frac{x^2}{2}\right) x \\
&\approx \frac{2x}{\sqrt{2\pi}} = \frac{\sigma\sqrt{T}}{\sqrt{2\pi}} .
\end{aligned}$$

Bibliography

[1] J. Abernethy, P. L. Bartlett, and E. Hazan. Blackwell approachability and no-regret learning are equivalent. *Journal of Machine Learning Research - Proceedings Track*, 19:27–46, 2011.

[2] J. Abernethy, R. M. Frongillo, and A. Wibisono. Minimax option pricing meets black-scholes in the limit. In *Proceedings of the 44th Symposium on Theory of Computing*, pages 1029–1040, 2012.

[3] A. Agarwal, E. Hazan, S. Kale, and R. E. Schapire. Algorithms for portfolio management based on the newton method. In *Proceedings of the 23rd International Conference on Machine Learning*, pages 9–16, 2006.

[4] J.-Y. Audibert and S. Bubeck. Minimax policies for adversarial and stochastic bandits. In *Proceedings of the 22nd Annual Conference on Learning Theory*, pages 217–226.

[5] J.-Y. Audibert and S. Bubeck. Regret bounds and minimax policies under partial monitoring. *Journal of Machine Learning Research*, 11:2785–2836, 2010.

[6] J.-Y. Audibert, S. Bubeck, and G. Lugosi. Minimax policies for combinatorial prediction games. *Journal of Machine Learning Research - Proceedings Track*, 19: 107–132, 2011.

[7] P. Auer, N. Cesa-Bianchi, and P. Fischer. Finite-time analysis of the multiarmed bandit problem. *Machine Learning*, 47(2-3):235–256, 2002.

[8] P. Auer, N. Cesa-Bianchi, Y. Freund, and R. E. Schapire. The nonstochastic multiarmed bandit problem. *SIAM Journal on Computing*, 32(1):48–77, 2002.

[9] P. Auer, N. Cesa-Bianchi, and C. Gentile. Adaptive and self-confident on-line learning algorithms. *Journal of Computer and System Sciences*, 64(1):48–75, 2002.

[10] R. Bhatia and C. Davis. A better bound on the variance. *The American Mathematical Monthly*, 107(4):353–357, 2000.

[11] F. Black and M. Scholes. The pricing of options and corporate liabilities. *Journal of Political Economy*, 81(3):637–654, 1973.

[12] D. Blackwell. An analog of the minimax theorem for vector payoffs. *Pacific Journal of Mathematics*, 6(1):1–8, 1956.

[13] A. Blum and C. Burch. On-line learning and the metrical task system problem. *Machine Learning*, 39(1):35–58, 2000.

[14] A. Blum and A. Kalai. Universal portfolios with and without transaction costs. *Machine Learning*, 35(3):193–205, 1999.

[15] A. Blum and Y. Mansour. From external to internal regret. *Journal of Machine Learning Research*, 8:1307–1324, 2007.

[16] A. Borodin and R. El-Yaniv. *Online Computation and Competitive Analysis*, volume 53. Cambridge University Press, 1998.

[17] A. Borodin, R. El-Yaniv, and V. Gogan. Can we learn to beat the best stock. *Journal of Artificial Intelligence Research*, 21:579–594, 2004.

[18] O. Bousquet and M. Warmuth. Tracking a small set of experts by mixing past posteriors. *Journal of Machine Learning Research*, 3:363–396, 2002.

[19] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[20] D. T. Breeden and R. H. Litzenberger. Prices of state-contingent claims implicit in option prices. *Journal of Business*, 51(4):621–651, 1978.

[21] S. Bubeck and N. Cesa-Bianchi. Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Foundations and Trends in Machine Learning*, 5 (1):1–122, 2012.

[22] N. Buchbinder, S. Chen, J. Naor, and O. Shamir. Unified algorithms for online learning and competitive analysis. *Journal of Machine Learning Research - Proceedings Track*, 23:5.1–5.18, 2012.

[23] N. Cesa-Bianchi and G. Lugosi. *Prediction, Learning, and Games*. Cambridge University Press, 2006.

[24] N. Cesa-Bianchi, Y. Freund, D. Haussler, D. P. Helmbold, R. E. Schapire, and M. K. Warmuth. How to use expert advice. *Journal of the ACM*, 44(3):427–485, 1997.

[25] N. Cesa-Bianchi, Y. Mansour, and G. Stoltz. Improved second-order bounds for prediction with expert advice. *Machine Learning*, 66(2-3):321–352, 2007.

[26] K. Chaudhuri, Y. Freund, and D. Hsu. A parameter-free hedging algorithm. In *Advances in Neural Information Processing Systems 22*, pages 297–305, 2009.

[27] A. V. Chernov and V. Vovk. Prediction with advice of unknown number of experts. In *Proceedings of the 26th Conference on Uncertainty in Artificial Intelligence*, pages 117–125, 2010.

[28] C.-K. Chiang, T. Yang, C.-J. Lee, M. Mahdavi, C.-J. Lu, R. Jin, and S. Zhu. Online optimization with gradual variations. *Journal of Machine Learning Research - Proceedings Track*, 23:6.1–6.20, 2012.

[29] R. Cont and P. Tankov. *Financial Modelling with Jump Processes*, volume 133. Chapman & Hall/CRC, 2004.

[30] A. Conze and Viswanathan. Path dependent options: the case of lookback options. *Journal of Finance*, 46(5):1893–1907, 1991.

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

[32] T. M. Cover and E. Ordentlich. Universal portfolios with side information. *IEEE Transactions on Information Theory*, 42(2):348–363, 1996.

[33] A. P. Dawid, S. de Rooij, P. Grunwald, W. M. Koolen, G. Shafer, A. Shen, N. Vereshchagin, and V. Vovk. Probability-free pricing of adjusted american lookbacks. *arXiv preprint arXiv:1108.4113*, 2011.

[34] A. P. Dawid, S. de Rooij, G. Shafer, A. Shen, N. Vereshchagin, and V. Vovk. Insuring against loss of evidence in game-theoretic probability. *Statistics and Probability Letters*, 81(1):157–162, 2011.

[35] P. DeMarzo, I. Kremer, and Y. Mansour. Online trading algorithms and robust option pricing. In *Proceedings of the 38th Annual ACM Symposium on Theory of Computing*, pages 477–486, 2006.

[36] R. El-Yaniv, A. Fiat, R. M. Karp, and G. Turpin. Optimal search and one-way trading online algorithms. *Algorithmica*, 30(1):101–139, 2001. A preliminary version appeared in FOCS, 1992.

[37] E. Even-Dar, M. Kearns, Y. Mansour, and J. Wortman. Regret to the best vs. regret to the average. *Machine Learning*, 72(1-2):21–37, 2008.

[38] Y. Freund and R. E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55 (1):119–139, 1997.

[39] Y. Freund, R. Schapire, Y. Singer, and M. Warmuth. Using and combining predictors that specialize. In *Proceedings of the 29th Annual ACM Symposium on the Theory of Computing*, pages 334–343, 1997.

[40] E. Gofer and Y. Mansour. Regret minimization algorithms for pricing lookback options. In *Algorithmic Learning Theory*, pages 234–248, 2011.

[41] E. Gofer and Y. Mansour. Pricing exotic derivatives using regret minimization. In *Algorithmic Game Theory*, pages 266–277, 2011.

[42] E. Gofer and Y. Mansour. Lower bounds on individual sequence regret. In *Algorithmic Learning Theory*, pages 275–289, 2012.

[43] E. Gofer, N. Cesa-Bianchi, C. Gentile, and Y. Mansour. Regret minimization for branching experts. *Journal of Machine Learning Research - Proceedings Track*, 30: 618–638, 2013.

[44] L. Györfi, G. Lugosi, and F. Udina. Nonparametric kernel-based sequential investment strategies. *Mathematical Finance*, 16(2):337–357, 2006.

[45] J. Hannan. Approximation to Bayes risk in repeated play. In M. Dresher, A. Tucker, and P. Wolfe, editors, *Contributions to the Theory of Games*, volume 3, pages 97–139. Princeton University Press, 1957.

[46] E. Hazan. *Efficient algorithms for online convex optimization and their applications*. PhD thesis, Princeton University, 2006.

[47] E. Hazan. The convex optimization approach to regret minimization. In S. Sra, S. Nowozin, and S. J. Wright, editors, *Optimization for Machine Learning*. MIT Press, 2011.

[48] E. Hazan and S. Kale. On stochastic and worst-case models for investing. In *Advances in Neural Information Processing Systems 22*, pages 709–717, 2009.

[49] E. Hazan and S. Kale. Extracting certainty from uncertainty: regret bounded by variation in costs. *Machine Learning*, 80(2-3):165–188, 2010.

[50] E. Hazan and S. Kale. Better algorithms for benign bandits. *Journal of Machine Learning Research*, 12:1287–1311, 2011.

[51] E. Hazan, A. Agarwal, and S. Kale. Logarithmic regret algorithms for online convex optimization. *Machine Learning*, 69(2-3):169–192, 2007.

[52] D. P. Helmbold, R. E. Schapire, Y. Singer, and M. K. Warmuth. On-line portfolio selection using multiplicative updates. *Mathematical Finance*, 8(4):325–347, 1998.

[53] M. Herbster and M. Warmuth. Tracking the best expert. *Machine Learning*, 32 (2):151–178, 1998.

[54] D. Hobson. The Skorokhod embedding problem and model-independent bounds for option prices. In *Paris-Princeton Lectures on Mathematical Finance 2010*, pages 267–318. Springer, 2011.

[55] D. G. Hobson. Robust hedging of the lookback option. *Finance and Stochastics*, 2(4):329–347, 1998.

[56] W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American statistical association*, 58(301):13–30, 1963.

[57] J. C. Hull. *Options, Futures, and Other Derivatives*. Prentice-Hall, sixth edition, 2006.

[58] A. Kalai and S. Vempala. Efficient algorithms for universal portfolios. *Journal of Machine Learning Research*, 3:423–440, 2002.

[59] A. Kalai and S. Vempala. Efficient algorithms for online decision problems. *Journal of Computer and System Sciences*, 71(3):291–307, 2005.

[60] R. Kleinberg, A. Niculescu-Mizil, and Y. Sharma. Regret bounds for sleeping experts and bandits. *Machine Learning*, 80(2-3):245–272, 2010.

[61] W. Koolen, D. Adamskiy, and M. Warmuth. Putting Bayes to sleep. In *Advances in Neural Information Processing Systems 25*, pages 135–143, 2012.

[62] W. M. Koolen and S. de Rooij. Switching investments. *Theoretical Computer Science*, 473:61–76, 2013.

[63] W. M. Koolen and V. Vovk. Buy low, sell high. In *Algorithmic Learning Theory*, pages 335–349, 2012.

[64] T. L. Lai and H. Robbins. Asymptotically efficient adaptive allocation rules. *Advances in applied mathematics*, 6(1):4–22, 1985.

[65] N. Littlestone and M. Warmuth. The weighted majority algorithm. *Information and Computation*, 108:212–261, 1994.

[66] J. Lorenz, K. Panagiotou, and A. Steger. Optimal algorithms for k-search with application in option pricing. *Algorithmica*, 55(2):311–328, 2009.

[67] B. Mandelbrot. The variation of certain speculative prices. *Journal of Business*, 36(4):394–419, 1963.

[68] R. C. Merton. Theory of rational option pricing. *Bell Journal of Economics and Management Science*, 4(1):141–183, 1973.

[69] R. C. Merton. Option pricing when underlying stock returns are discontinuous. *Journal of financial economics*, 3(1):125–144, 1976.

[70] A. Nemirovski and D. Yudin. *Problem Complexity and Method Efficiency in Optimization*. Wiley, 1983.

[71] Y. Nesterov. *Introductory Lectures on Convex Optimization: A Basic Course*. Kluwer Academic Publishers, 2004.

[72] N. Nisan, T. Roughgarden, E. Tardos, and V. V. Vazirani. *Algorithmic Game Theory*. Cambridge University Press, 2007.

[73] E. Ordentlich and T. M. Cover. The cost of achieving the best portfolio in hindsight. *Mathematics of Operations Research*, 23(4):960–982, 1998.

[74] T. Popoviciu. Sur les équations algébriques ayant toutes leurs racines réelles. *Mathematica (Cluj)*, 9:129–145, 1935.

[75] A. Rakhlin, O. Shamir, and K. Sridharan. Relax and localize: From value to algorithms. *arXiv preprint arXiv:1204.0870*, 2012.

[76] H. Robbins. Some aspects of the sequential design of experiments. *Bulletin of the American Mathematical Society*, 58(5):527–535, 1952.

[77] R. T. Rockafellar. *Convex Analysis*. Princeton University Press, 1970.

[78] B. Russell. *The Problems of Philosophy*. Oxford University Press, 1912.

[79] W. Schoutens. *Lévy Processes in Finance: Pricing Financial Derivatives*. Wiley, 2003.

[80] G. Shafer and V. Vovk. *Probability and Finance: It's Only a Game!* John Wiley & Sons, 2001.

[81] S. Shalev-Shwartz. *Online Learning: Theory, Algorithms, and Applications*. PhD thesis, The Hebrew University, 2007.

[82] S. Shalev-Shwartz and Y. Singer. A primal-dual perspective of online learning algorithms. *Machine Learning*, 69(2-3):115–142, 2007.

[83] C. R. Shalizi, A. Z. Jacobs, and A. Clauset. Adapting to non-stationarity with growing expert ensembles. *arXiv preprint arXiv:1103.0949*, 2011.

[84] Y. Singer. Switching portfolios. In *Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence*, pages 488–495, 1998.

[85] D. Sleator and R. E. Tarjan. Amortized efficiency of list update and paging rules. *Communications of the ACM*, 28(2):202–208, 1985.

[86] N. N. Taleb. *The Black Swan: The Impact of the Highly Improbable.* Penguin Books, 2008.

[87] W. R. Thompson. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, 25(3/4):285–294, 1933.

[88] J. von Szökefalvi Nagy. Über algebraische gleichungen mit lauter reellen wurzeln. *Jahresbericht der Deutschen Mathematiker-Vereinigung*, 27:37–43, 1918.

[89] V. Vovk. Aggregating strategies. In *Proceedings of the 3rd Annual Workshop on Computational Learning Theory*, pages 371–383, 1990.

[90] V. Vovk. A game of prediction with expert advice. *Journal of Computer and System Sciences*, 56(2):153–173, 1998.

[91] V. Vovk and C. Watkins. Universal portfolio selection. In *Proceedings of the 11th Annual Conference on Computational Learning Theory*, pages 12–23, 1998.

[92] M. Zinkevich. Online convex programming and generalized infinitesimal gradient ascent. In *ICML*, pages 928–936, 2003.

TEL AVIV UNIVERSITY אוניברסיטת תל-אביב

הפקולטה למדעים מדויקים ע״ש ריימונד ובברלי סאקלר

בית הספר למדעי המחשב ע״ש בלבטניק

אלגוריתמי למידה חישובית
עם יישומים במימון

חיבור לשם קבלת תואר ״דוקטור לפילוסופיה״

מאת

איל גופר

בהנחייתו של

פרופסור ישי מנצור

הוגש לסנאט של אוניברסיטת תל-אביב

מרץ 2014

תמצית

קבלת החלטות ולמידה מקוונות הן מרכיב מרכזי בבעיות רבות ומגוונות. ההחלטות עשויות להיות כרוכות במסחר במניות, שיבוץ פרסומות, ניתוב, בחירת היוריסטיקות, או ביצוע מהלכים במשחק. המצבים הללו שונים אלה מאלה גם ברמת הסיבוך של סביבת ההחלטה או היריב, סוג ההיזון החוזר וטבען של ההחלטות האפשריות. באופן מפתיע, פותחו בעשורים האחרונים במסגרת תורת הלמידה המקוונת אלגוריתמים המסוגלים להתמודד עם מגוון בעיות עשיר זה. אלגוריתמים אלה הם בעלי שתי תכונות רצויות. ראשית, הנחות המוצא שלהם על סביבת הלמידה הן מינימליות, ועל פי רוב הם אף מתוכננים להתמודד מול יריב, תכונה ההופכת אותם לגמישים ועמידים במיוחד. שנית, מידת איכות ההחלטות שלהם מתכנסת תמיד לזו של האסטרטגיה הטובה ביותר מתוך קבוצת אסטרטגיות מסוימת, המשמשת קנה מידה להשוואה. תכונה אחרונה זו נקראת מזעור חרטה.

עבודה זו עוסקת הן בתיאוריה הכללית של מזעור חרטה והן בהשלכות שלה בתחום תמחור הנגזרים הפיננסיים.

תרומה אחת לתיאוריה של מזעור חרטה היא תוצאה המראה שלחלק מאלגוריתמי מזעור החרטה החשובים ביותר יש בו זמנית גם תכונה הפוכה, לפיה הם בהכרח מבטיחים קיום רמת חרטה אי-שלילית ואף חיובית בכל תסריט אפשרי. תרומה אחרת כוללת פיתוח אלגוריתמי מזעור חרטה לנסיבות בהן קבוצת האסטרטגיות להשוואה היא בעלת רמה גבוהה של יתירות; מקרים אלה מטופלים במסגרת מודל המאפשר לאסטרטגיות להתפצל באופן דינאמי.

התרומות לתחום תמחור הנגזרים מתבססות על רדוקציה מבעיית התמחור לבעיה של מציאת חסמים לחרטה של אלגוריתמי מסחר. תרומות אלה כוללות חסמים מבוססי מזעור חרטה למחירים של מגוון נגזרים פיננסיים, תוך שימוש הן באלגוריתמים קיימים והן באלגוריתמים שפותחו במיוחד. יתרה מכך, באמצעות פיתוח שיטה ישירה לתרגום חסמים על איכות הביצועים של אלגוריתמי מזעור חרטה כלליים לחסמים על ביצועיהם כאלגוריתמי מסחר, מוכחים חסמים עליונים ותחתונים חדשים על מחירי נגזרים.

תקציר

מבוא

עבודה זו עוסקת בשני נושאים עיקריים: נושא אחד הוא התיאוריה הכללית של אלגוריתמי למידה מקוונים ומזעור חרטה, והנושא השני הוא יישומים של אלגוריתמים אלה בתחום תמחור הנגזרים הפיננסיים. מבוא זה יסקור בקצרה את האלמנטים הבסיסיים של שני נושאים אלה.

למידה מקוונת

למידה מקוונת היא ענף חשוב בתחום הלמידה החישובית המודרנית, ורעיונותיו העיקריים הופיעו לראשונה בעבודותיהם של האנאן ובלקוול בשנות החמישים. בתסריט למידה מקוונת אופייני מתקיים משחק בכמה סיבובים בין אלגוריתם לומד לבין יריב. לכל אחד מהשחקנים יש קבוצת פעולות (או החלטות) מותרות, ובכל סיבוב שני השחקנים בוחרים את פעולותיהם בו-זמנית. לאחר שהפעולות נבחרות, הלומד סופג **הפסד**, שהוא פונקציה קבועה של זוג הפעולות. בסיום המשחק, ניתן להשוות את סכום ההפסדים של הלומד לסכום ההפסדים שהיה סופג אילו בחר פעולה קבועה מסוימת במהלך המשחק. ה**חרטה** של הלומד מוגדרת כהפרש בין סכום הפסדיו לבין סכום ההפסדים של הפעולה הקבועה הטובה ביותר. מטרת הלומד היא להבטיח חרטה מועטה ללא תלות בפעולות שבחר היריב. המהות המדויקת של החסמים שניתן להבטיח ביחס לחרטה תלויים בפרטי המשחק. למרות זאת, ביחס למחלקות גדולות של משחקים, באפשרותו של הלומד להבטיח שהחרטה הממוצעת לסיבוב משחק תשאף לאפס ככל שאורך המשחק שואף לאינסוף, מצב המוגדר כ**למידה ללא חרטה**.

מודל המומחה הטוב ביותר

התכונות של משחק למידה מקוונת תלויים בפרטי קבוצות ההחלטות האפשריות, פונקצית ההפסד, הטבע המדויק של המידע המתקבל על-ידי הלומד בכל סיבוב, והגבלות אפשריות נוספות.

מודל **המומחה הטוב ביותר** הוא אולי מודל המשחק המקוון הנחקר ביותר. במודל זה, היריב בוחר וקטור של מספרים חסומים, והלומד בוחר וקטור הסתברות באורך זהה. ההפסד של הלומד מוגדר כמכפלה הסקלארית של שני הוקטורים. ניתן לראות את הוקטור של היריב כמחירים של קבלת העצות של שורת מומחים (עליהם יש ליריב שליטה). הבחירה של הלומד ניתנת לפירוש כבחירה אקראית באחד המומחים. המושג 'מומחים' עשוי להתייחס להיוריסטיקות שונות, דרכים אפשריות לנסיעה עבודה, פרסומות שונות לשיבוץ באתר מרשתת, וכיוצא באלה.

האלגוריתם הידוע ביותר למודל זה הוא אלגוריתם Hedge או Randomized Weighted Majority (Vovk, 1990; Littlestone and Warmuth, 1994; Freund and Schapire, 1997). אלגוריתם זה נותן לכל מומחה משקל היורד מעריכית עם סכום הפסדיו, ולאחר מכן מנרמל את המשקלים לקבלת ערכי הסתברות. ניתן לשלוט על קצב הירידה המעריכית הזו באמצעות פרמטר מספרי, הקרוי **קצב הלימוד**. סכמת משקול זו ניתנת למימוש באמצעות עדכון כפלי המתבצע בכל סיבוב ויורד מעריכית עם ההפסד בסיבוב האחרון. אלגוריתם זה משיג חרטה החסומה על-ידי $O\left(\sqrt{T \ln N}\right)$, כאשר T הוא אורך המשחק, N הוא מספר המומחים, וקצב הלימוד נבחר כפונקציה מתאימה של שניהם. חסם זה הוא אופטימלי, שכן

כל לומד מקוון משיג סדר גודל דומה של חרטה בתוחלת כנגד הפסדים, הנבחרים כרצף אקראי לחלוטין של משתני ברנולי. חסמים אלה, התלויים באורך המשחק, נקראים חסמים מסדר אפס.

חסמים מסדר אפס מתעלמים לחלוטין ממאפייני סדרת ההפסדים. בפרט, ניתן לבחון מקרים שבהם סכום ההפסדים של המומחה הטוב ביותר הוא קטן. מסתבר שבאמצעות בחירת קצב לימוד שונה, ניתן להשיג חסם על החרטה, שבו אורך המשחק מוחלף בסכום ההפסדים של המומחה הטוב היותר, וידוע גם כחסם מסדר ראשון.

בחסמים נוספים, הידועים כחסמים מסדר שני, מוחלף אורך המשחק בגודל המודד את הווריאביליות של סדרת ההפסדים. התוצאה הראשונה מסוג זה הוכחה עבור אלגוריתם Polynomial Weights או Prod (Cesa-Bianchi et al., 2007), שהוא מודיפיקציה קטנה אך משמעותית של Hedge. ספציפית, גורם העדכון הכפלי ב-Hedge מוחלף באלגוריתם Polynomial Weights בקירוב טיילור הלינארי שלו. בחסם החרטה של אלגוריתם זה מוחלף אורך המשחק בהשתנות הריבועית המקסימלית של מומחה כלשהו, כשההשתנות הריבועית מוגדרת כסכום ריבועי ההפסדים.

אופטימיזציה קמורה מקוונת

המודל של **אופטימיזציה קמורה מקוונת** (Zinkevich, 2003),עוסק בבעיות קבלת החלטות סדרתיות שבהן הלומד בוחר נקודה (החלטה) מתוך קבוצה קומפקטית וקמורה במרחב האוקלידי, וההפסד שלו הוא הערך של פונקציה קמורה, הנבחרת על-ידי היריב, בנקודה שבחר. החרטה של הלומד נמדדת ביחס להחלטה הקבועה הטובה ביותר. מודל זה הוא הכללה של מודל המומחה הטוב ביותר, שבו ההחלטות הן וקטורי הסתברות ופונקציות ההפסד הן לינאריות. הוא מקיף גם סוגים נוספים של בעיות. דוגמא אחת היא בעיית בחירת תיק השקעות, שבה הלומד מחליט בכל סיבוב כיצד לחלק את הונו בין נכסים שונים. בדוגמא זו, ההחלטה היא עדיין וקטור הסתברות, אך פונקצית ההפסד היא לוגריתמית ולא לינארית. דוגמא אחרת היא הבעיה של בחירת נתיב מקוונת, בה נהג בוחר בכל יום באיזו דרך לנסוע לעבודה, והפסדו נמדד בזמן שהוא מבזבז בגין קטעי הכביש שבחר. כאן ההחלטות הן וקטורים עם ערכי 0/1, המציינים אם קטע כביש נבחר או לא, ולא וקטורי הסתברות. למרות שניתן לייצג בעיה זו במודל המומחה הטוב ביותר, כאשר כל צירוף של קטעי כביש נחשב כמומחה, הרי שייצוג זה דורש מספר מעריכי של מומחים. חשוב לציין שהמודל הכללי יותר של Zinkevich מאפשר ליישם כלים חזקים מתחום האופטימיזציה הקמורה בפיתוח אלגוריתמי מזעור חרטה.

בעבודה זו נתעניין במיוחד בתת-המודל שבו פונקציות היריב הן לינאריות, הקרוי **אופטימיזציה לינארית מקוונת**. ניתן להשיג למידה ללא חרטה במודל של אופטימיזציה לינארית מקוונת באמצעות אלגוריתם Regularized Follow the Leader (RFTL). אלגוריתם זה הוא עידון של האלגוריתם החמדני הבוחר בכל סיבוב את ההחלטה שהייתה ממזערת את ההפסד המצטבר עד לאותו רגע, או במילים אחרות, אלגוריתם העוקב אחרי המוביל. לעומתו, אלגוריתם RFTL בוחר את ההחלטה הממזערת את ההפסד עד כה בתוספת גורם רגולריזציה, שהוא פונקציה קמורה חזק של ההחלטה, המחולק בפרמטר מספרי, קצב הלימוד, הקובע את מידת השפעת הרגולריזציה. אלגוריתם זה מכליל הן את Hedge, והן את אלגוריתם Online Gradient Descent, שהוצג על-ידי Zinkevich, בבחירה מתאימה של פונקציות הרגולריזציה.

נציין כי חסמי החרטה הבסיסיים ביותר עבור RFTL במודל האופטימיזציה הלינארית המקוונת תלויים בשורש הריבועי של משך המשחק. הקבועים תלויים בקוטר של קבוצת ההחלטות ובתכונות של פונקציות ההפסד והרגולריזציה. תלות זו היא אופטימלית, כפי שניתן להוכיח, בדומה למודל המומחה הטוב ביותר, באמצעות בניית יריבים הבוחרים הפסדים אקראיים לחלוטין. נעיר בקצרה כי בדומה למודל המומחה הטוב ביותר, גם במודל זה פותחו חסמי חרטה מסדר שני, בהם מוחלף אורך המשחק במדדים שונים של הווריאביליות של סדרת ההפסדים.

היבטים נוספים של המשחק בין הלומד ליריב

בכל המודלים שתוארו לעיל קיימת אבחנה חשובה ביחס להיזון החוזר שמקבל הלומד בכל סיבוב. למשל, בהינתן קבוצת מומחים, החוזים את מזג האוויר, ללומד יש גישה להפסדים של כל המומחים, כלומר **מידע מלא**. בניגוד לכך, במקרה של שיבוץ פרסומות או בחירה מתוך אוסף מכונות מזל, הלומד מודע רק להפסד של הפעולה שבחר. סוג היזון חוזר זה ידוע כמודל **השודד מרובה הזרועות**, ונקרא כך בהתייחס לסוג של מכונת מזל.

ניתן גם לעשות אבחנות ביחס לרמת הקושי של סדרת הבחירות של היריב. כפי שכבר צוין, מידת הווריאביליות של הסדרה (הניתנת למדידה בדרכים רבות) משפיעה על חסמי החרטה שהלומד יכול להבטיח. אינטואיטיבית, רמה נמוכה של ווריאביליות מסייעת ללומד לעקוב אחרי מהלכי היריב. מלאכתו של הלומד עשויה להיות קלה יותר גם אם יש יתירות בקבוצת המומחים. למשל, אם יש רק מספר קטן של מומחים איכותיים או אם מומחים רבים הם כמעט זהים למומחים אחרים, ניתן להבטיח חסמי חרטה טובים יותר. היבטים אלה ונוספים של נושא הלמידה המקוונת נדונים בפירוט בחלקה הראשון של עבודה זו, בפרקים 3 ו-4.

תמחור נגזרים

בחלק השני של העבודה מיושמת המתודולוגיה של מזעור חרטה לתמחור נגזרים, אחת הבעיות המרכזיות במימון. **נגזר** הוא נייר ערך שמחירו נקבע על סמך מחיר של נכס בסיס אחד או יותר, למשל, מניה. סוג חשוב של נגזר הוא **אופציה**, שהיא מכשיר פיננסי המאפשר למחזיק בו לקנות או למכור נכס מסוים במחיר ובזמן נתונים. למשל, אופציית רכש אירופאית מאפשרת לבעליה במועד הפקיעה שלה, T, לקנות נכס עבור מחיר K, הנקרא מחיר המימוש. כלומר, האופציה מעניקה לבעליה סכום של $\max\{S_T - K, 0\}$ בזמן T, כאשר S_T הוא מחיר הנכס (מניה) בזמן T. אופציה תקנית אחרת היא אופציית מכר אירופאית, המאפשרת לבעליה למכור נכס עבור מחיר K, או באופן שקול, לקבל סכום של $\max\{K - S_T, 0\}$ בזמן T.

מלבד אופציות הרכש והמכר התקניות, קיימות אופציות רבות אחרות, הנסחרות ומפותחות כדי לענות על צרכים מימוניים פרטניים. אופציות אלה נקראות **אופציות אקזוטיות**. דוגמא אחת היא אופציית רכש אירופאית מסוג lookback עם מחיר מימוש קבוע, אשר במועד הפקיעה שלה, T, מאפשרת לבעליה לבחור את הזמן הטוב ביותר בדיעבד לקנות נכס בסיס מסוים במחיר K. במילים אחרות, אופציית ה-lookback מעניקה לבעליה סכום של $\max\{M_T - K, 0\}$ בזמן T, כאשר M_T הוא מחיר הנכס המקסימלי בתקופת חיי האופציה.

תחום תמחור הנגזרים הושפע מאוד, הן תיאורטית והן מעשית, מנוסחת התמחור של בלאק, שולס ומרטון (Black and Scholes, 1973; Merton 1973). בעבודותיהם, שזיכו את כותביהן בפרס נובל לכלכלה (למעט בלאק, שנפטר טרם הענקת הפרס), הם מידלו את מחירן של מניות כתנועה בראונית גיאומטרית והניחו שהשוק הוא ללא הזדמנויות ארביטראז׳, כלומר, אין בו אפשרות לרווח חסר סיכון.

עם זאת, בהנחות של מודל בלאק-שולס-מרטון יש כמה בעיות ידועות. ראשית, המודל הוא רק בגדר הפשטה של שינויי המחיר בשוק, בעוד שבפועל המחירים הם בדידים ועשויים לקפוץ בחדות, והתשואות היומיות אינן בלתי-תלויות ושוות התפלגות. שנית, **התנודתיות** של המניה, שהיא הפרמטר העיקרי הנדרש במודל, אינה נתונה ויש לשערכה. למעשה, חישוב **התנודתיות הנגזרת** של המודל על סמך מחירי השוק האמיתיים של אופציות רכש אירופאיות מניב ערכים שונים עבור מחירי מימוש שונים, אפילו ביחס למועדי פקיעה זהים.

על הקשר בין מזעור חרטה לתמחור

אי-ההתאמות האמפיריות של מודל בלאק-שולס-מרטון שימשו מוטיבציה להצגת מודל מסחר על בסיס למידה מקוונת בעבודתם של דה מרזו, קרמר ומנצור (DeMarzo et al., 2006). במודל זה, שבו נעשה שימוש גם בתזה זו, המסחר מתקיים בזמן בדיד ובהנחה שהשוק הוא ללא ארביטראז׳. ההגבלות היחידות על מחיר המניה הן חסם על סכום ריבועי התשואות של המניה בכל תקופות המסחר (ההשתנות הריבועית) וחסם על הערך המוחלט של התשואה בתקופת מסחר אחת. חסם אפשרי נוסף הנבחן במודל ואשר מאומץ בתזה זו, הוא חסם עליון על מחיר המניה. מודל זה מאפשר בבירור קפיצות ותלות בין מחירים, תופעות המאפיינות שווקים אמיתיים.

העבודה של (DeMarzo et al., 2006) הראתה שבשוק ללא ארביטראז׳, ניתן לגזור חסמים עליונים על מחירי אופציות רכש אירופאיות בעזרת אלגוריתמי מסחר, המתבססים על אלגוריתמי מזעור חרטה במודל המומחה הטוב ביותר. תוצאה זו מתבססת על קשר יסודי בין אלגוריתמי מסחר לבין תמחור נגזרים, המתואר להלן. נניח שנתון נגזר, המבטיח לבעליו תשלום במועד עתידי נתון עבור מחיר כלשהו, כאשר התשלום תלוי בהתפתחויות בשוק. יהי עתה נתון אלגוריתם מסחר, המצריך השקעה התחלתית כלשהי, ובאותו מועד עתידי מניב סכום, התלוי הן באסטרטגיית המסחר והן בהתפתחויות בשוק. היות שהשוק הוא ללא ארביטראז׳, הרי שאילו הראינו, שבכל התפתחות עתידית אפשרית בשוק שווי נכסי האלגוריתם עולה על הסכום המשולם למחזיק בנגזר, הרי שהסכום ההתחלתי המושקע באלגוריתם בהכרח מהווה חסם עליון למחיר הנגזר. יש לציין שעקרון דומה מאפשר הוכחת חסמים תחתונים על מחירי נגזרים.

כדי להראות חסם עליון על המחיר במקרה הפשוט במיוחד של אופציית רכש אירופאית, מספיק למצוא אלגוריתם מסחר שסך נכסיו הסופיים עולה הן על מחיר המימוש והן על מחיר המניה הסופי. למעשה, עליו להתחרות תמיד בהצלחה מול הטובה מבין שתי אסטרטגיות: זו המחזיקה בסכום המימוש, וזו המחזיקה במניה מבלי לסחור בה. באמצעות הסבת אלגוריתם מזעור חרטה לתסריט של מסחר, דה מרזו, קרמר ומנצור הראו שניתן לתרגם חרטה נמוכה ליכולת מובטחת להתחרות בהצלחה מול שתי אסטרטגיות אלה. בחירתם הספציפית באלגוריתם Polynomial Weights אפשרה להם להוכיח חסמי מחיר התלויים בהשתנות הריבועית של מחיר המניה אך לא במספר תקופות המסחר. בכך הם קשרו את המחיר בתנודתיות, בדומה לתמחור בלאק-שולס-מרטון, וכן אפשרו הגברת תדירות המסחר ללא השפעה מפורשת על המחיר.

תזה זו מרחיבה באופן ניכר את תחום תמחור הנגזרים באמצעות מזעור חרטה. בחלקה השני נעשה שימוש באלגוריתם מזעור החרטה של (DeMarzo et al., 2006) לתמחור מגוון רחב של נגזרים אקזוטיים כפונקציה של ההשתנות הריבועית. בפרט, אלגוריתם זה משולב עם אלגוריתמי מסחר חד-כיווניים (היכולים רק למכור מניות) לצורך הוכחת חסמים עליונים על מחירי אופציות lookback. בנוסף, מוכחת נוסחה לתרגום ישיר של חסמים על ביצועי אלגוריתמים במודל המומחה הטוב ביותר לחסמים בתסריט של מסחר. בעזרת נוסחה זו מתאפשרת הוכחה של חסמים עליונים ותחתונים חדשים על מחירי נגזרים, בפרט כאלה המתבססים על תוצאות בחלקה הראשון של התזה.

תוכן העבודה

התוצאות המובאות בתזה זו מאורגנות בשני חלקים. החלק הראשון עוסק בתיאוריה של אלגוריתמי מזעור חרטה והחלק השני עוסק ביישומם לתמחור נגזרים.

חלק ראשון: מזעור חרטה

חלק זה כולל את פרקים 2-4, כאשר פרק 2 כולל רקע ואת עיקרי המודל.

פרק 3: חסמים תחתונים על חרטה של סדרות אינדיווידואליות

בפרק זה אנו מוכיחים חסמים תחתונים על חרטה, אשר תקפים **לכל** סדרת הפסדים שבוחר היריב. חסמים אלה תלויים בהשתנות הריבועית של סדרת ההפסדים ותקפים למשפחות גדולות של אלגוריתמי למידה במודל אופטימיזציה לינארית מקוונת. תוצאות אלה שונות באופיין מהחסמים התחתונים המקובלים, שמטרתם להראות את האופטימליות של חסמים עליונים על החרטה, ועל כן מסתפקים בהוכחת **קיומן** של סדרות הפסדים קשות.

ראשית, אנו נדרשים לשאלת אפיון אלגוריתמים שהם בעלי חרטה אי-שלילית לכל סדרה. אנו מתמקדים באלגוריתמים אשר בכל זמן t מגדירים את החלטתם הבאה $\mathbf{x}_{t+1}$ כפונקציה רציפה g של וקטור ההפסדים המצטברים עד לאותו רגע, $\mathbf{L}_t$. אנו מראים שלאלגוריתמים אלה יש חרטה אי-שלילית לכל סדרה, אםם g היא גרדיינט של פונקצית פוטנציאל קעורה, וכי תכונה זו מאפיינת את כל האלגוריתמים ממשפחת RFTL.

התוצאה המרכזית בפרק זה מראה שקיום חסם עליון על החרטה גורר קיום חסם תחתון על החרטה המקסימלית במהלך המשחק. בפרט, אם החסם העליון הוא מהצורה $O\left(\sqrt{Q}\right)$ לכל סדרה עם השתנות ריבועית Q, אזי החרטה המקסימלית לאורך המשחק היא בהכרח $\Omega\left(\sqrt{Q}\right)$, לכל סדרה עם השתנות ריבועית $\Theta(Q)$. תוצאה זו מוכחת לשתי מחלקות של פוטנציאלים גזירים פעמיים ברציפות. מחלקה אחת כוללת פוטנציאלים עם הסיאן מוגדר שלילית בסביבה של וקטור האפס, ומכילה את Online Gradient Descent. מחלקה שנייה כוללת פוטנציאלים במודל המומחה הטוב ביותר, המקיימים דרישות רגולריות טבעיות, ומכילה את Hedge. בהקשר של המחלקה הראשונה אנו מודדים את ההשתנות הריבועית כ-$\sum_{t=1}^{T}\|\mathbf{l}_t\|_2^2$, כאשר $\mathbf{l}_t$ הוא וקטור ההפסדים שבוחר היריב בזמן t. בהקשר של המחלקה השנייה, אנו משתמשים בהשתנות הריבועית היחסית, המהווה גודל מתאים יותר להקשר זה, ומוגדרת כ-

$\sum_{t=1}^{T}\left(\max_i\{l_{i,t}\}-\min_i\{l_{i,t}\}\right)^2$. נציין כי החסמים התחתונים עבור Hedge משמשים בפרק 8 להוכחת חסמים תחתונים למחירי אופציות.

התוצאות בפרק זה פורסמו במאמר הבא:

E. Gofer and Y. Mansour.
Lower bounds on individual sequence regret.
In Algorithmic Learning Theory, pages 275–289, 2012.

פרק 4: מזעור חרטה למומחים מתפצלים

פרק זה מציג אלגוריתמים לווריאנט של מודל המומחה הטוב ביותר, שבו קבוצת המומחים יכולה לגדול במהלך המשחק. ספציפית, בווריאנט זה מומחים חדשים עשויים להתפצל בכל סיבוב מכל מומחה קיים, על-פי בחירת היריב, תהליך היוצר מבנה של עץ. אנו מראים כי אלגוריתם הממזער חרטה במודל זה מאפשר גם השגת חסמי חרטה משופרים במודל המומחה הטוב ביותר הרגיל, במקרים של יתירות גבוהה בקרב קבוצת המומחים.

אנו בוחנים את מודל המומחים המתפצלים הן בתסריט של מידע מלא, והן בתסריט השודד מרובה הזרועות. לתסריט הראשון אנו מציגים מודיפיקציה של אלגוריתם Hedge, ומוכיחים עבורה חסמי חרטה אופטימליים. לתסריט השני אנו מוכיחים חסמי חרטה עבור מודיפיקציה של אלגוריתם Exp3, שהוא גרסת Hedge המותאמת לתסריט השודד מרובה הזרועות (Auer et al., 2002).

התוצאות במודל של מידע מלא מיושמות לשני תסריטים של יתירות בקבוצת המומחים. בתסריט הראשון, המומחה המוביל (כלומר, בעל ההפסד המצטבר הנמוך ביותר) בכל רגע במשחק מגיע מקבוצה קטנה של מומחים איכותיים. בתסריט זה אנו מראים חסם חרטה של $O\left(\Lambda_T(1+\ln L_T^*)+\sqrt{L_T^*\Lambda_T}\right)$, כאשר Λ_T מציין את מספר המובילים השונים במהלך המשחק, ו- L_T^* מציין את ההפסד הכולל של המומחה הטוב ביותר בסוף המשחק, כשהחסם עצמו אינו תלוי במספר המומחים. יש לציין, כי תוצאה זו תקפה גם אם נדרש סף מסוים של יתרון על מנת להפוך למוביל.

בתסריט אחר, ההפסדים הכוללים של כל המומחים מתרכזים סביב הערכים של מספר מומחים מצומצם. ספציפית, קיימות $N_{\alpha,T}$ קבוצות מומחים כך שההפרש בין ההפסדים הכוללים של זוג מומחים באותה קבוצה אינו עולה על α במהלך המשחק. עבור תסריט זה אנו מראים חסם של

$$O\left(N_{\alpha,T}\left(1+\alpha N_{\alpha,T}\right)(1+\ln L_T^*)+\sqrt{L_T^* N_{\alpha,T}}\right)$$

על החרטה.

חשוב להדגיש שהלומד אינו זקוק לכל מידע מוקדם לגבי זהות המובילים או הקבוצות, או אף לגבי מספרם. כמו כן, החסמים בשני התסריטים הם למעשה אופטימליים: הגורמים העיקריים, $\sqrt{L_T^*\Lambda_T}$ ו- $\sqrt{L_T^* N_{\alpha,T}}$, ניתנים לשיפור רק בפקטורים קבועים.

פרק זה מתבסס על המאמר הבא:

E. Gofer, N. Cesa-Bianchi, C. Gentile, and Y. Mansour.
Regret minimization for branching experts.
Journal of Machine Learning Research - Proceedings Track, 30:618–638, 2013.

חלק שני: תמחור נגזרים

חלק זה כולל את פרקים 5-8, כאשר פרק 5 משמש למתן רקע והצגת עיקרי המודל.

פרק 6: תמחור נגזרים אקזוטיים

פרק זה מיישם מתודולוגית מזעור חרטה לתמחור מגוון נגזרים אקזוטיים, ומכליל בכך את התוצאה של (DeMarzo et al., 2006). הפרק כולל חסמים עליונים למחירים של אופציות exchange, אופציות shout, אופציות lookback, אופציות רכש עם מחיר מימוש ממוצע ואופציות רכש על מחיר ממוצע. חסמים אלה מתבססים על חסם עליון על המחיר של אופציה המשלמת בזמן עתידי את המקסימום של מחירי מספר נגזרים. אופציה זו מתומחרת על בסיס חסמי חרטה כפליים, כלומר חסמים תחתונים על היחס בין ההון הסופי של אלגוריתם מסחר לשווי הסופי של הנגזרים האמורים. יתר האופציות מובעות בלשון האופציה הזו ומתומחרות על סמך התמחור שלה. הניתוח משתמש ברכיב מזעור חרטה זהה לזה של (DeMarzo et al., 2006), וחסמי המחיר תלויים בהשתנות הריבועית של מחיר נכסי הבסיס.

בפרק זה מוצג גם תמחור מבוסס מזעור חרטה של כל נגזר המעניק לבעליו בזמן עתידי תשלום שהוא פונקציה קמורה של מחירה של מניה באותו זמן. נגזרים אלה שקולים לתיק של אופציות רכש ללא פוזיציות בחסר, ולכן ניתן להשתמש בחסמים עליונים על מחירי אופציות רכש, כדוגמת זה של (DeMarzo et al., 2006), להוכחת חסם עליון על מחירם.

פרק זה מתבסס על המאמר הבא:

E. Gofer and Y. Mansour.
Pricing exotic derivatives using regret minimization.
In Algorithmic Game Theory, pages 266–277, 2011.

פרק 7: בחינה מקרוב של אופציות Lookback

בפרק זה אנו מציגים משפחה של אלגוריתמי מזעור חרטה, המשלבת שני רכיבים אלגוריתמיים: רכיב מזעור חרטה, ורכיב **מסחר חד-כיווני**, אשר כשמו, יכול רק למכור בהדרגה אחזקות במניה נתונה. ביצועי שני הרכיבים יחדיו מתורגמים לחסמים עליונים על המחיר של אופציות lookback.

הניתוח המובא משתמש באלגוריתם של (DeMarzo et al., 2006) בתור רכיב מזעור החרטה, ובוחן שני כיוונים עיקריים באשר לרכיב המסחר החד-כיווני. כיוון אחד הוא הפעלת כלל מכירה ספציפי מבוסס מחיר, אשר בעזרתו מתקבל חסם מחיר קונקרטי. כיוון אחר מנתח את מחיר האופציה כפונקציה של **יחס התחרותיות** של אלגוריתם מסחר חד-כיווני כללי, כלומר, חסם עליון על היחס בין המחיר המקסימלי של המניה לבין ההון המתקבל משימוש באלגוריתם המסחר. בפרט, הדבר מאפשר שימוש באלגוריתם מסחר

חד-כיווני אופטימלי עבור המודל שהוצג בעבודתם של אל-יניב, פיאט, קארפ וטרפין (El-Yaniv et al., 2001). שני סוגי החסמים תלויים גם בהשתנות הריבועית של מחיר המניה.

אנו מראים שהאלגוריתמים המשולבים שאנו מציגים עשויים לשפר את יחס התחרותיות האופטימלי המוכח במודל של (El-Yaniv et al., 2001). הסיבה לכך כפולה: ככלל, האלגוריתמים המשולבים גם קונים וגם מוכרים את המניה, וכמו כן, המודל בו אנו עובדים מניח חסם על ההשתנות הריבועית.

פרק זה מתבסס על המאמר הבא:

E. Gofer and Y. Mansour.
Regret minimization algorithms for pricing lookback options.
In Algorithmic Learning Theory, pages 234–248, 2011.

פרק 8: תמחור המבוסס על חרטה חיבורית

תוצאות תמחור הנגזרים המובאות בפרקים 6 ו-7 מתבססות על חסמים תחתונים על היחס בין ההון הסופי של אלגוריתמי מסחר לבין השווי הסופי של נגזרים מסוימים. מכיוון שהשוק הוא ללא ארביטראז׳, נובעים מחסמים אלה גם חסמים עליונים על מחירי אופציות.

בליבה של שיטה זו מצוי אלגוריתם Polynomial Weights למודל המומחה הטוב ביותר, שהוסב לאלגוריתם מסחר (הנקרא Generic) בעבודה של (DeMarzo et al., 2006). הסבה זו, שהצריכה שינוי של האלגוריתם ואנליזה חדשה של ביצועיו, מצביעה הן על הקשר היסודי בין אלגוריתמים למודל המומחה הטוב ביותר לאלגוריתמי מסחר והן על הקושי בתרגום התוצאות ממודל אחד למשנהו.

בפרק 8 אנו מפתחים שיטות לתרגום ישיר של חסמי חרטה של אלגוריתמים במודל המומחה הטוב ביותר לחסמים על היחס בין ההון הסופי של אלגוריתמי מסחר לבין השווי הסופי של הנכסים הנסחרים. באופן זה, אנו מסוגלים להסב תוצאות קיימות מעולם מזעור החרטה לעולם התמחור ללא צורך בשינויים אלגוריתמיים או בניתוח מחודש. בדרך זו אנו מוכיחים חסמים עליונים ותחתונים על מחירי אופציות ועושים זאת אף במודלים כלליים יותר מאשר המודל של דה מרזו, קרמר ומנצור.

בפרט, אנו מוכיחים חסמים תחתונים על המחיר של אופציות רכש אירופאיות **בכסף**, כלומר עם מחיר מימוש השווה למחיר ההתחלתי של המניה. תוצאות אלה עושות שימוש בחסמים תחתונים על החרטה המקסימלית של אלגוריתם Hedge, אשר פותחו בפרק 3.

עבור אופציות אלה אנו מוכיחים, בין השאר, שאם נתון מראש הערך $q = \sum_{t=1}^{T} \ln^2 (1 + r_t)$, כאשר r_t הוא התשואה של המניה בזמן t, ו-$q < 0.5$, הרי שמחיר האופציה חסום מלרע על-ידי $0.1\sqrt{q}$.

בהשוואה, במודל בלאק-שולס-מרטון המחיר האסימפטוטי (לערכי q קטנים) מקביל ל- $\frac{\sqrt{q}}{\sqrt{2\pi}} \sim 0.4\sqrt{q}$.

כלומר, ההתנהגות האסימפטוטית של שני החסמים זהה, למרות ההנחות המחמירות יותר, ובפרט, הנחת היריב, במודל שלנו. תוצאה זו משפרת חסם קודם של $\Omega(q)$ המוכח בעבודה של (DeMarzo et al., 2006).

פרק זה מתבסס בחלקו על תוצאות מהמאמר:

E. Gofer and Y. Mansour.
Lower bounds on individual sequence regret.
In Algorithmic Learning Theory, pages 275–289, 2012.