

**POLITECNICO DI MILANO**  
**Mathematical Engineering Master Degree**  
**Dipartimento di Matematica**



**POLITECNICO**  
**MILANO 1863**

## **Titolo Serio e Importante**

**Artificial Intelligence and Robotic Laboratory  
of Politecnico di Milano**

**Supervisor: Prof. Marcello Restelli**

**Co-supervisors: Francesco Trovò Ph.D., Edoardo Vittori,**

**Master thesis of:  
Martino Bernasconi de Luca, 10465492**

**Academic Year 2018-2019**







# Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Online Learning</b>	<b>7</b>
2.1	Online Learning . . . . .	8
2.1.1	Regret and Experts . . . . .	9
2.1.2	Existence of No-Regret Strategies . . . . .	11
2.1.3	Experts . . . . .	13
2.1.4	Uncountable Experts . . . . .	13
2.1.5	Exp-Concave loss functions . . . . .	14
2.2	Regret Minimization in Games . . . . .	15
2.2.1	Mixed extension . . . . .	16
2.2.2	MinMax Consistency . . . . .	17
2.3	Online Convex Optimization for Regret Minimization . . . . .	18
2.3.1	A General Algorithm for Online Convex Optimization . . . . .	19
2.3.2	Geometric Convexity and Why OMD works . . . . .	21
2.3.3	From Online Learning to Statistical Learning . . . . .	22
<b>3</b>	<b>Information, Prediction and Investing</b>	<b>25</b>
3.1	Probability assignment . . . . .	25
3.1.1	Connection to Information Theory . . . . .	27
3.1.2	Horse Races . . . . .	28
3.2	From Horse Races to Online Portfolio Optimization . . . . .	29
3.2.1	The Online Portfolio Optimization Model . . . . .	29
3.2.2	Effectiveness of Constant Rebalancing Portfolios . . . . .	30
<b>4</b>	<b>Algorithms for the Online Portfolio Optimization Problem</b>	<b>33</b>
4.1	Universal Portfolios . . . . .	33
4.2	Exponential Gradient . . . . .	33
4.3	Online Newton Step . . . . .	33
	<b>Bibliography</b>	<b>35</b>



# List of Figures

2.1	Online Learning with Expert Advice as Multi Agent-Environment interaction. . . . .	9
2.2	Rock Paper Scissor Dynamics Exponentially Weighted Majority	19





# Chapter 1

## Introduction

Classical investment techniques for the portfolio management problem derive from the knowledge of the statistical distribution of the assets return. Then, once the statistical model has been chosen, the problem get solved by optimizing the expected value of the utility of some random variable (usually accounting for the trade-off between risk and return), that describes the value of the portfolio in some fixed time in the future. This line of thinking has been proposed and sustained by Markovitz, Samuelson, Fama ecc... , and it is now called Modern Portfolio Theory (MPT).

This approach is known to be very susceptible to the errors in the modelling of the random variable that model the asset return. In fact, it is known that the markets have a non stationary behaviour, which means that every statistical assumption is ephemeral and unreliable. and they are usually referred to to backward looking, i.e. that they optimize

A different approach has been originated from the fields of information theory at the Bell Labs in the 1950, from the works of Shannon, Kelly and Cover. This methods were first included in the classical portfolio theory framework, under the name of Capital Growth Theory (CGT) [Hakansson et al., 1995] and then got included in the machine learning literature under the framework of Online Game Playing. Only recently this field has been taken into the Online Optimization This formulation has very interesting properties such as stability in a game theory fashion (equilibrium) and robustness versus adversarial manipulation.

One of the strongest points in favor of this techniques are the strong theoretical guarantees that algorithms developed under this framework can give. This guarantees come from the game theory concept of Regret, which is a form of dissatisfaction originated from having taken an action, instead of another action.

Principal in this thesis will be the extension of the modelling of the financial applications of this methodologies to the presence of transaction costs and to provide strong theoretical assurance even in the presence of transaction costs. In fact in many financial situations, transaction costs are not modelled and





## Chapter 2

# Online Learning

Non hai corretto un sacco di cose che ti avevo segnato l'altra volta, soprattutto virgole, che non sto a segnarti di nuovo — ti segno solo quelli gravi

Online Learning is a theoretical framework to formalize a sequential decision problem in which an agent has to take consecutive actions in an environment. Every time the agent takes an action, the environment returns a loss signal (or reward depending on the convention on the sign). This framework is similar to other sequential decision problems such as Reinforcement Learning [Sutton et al., ], with the exception that the loss function is decided by an adversary which has complete knowledge of your strategy in advance, rather than being described by a stochastic probability kernel. The purpose of this section is to present the general framework of Online Game Playing and to introduce the notation necessary for the development of the theory. We will define formally the framework of Online Learning with Expert Advice, which is one of the most studied frameworks of Online Learning, due to its ability to include many other frameworks, such as Multi Armed Bandit and Online Convex Optimization. Then we will present the concept of *regret* and present the relationship of Online Learning to classical repeated games, a classical framework coming from the field of Game Theory. We are interested in this framework in order to model classical repeated investments in this framework. Modern finance has more and more the need for a Game Theoretic approach, this is evident when one looks at the field of *Online Venue Market Making*, that can be modeled naturally as a repeated game, or in merger and acquisition that can be modeled as a normal form game [Jiang et al., 2016]. Finally we will introduce Online Convex Optimization as a special case of Online Learning with expert advice and its interesting relationship to theoretical statistical learning. The choice of this path, from Online Learning to Online Convex Optimization, has been done to show how general and powerful Online Learning is in its simplicity, and why Online Convex Optimization is the most suitable framework to present our

F: è sbagliato.

define

F: manca coesione  
OL->finance. M:  
meglio?

add ref e ag-  
giustare

contribution to Online Portfolio Selection, that will be presented in later chapters.

In fact, even if we will focus on the portfolio problem, the apparently simple formulation of this framework is capable to encompass many other applications and problems, such as network routing [Belmega et al., 2018] and dark pool order allocation [Agarwal et al., 2010]. A thorough dissertation of the techniques that have been developed in the field of Online Learning can be found in [Cesa-Bianchi and Lugosi, 2006].

## 2.1 Online Learning

**Definition 2.1.1.** (*Online Game Playing*). Let  $\mathcal{Y}$  be the outcome space,  $\mathcal{D}$  the prediction space and  $f : \mathcal{D} \times \mathcal{Y} \rightarrow \mathbb{R}$  is a loss function, an Online Game is the following sequential game played by the forecaster  $\mathcal{A}$  and the environment:

For each round  $t \in \mathbb{N}$

1. The learner  $\mathcal{A}$  chooses an element of the decision space  $x_t \in \mathcal{D}$ .
2. The environment chooses the element  $y_t \in \mathcal{Y}$ , and subsequently determines the loss function  $f(\cdot, y_t)$ .
3. The agent  $\mathcal{A}$  incurs in a loss  $f(x_t, y_t)$ .
4. The agent updates its cumulative losses  $L_t = L_{t-1} + f(x_t, y_t)$  with  $L_0 = 0$ .

In Online Learning an agent  $\mathcal{A}$  has to guess the outcome  $y_t$  based on a the past outcomes  $y_1, y_2, \dots, y_{t-1}$  of some events that are in the outcome space  $\mathcal{Y}$ , at each time step she will play (sometimes we will also say *predict*)  $x_t$ , that is an element of the prediction space  $\mathcal{D}$ , and the environment will choose a loss function  $f(\cdot, y_t)$  by determining the outcome  $y_t$ . Sometimes it is not important to know the exact outcome of the round and so we can identify the function  $f(x, y_t)$  with  $f_t(x)$ . The agent  $\mathcal{A}$  is essentially the identification of the functions that maps the history of past outcomes to the new prediction:

$$\mathcal{A} \equiv \{h_{t-1} := (y_1, \dots, y_{t-1}) \mapsto x_t\}_{t \geq 1}$$

The simplest case is for  $\mathcal{Y} = \mathcal{D}$  and both of finite cardinality, meaning that there are only a finite number of actions that the agent  $\mathcal{A}$  can choose from. We will sometimes refer to the environment defined in Section 2.1.1 as

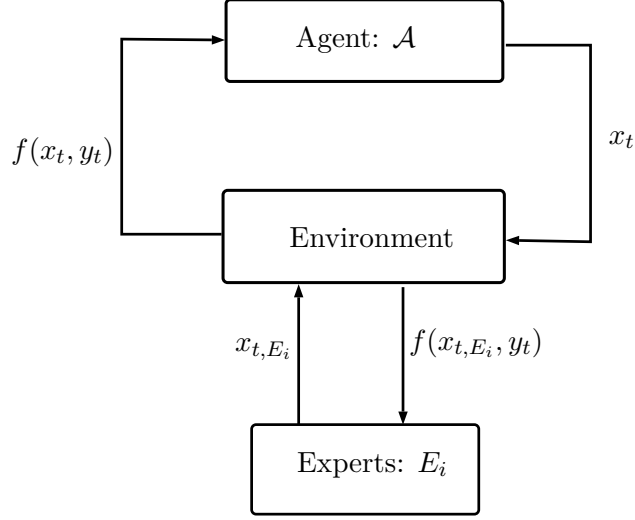


Figure 2.1: Online Learning with Expert Advice as Multi Agent-Environment interaction.

"adversarial", since no stochastic characterization is given to the outcome sequence  $y_t$  and the analysis of the regret is done assuming a worst case scenario. Since the adversary knows the prediction  $x_t$ , before deciding the outcome  $y_t$ , designing an algorithm which tries to minimize the loss is an hopeless task and so we have to set an easier scope. We will also present the counterexample to why the absolute minimization of the loss is an hopeless task, and present the adapt minimal framework to successful Online Learning in Adversarial Environment.

### 2.1.1 Regret and Experts

We stated that the objective of absolute loss minimization is hopeless in an adversarial framework, as the adversary can always choose the outcome  $y_t$  that maximizes the loss  $f(x, y_t)$  regardless of the decision  $x \in \mathcal{D}$  taken by the learner. We shall present a simple counterexample in this setting.

Take  $\mathcal{D}$  as a space of binary outcomes, *i.e.*  $|\mathcal{D}| = 2$  absolute loss as  $f(x, y) = |x - y|$ . Since the adversary plays after the learner  $\mathcal{A}$ , it can make the loss of the learner  $L_T = T$  by outputting the bit non predicted by the learner. Notice that no assumption has been made on the strategy followed by the learner  $\mathcal{A}$ . From this example it is clear that the learner has to set a less ambitious goal.

We do so by extending the theoretical formulation in Section 2.1 by including a set  $\mathcal{E}$  of other players, this setting is called "prediction with

expert advice". At each time step of the prediction game, each expert  $e \in \mathcal{E}$ , predicts an element  $x_{t,e} \in \mathcal{D}$ , and incurs in a loss  $f(x_{t,e}, y_t)$ , just as the agent  $\mathcal{A}$ , creating a general multi-agent interaction as in Figure 2.1. Now the goal that the learner sets itself to solve is to obtain small losses with respect to the best expert in the class  $\mathcal{E}$ . This concept is captured by the definition of regret. Formally, we define the regret  $R_{T,e}$  for the agent  $\mathcal{A}$  with respect to expert  $e \in \mathcal{E}$  (assumed finite for the moment) as follows:

$$R_{T,e} = L_T - L_{T,e} \quad (2.1)$$

The regret observed by the agent  $\mathcal{A}$  with respect to the entire class of experts  $\mathcal{E}$  is defined as:

$$R_T = \max_{e \in \mathcal{E}} R_{T,e} = L_T - \min_{e \in \mathcal{E}} L_{T,e}. \quad (2.2)$$

The task the agent  $\mathcal{A}$  is set to solve is to find a sequence  $x_t$  function of the information obtained up to the time  $t$  in order to obtain small regret  $y_T$  with respect to any sequence  $y_1, y_2, \dots$  chose by the environment.

In particular we aim to achieve sub-linear regret  $R_T = o(T)$ , meaning that the per-round regret  $R_T/T$  will asymptotically vanish:

$$R_T = o(T) \implies \lim_{T \rightarrow \infty} \frac{R_T}{T} = 0, \quad (2.3)$$

where  $o(T)$  is the space of sub-linear affine functions. A strategy  $\mathcal{A}$  that attains sub-linear regret is called *Hannan-Consistent* [Hannan, 1957].

The regret is a measure of the distance between our online performance and the best offline (in retrospect) performance among the expert class  $\mathcal{E}$ , this is also called *external regret* since it is compared to the external set of experts  $\mathcal{E}$ . A surprising fact is even that such algorithms do even exist. Indeed a first result is that in general there are no Hannan Consistent strategies, and just introducing the concept of regret is not enough for successful Online Learning. ✖

A first simple counterexample can be found in [Cover, 1966]. If the decision space  $\mathcal{D}$  is finite then there exists a sequence of loss function such that  $R_T = \Omega(T)$ . Again take  $\mathcal{D}$  as a space of binary outcomes, absolute loss as  $f(x, y) = |x - y|$ , and the class of experts is composed by two experts, one predicting always 0 and the other always 1. Taking  $T$  odd, we have that the loss of the best expert is  $L_{T,e} < \frac{T}{2}$ , and we have already shown that the adversary can make the loss of the learner  $L_T = T$ . It is now evident that the regret is  $R_T > T - \frac{T}{2}$ , which do not allow  $R_T/T \rightarrow 0$ . This argument is easily extended in the case of any finite decision space  $\mathcal{D}$ .



questa frase non funziona grammaticalmente ma  
non so come correggerla perché non capisco  
cosa devi dire

In order for the learner to achieve sub-linear regret is to randomize its predictions, the learner, at each turn  $t$ , holds a probability distribution on the decision space and plays  $x_t$  according to this distribution. Clearly the adversary has knowledge of the probability distribution of the learner  $\mathcal{A}$ , but has no knowledge of the random seed used by the agent  $\mathcal{A}$ , *i.e.* does not know the actual decision taken according to the distribution held by the agent. If the original decision space was  $\mathcal{D}$  with  $|\mathcal{D}| = N$  after the randomization of the decision, we effectively transformed the decision space  $\mathcal{D}$  into the  $\Delta_{N-1} \in \mathbb{R}^N$  probability simplex. By doing so we are formally extending the game into its mixed extension, as will be discussed further in Section 2.2. It can be viewed also as a *convexification* of the domain, pointing to the undeniably necessity of convex geometry in this context, that will be discussed in 2.3. Therefore, from now on the domain  $\mathcal{D}$  will be convex, either by the problem specification or by randomized convexification if the problem has discrete decision space.

### 2.1.2 Existence of No-Regret Strategies

In this section we will show the existence of Hannan-consistent strategies in the case of finite experts and provide a general form to generate sub-linear regret strategies. The general idea with a finite class of experts is given by the Weighted Average Forecaster. A natural class of algorithm to explore is the predicting as the weighted average prediction of the experts predictions, weighted on the accrued regret suffered by the agent  $\mathcal{A}$ :

**Definition 2.1.2.** (*Weighted Average Forecaster*). For a finite class of experts  $\mathcal{E}$  the weighted average prediction is defined as

$$x_t = \frac{\sum_{i=1}^N w_{t-1,i} x_{t,i}}{\sum_{i=1}^N w_{t-1,i}}, \quad (2.4)$$

where  $w_{t-1,i} > 0$  and  $x_{t,i}$  is the prediction of expert  $E_i \in \mathcal{E}$  at round  $t$ .

Since  $\mathcal{D}$  is convex we have that  $x_t \in \mathcal{D}$ . Then it is natural to assume that the weights are a function of the cumulated regret suffered by the agent with respect to the experts, and also that the change in weight is proportional to the change in a potential function. We can generalize the simple weighted average prediction in Equation (2.1.2) in the following general form, introduced in [Cesa-Bianchi and Lugosi, 2003]:

$$x_t = \frac{\sum_{i=1}^N \partial_i \Phi(\mathbf{R}_{t-1}) x_{t,i}}{\sum_{i=1}^N \partial_i \Phi(\mathbf{R}_{t-1})}, \quad (2.5)$$

where  $\Phi(\mathbf{u}) = \varphi \left( \sum_{i=1}^N \phi(u_i) \right)$  is a function  $\Phi : \mathbb{R}^N \rightarrow \mathbb{R}^+$  defined through two increasing functions  $\phi, \varphi : \mathbb{R} \rightarrow \mathbb{R}^+$ ,  $\varphi, \phi \in \mathcal{C}^2(\mathbb{R})$  and concave and convex, respectively and  $\mathbf{R}_T = (R_{T,1}, \dots, R_{T,N})$ . By specializing the two functions  $\varphi, \phi$  we can derive most of the algorithm for dealing with prediction under expert advice. The reasons behind the general form of Equation (2.5) and an extended discussion can be found in [Hart and Mas-Colell, 2001], [Cesa-Bianchi and Lugosi, 2003] and [Blackwell et al., 1956], but the general idea is that the form of Equation (2.5) has the following property:

**Theorem 2.1.1.** *If  $x_t$  is given by Equation (2.5) and the loss  $f(\cdot, y)$  is convex in the first argument then the instantaneous weighted regret satisfies:*

$$\sup_{y_t \in \mathcal{Y}} \sum_{i=1}^N [f(x_t, y_t) - f(x_{t,i}, y_t)] \partial_i \Phi(\mathbf{y}_{t-1}) \leq 0$$



*Proof.* By convexity of  $f(\cdot, y_t)$  we have that

$$f(x_t, y_t) \leq \frac{\sum_{i=1}^N \partial_i \Phi(\mathbf{R}_{t-1}) f(x_{t,i}, y_t)}{\sum_{i=1}^N \partial_i \Phi(\mathbf{R}_{t-1})}, \forall y_t \in \mathcal{Y} \quad (2.6)$$

And since  $\Phi(\mathbf{x}) = \varphi \left( \sum_{i=1}^N \phi(x_i) \right)$  we have that

$$\partial_i \Phi(\mathbf{x}) = \varphi' \left( \sum_{i=1}^N \phi(x_i) \right) \phi'(x_i) \geq 0$$

Hence we can rearrange the terms in Equation (2.6) to obtain the statement.  $\square$

Note that fixing the structure for the weights as in Equation (2.5) we have that  $w_{t,i} \propto \phi'(R_{t,i})$  ~~that~~ is an increasing function in  $R_{t,i}$  (since  $\phi$  is convex and increasing) that essentially states that we are increasing the probability of playing actions on which we saw high regret  $R_{t,i}$ .

check

F: che schifo di definizione. Si

**Definition 2.1.3.** *The exponentially weighted algorithm is Equation (2.5) where we defined  $\varphi(x) = \frac{1}{\eta} \ln(x)$  and  $\phi(x) = e^{\eta x}$  giving weights of the form*

$$w_{t-1,i} = e^{\eta y_{t-1,i}} / \sum_{j=1}^N e^{\eta y_{t-1,j}}$$

It can be shown that the algorithm defined by the update rule in Equation (2.1.3), and for a convex loss function  $f(\cdot, y_y)$ , gives the following guarantee on the regret:

e allora show + cit del toerema

$$R_T \leq \frac{\log(N)}{\eta} + \frac{T\eta}{8} \quad (2.7)$$



By choosing  $\eta = O\left(\sqrt{\frac{1}{T}}\right)$  we obtain a sub-linear regret  $R_T = \mathcal{O}(\sqrt{T})$ .

### 2.1.3 Experts

The theoretical framework described in Section 2.1 is very general and most suited for a game theory analysis of the problem. This helps us describe many other frameworks, such as Online Optimization, or Multi Armed Bandit (MAB) as embedded into a Game Playing framework with expert advice. It can then be specialized by fixing many elements of the definition, in order to be applied to the specific problem we are willing to solve. For instance, the class of experts  $\mathcal{E}$  is most of the time completely fictitious, meaning that the experts are not real players of the game, but most of the time they are *simulable* meaning that the agent  $\mathcal{A}$  is able to compute  $x_{t,e}$  for each expert  $e \in \mathcal{E}$  and most of the times the class of expert is very limited in its actions, *e.g.*  $\mathcal{E}$  is the class of experts for which  $x_{t,e}$  is constant in  $t$ . In this case, which is the most studied class of experts, we are basically just comparing our learner  $\mathcal{A}$  to the best fixed action  $x^*$  in hindsight. This is a clairvoyant strategy that attains the minimum cumulative loss over the entire length of the game  $T$ .

### 2.1.4 Uncountable Experts

In the case of uncountable experts the Exponentially Averaged Prediction cannot be applied directly, but can be extended to a continuous mixture of experts predictions. More specifically we need the case of the class  $\mathcal{E}$  being generated by a convex hull of a finite number of a base class of experts,  $\mathcal{E}_N$ . With continuous class of experts  $\mathcal{E}$  defined in this way, the regret definition becomes:

$$R_T = \sup_{q \in \Delta_{N-1}} R_{T,q} := L_T - \inf_{q \in \Delta_{N-1}} L_{T,q}, \quad (2.8)$$

where  $\Delta_{N-1} \subset \mathbb{R}^N$  is the  $N$ -simplex, and

$$L_{T,q} = \sum_{t=1}^T f(\langle q, x_{t,e} \rangle, y_t),$$

where  $x_{t,e} = (x_{t,1}, \dots, x_{t,N}) \in \mathbb{R}^N$  is the vector of expert predictions at time  $t$ .

### 2.1.5 Exp-Concave loss functions

It will be important for the study of Portfolio Optimization the exp-concave class of loss functions.  $f(\cdot, y)$  is  $\nu$ -exp concave if  $e^{-\nu f(\cdot, y)}$  is concave.

**Theorem 2.1.2.** *The Exponentially Weighted Average forecaster, for  $\nu$ -exp concave loss functions and for  $\eta = \nu$  has the following property:*

$$\Phi(\mathbf{R}_T) \leq \Phi(\mathbf{R}_0)$$

where  $\Phi(x) = \varphi\left(\sum_{i=1}^N \phi(x_i)\right)$  is chosen as  $\varphi(x) = \frac{1}{\nu} \log(x)$  and  $\phi(x) = e^{\nu x}$

*Proof.* The weights are given by  $w_{t-1,i} = e^{\nu y_{t-1,i}} / \sum_{j=1}^N e^{\nu y_{t-1,j}}$ . By exp-concavity we have that

$$e^{-\nu f(x_t, y_t)} = \exp \left\{ -\nu f \left( \frac{\sum_{i=1}^N w_{t-1,i} x_{t,i}}{\sum_{i=1}^N w_{t-1,i}}, y_t \right) \right\} \geq \frac{\sum_{i=1}^N w_{t-1,i} e^{-\nu f(x_{t,i}, y_t)}}{\sum_{i=1}^N w_{t-1,i}} \quad (2.9)$$

this can be rewritten as

$$\sum_{i=1}^N e^{\nu y_{t-1,i}} e^{\nu [f(x_t, y_t) - f(x_{t,i}, y_t)]} \leq \sum_{i=1}^N e^{\nu y_{t-1,i}} \quad (2.10)$$

Applying  $\varphi(x) = \frac{1}{\nu} \log(x)$  to both sides of equation (2.10) we obtain that

$$\Phi(\mathbf{R}_t) \leq \Phi(\mathbf{R}_{t-1})$$

that proves the thesis.  $\square$

The case of exp-concave functions is very special, since we can obtain Theorem 2.1.2 that can be used to prove regret bounds very easily:

$$R_T \leq \frac{1}{\eta} \log \left( \sum_{i=1}^N e^{\nu R_{T,i}} \right) = \Phi(\mathbf{R}_T) \leq \Phi(\mathbf{R}_0) = \frac{\log N}{\eta} \quad (2.11)$$

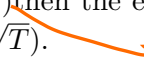
The case of exp-concave losses is also useful for the case of uncountable experts sketched in Section 2.1.4. This formulation will be of central importance for the portfolio optimization problem.

It is natural to extend the Exponential Weighted Majority algorithm described by equation (2.1.2) into its continuous case by:

$$x_t = \frac{\int \Delta_{N-1} w_{q,t-1} \langle q, x_{t,e} \rangle dq}{\int \Delta_{N-1} w_{q,t-1} dq} \quad (2.12)$$

**Theorem 2.1.3.** (*Mixture forecaster for exp-concave losses*). *Choosing  $w_{q,t-1} = \exp \left\{ -\eta \sum_{s=1}^{t-1} f(\langle q, x_{t,e} \rangle, y_s) \right\}$  in Equation (2.12), for a bounded  $\nu$ -exp concave loss function  $f(\cdot, y)$ , we obtain*

$$R_T \leq N\nu \left( \log \left( \frac{\nu T}{N} \right) + 1 \right)$$

Even in the case of uncountable many experts, exp-concavity of the loss function gives a better convergence rate of  $\mathcal{O}(\log T)$  then the exponentially weighted algorithm in Equation (2.7), which is  $\mathcal{O}(\sqrt{T})$ .  **manca spazio**

## 2.2 Regret Minimization in Games

In this section we explore the connection of the framework of Section 2.1 into a more classical repeated game framework. In the previous Section we looked at the adversary as a black box, without any specific model in mind. The reason of this chapter is to clarify its role as a player in the game and to show the game theoretical properties of Hannan-consistent agents. Since in Online Learning the convention is to speak about losses, we shall speak about losses (minimization) also in the classical definitions of game theory instead of payoffs (maximization).

**Definition 2.2.1.** (*Strategic Form  $K$ -Player Game*). *A Strategic form  $K$ -player game is a tuple  $\langle \mathcal{K}, \{X_i\}_{i \in \mathcal{K}}, \{l_i\}_{i \in \mathcal{K}} \rangle$  where*

1.  $\mathcal{K} = \{1, \dots, K\}$  is the finite set of players
2.  $X_i$  is the set of actions available to player  $i \in \mathcal{K}$

Forse da togliere tutto questo sub-chpt

3.  $l_i : \bigotimes_{k=1}^K X_k \rightarrow \mathbb{R}$  is the loss observed by player  $i \in \mathcal{K}$

The game is called finite if  $|X_i| < +\infty$  for all  $i \in \mathcal{K}$ .

### 2.2.1 Mixed extension

As in Section 2.1 we saw that it is impossible to obtain sub-linear regret in adversarial environment with finite decision space  $\mathcal{D}$ . A first step to solve this has been the *randomized convexification*, where finite action spaces are extended into convex sets, given by their probability simplex, also losses are to be interpreted as expected losses when the mixed extension is applied to the formal game. More formally:

secondo me devi mettere ; o .

**Definition 2.2.2.** (*Mixed-extension for finite games*). A finite game  $\langle \mathcal{K}, \{X_i\}_{i \in \mathcal{K}}, \{l_i\}_{i \in \mathcal{K}} \rangle$  can be extended into the game  $\langle \mathcal{K}, \{\tilde{X}_i\}_{i \in \mathcal{K}}, \{\tilde{l}_i\}_{i \in \mathcal{K}} \rangle$

1.  $\tilde{X}_i = \Delta_{|X_i|-1} \subset \mathbb{R}^{|X_i|}$  for all  $i \in \mathcal{K}$

2.  $\tilde{l} : \bigotimes \tilde{X}_i \rightarrow \mathbb{R}$  is defined as

$$\tilde{l}(x_1, \dots, x_K) = \sum_{i_1=1}^N \dots \sum_{i_K=1}^N p_{i_1} \dots p_{i_K} l(i_1, \dots, i_K)$$

Due to the impossibility result of Cover [Cover, 1966], we have to work with the mixed extension formulation of the game. So from now on we take this step implicitly. The taxonomy of game definition is quite extended and complex, we will focus on non-cooperative games since they are closely related to the setting tackled in the Online Learning field. More specifically, we will need the model for *Zero Sum Game*.

**Definition 2.2.3.** (*2-Player Zero-Sum Game*). A Zero Sum game is a tuple  $\langle \{X_1, X_2\}, l : X_1 \times X_2 \rightarrow \mathbb{R} \rangle$ . As in Definition 2.2.1  $X_1, X_2$  are the action spaces for Player 1 (row player) and Player 2 (columns player) respectively and  $l(x_1, x_2)$  for  $x_1, x_2 \in X_1 \times X_2$ , represents the losses for Player 1 and profits for player 2.

If this game is played for  $T$  turns, we can call it a repeated game, and the losses for each player will be  $L_1^{(T)} = \sum_{t=1}^T l_i(x_i^{(t)}, x_2^{(t)})$  and  $L_2^{(T)} = -L_1^{(T)}$ .

non capisco ma penso tu voglia dire: 'The question of what guarantees Hannan-consistent strategies do have brings to the game theoretical formulation of the problem...'

## 2.2.2 MinMax Consistency

The question of what guarantees does Hannan-consistent strategies bring to the game theoretical formulation of the problem, and why Online Learning is sometimes called *Learning in Games*. For such games we can define a *values* for the game as:

$$V_1 = \inf_{x_1 \in X_1} \sup_{x_2 \in X_2} l(x_1, x_2) \quad (2.13)$$

$$V_2 = \sup_{x_2 \in X_2} \inf_{x_1 \in X_1} l(x_1, x_2) \quad (2.14)$$

'This is the value' o  
'These are the values'

These is the value that the players can guarantees themselves, meaning that no matter the strategy of the columns player, the row player could guarantee himself a loss of at maximum  $V_1$ , the converse holds for the row player. It can be interpreted as the minimum loss (best payoff) that player could achieve if we know that the other player would play adversarially. It is clear that  $V_2 \leq V_1$ . In the case the zero sum-game is a mixed extension of a finite game, then the Von Neumann theorem states that  $V_1 = V_2$ .

Now we will embed the framework of Online Game Playing of Section 2.1 in a two player zero sum game. Online Learning is a special form of Zero Sum Game (possibly considering its mixed extension described in Definition 2.2.1) where  $X_1 \equiv \mathcal{D}$  and  $X_2 \equiv \mathcal{Y}$ . The loss function  $l : X_1 \times X_2 \rightarrow \mathbb{R}$  can be identified by the loss  $f : \mathcal{D} \times \mathcal{Y} \rightarrow \mathbb{R}$  of the Online Learning Agent  $\mathcal{A}$ . Now we will explore interesting properties of Hannan Consistent strategies. A surprising fact is that if the row player plays accordingly to a Hannan Consistent strategy then it achieve the value of the game  $V_1$ .

**Theorem 2.2.1.** *Hannan Consistent agents in Online Game Playing reach asymptotically the minmax value of the one shot game.*

$$\limsup_{T \rightarrow +\infty} \frac{1}{T} \sum_{t=1}^T f(x_t, y_t) \leq V_1$$

*Proof.* Let us suppose that player 1 plays an Hannan Consistent strategy and that  $y_1, y_2, \dots \subset \mathcal{Y}$  is a generic sequence played by the columns player.

$$\limsup_{T \rightarrow +\infty} \frac{y_T}{T} \leq 0 \quad (2.15)$$

that can be translate into

$$\limsup_{T \rightarrow +\infty} \frac{1}{T} \sum_{t=1}^T f(x_t, y_t) \leq \limsup_{T \rightarrow +\infty} \frac{1}{T} \inf_{x \in \mathcal{D}} \sum_{t=1}^T f(x, y_t) \quad (2.16)$$

secondo me è un po troppo colloquiale, forse dovresti usare 'We will call' o simile

Lets call  $\hat{y}_T$  the empirical distribution played by player 2 up to  $T$ :

$$\hat{y}_T(j) = \frac{1}{T} \sum_{t=1}^T y_t$$

by (2.16) we just need to show  $\frac{1}{T} \inf_{x \in \mathcal{D}} \sum_{t=1}^T f(x, y_t) \leq V$

$$\inf_{x \in \mathcal{D}} \frac{1}{T} \sum_{t=1}^T f(x, y_t) = \inf_{x \in \mathcal{D}} f(x, y_T) \leq \sup_{y \in \mathcal{Y}} \inf_{x \in \mathcal{D}} f(x, y) \leq V_1 \quad (2.17)$$

□

We showed that regardless of the strategy of player 2, a player playing an Hannan Consistent strategy achieves lower losses <sup>than</sup> that the value of the game  $V_1$ . Clearly using an Hannan Consistent strategy means that if player 2 <sup>was</sup> ~~were~~ not adversarial, then player 1 could potentially earn a significantly higher average payoff than the value  $V$  of the game. By symmetry, if both players play an Hannan Consistent strategy <sup>then</sup> ~~than~~ they will asymptotically reach the value of the game  $V = V_1 = V_2$ .

magari meglio 'a player using'

## 2.3 Online Convex Optimization for Regret Minimization

Ma non puoi dire così. O metti '..optimization ; this will be' oppure '..optimization , that will be..'

Let's compare ~~this~~ framework to an apparently unrelated problem, namely optimization, ~~this~~ will be the most suited framework to embed the Online Portfolio Optimization Problem. In online optimization an agent  $\mathcal{A}$  is set to optimize a sequence of functions  $f_t(x)$  where usually  $f_t : \mathcal{X} \rightarrow \mathbb{R}$  is a real valued function from the set  $\mathcal{X} \subset \mathbb{R}^n$ . In Online Convex Optimization literature, <sup>sometimes</sup> ~~some times~~ the loss functions are identified as  $f(x, y_t) \equiv f_t(x)$ . The decision space  $\mathcal{D}$  is assumed to be convex, as the are the functions  $f_t : \mathcal{D} \rightarrow \mathbb{R}$ . This framework was first devised in [Zinkevich, 2003], and has been later wildly used in the machine learning community to engineer optimization procedures [Shalev-Shwartz et al., 2012].

Convexity plays ~~a~~ central role in most of the analysis made in Online Learning, and Online Convex Optimization. Convexity of the domain  $\mathcal{D}$  and of the loss functions  $f(\cdot, y)$  bound the problem geometry and let us derive simple and efficient learning procedures. More generally in the subsequent section we will present the general learning.

Speigare meglio regret in OCO.



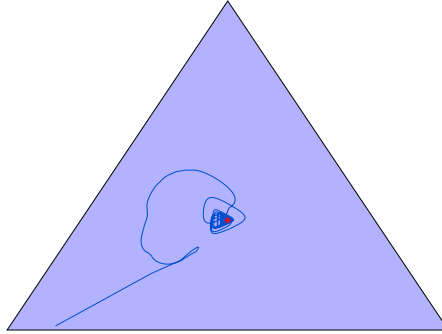


Figure 2.2: Rock Paper Scissor Dynamics Exponentially Weighted Majority

### 2.3.1 A General Algorithm for Online Convex Optimization

In this Section we will see an algorithm called *Online Mirror Descent* (OMD), that generalizes many Online Convex Optimization algorithms. It is a first order method that works in the dual space defined by the choice of some regularizator. The OMD algorithm is general and optimal in the sense that every Online Convex problem can be learned online nearly optimally with OMD; the precise formulation can be found in [Srebro et al., 2011].

regularizator?

mmmh non penso esista

OMD works with a class of regularizators called Bregman Divergences, [Banerjee et al., 2005].

**Definition 2.3.1.** (*Bregman divergence*). Given a differentiable convex function  $\psi : \mathcal{X} \rightarrow \mathbb{R}$  the Bregman divergence is defined as an operator  $d_\psi : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}^+$  defined for  $x, y \in \mathcal{X} \times \mathcal{X}$  as

$$d_\psi(x, y) = \psi(x) - \psi(y) - \langle x - y, \nabla \psi(y) \rangle \quad (2.18)$$

No ok questa a questa frase manca qualcosa. Ti ho messo quel punto ma nella frase 'We can see that....' manca un verbo, cioè è tipo 'Vediamo che per ... e per ...' e poi manca qualcosa. Forse questo qualcosa è la prima frase (quella cosa che dici con 'we have that..') ma allora devi ricostruire tutto il periodo

Since  $\psi$  is convex we have that  $d_\psi(x, y) \geq 0$ . We can see that by linearization of  $\psi(x)$  around  $y \in \mathcal{X}$  and by the inequality on the non standard inner product defined by the hessian of the function  $\psi$  which is positive thanks to the convexity. Since the operator defined in Equation (2.18) is not symmetric in its arguments, it does not define a metric in the space  $\mathcal{X}$ .

For  $\psi(x) = \|x\|_2^2$  then  $d_\psi(x, y) = \|x - y\|_2^2$ . For  $\psi(x) = \sum_{i=1}^N x_i \log(x_i)$  then  $d_\psi(x, y) = \sum_{i=1}^N x_i \log(x_i/y_i)$ , for  $x, y \in \Delta_{N-1} \subset \mathbb{R}^N$  which is the well know Kullback–Leibler divergence [Van Erven and Harremos, 2014].

The OMD algorithm for Online Convex Optimization uses the regularization given by a Bregman divergence to follow the best point in the convex set  $\mathcal{D}$  up to now, but it is kept close to the current one by the divergence operator. Formally:

**Definition 2.3.2.** (Online Mirror Descent). OMD for a Bregman Divergence induced by the differentiable, convex real values function  $\psi$ , and for a set of learning rates  $\{\eta_0, \dots, \eta_T\}$  has the following update rule:

$$x_{t+1} = \arg \min_{x \in \mathcal{X}} d_\psi(x, x_t) + \eta_t \langle \nabla f_t(x_t), x - x_t \rangle$$

Next we will show the idea for a general bound for the OMD algorithm, which is to show the geometric ideas behind the OMD algorithm. It is important to point that the analysis can be refined by fixing the loss function  $f_t$  or the convex function  $\psi$ .

The convex function  $\psi$  is assumed to be differentiable in the domain  $\mathcal{X}$ , but it is not required by the analysis, and in general we can work with subgradients rather than gradients.

**Theorem 2.3.1.** Let  $d_\psi : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$  the Bregman divergence associated to the convex smooth function  $\psi$ . Moreover, assume  $d_\psi$  is  $\alpha$ -strong convex, i.e.  $d_\psi(x, y) \geq \frac{\alpha}{2} \|x - y\|^2$ . Then  $\forall x \in \mathcal{X}$  we have

$$\eta_t (f_t(x_t) - f_t(x)) \leq d_\psi(x, x_t) - d_\psi(x, x_{t+1}) + \frac{\eta_t^2}{2} \|\nabla f_t(x_t)\|_*^2$$

With  $\|\cdot\|_*$  we defined the dual norm with respect to the norm  $\|\cdot\|$

**Definition 2.3.3.** (Dual Norm). Let  $x \in X$ , the dual norm  $\|\cdot\|_*$  of a norm  $\|\cdot\|$  is defined as:

$$\|x\|_* = \sup_{y: \|y\| \leq 1} \langle x, y \rangle$$

---

**Algorithm 1** OMD Online Convex Optimization

---

**Require:** learning rate sequence  $\{\eta_1, \dots, \eta_T\}$

- 1: Set  $\mathbf{x}_1 \leftarrow \frac{1}{M} \mathbf{1}$
  - 2: **for**  $t \in \{1, \dots, T\}$  **do**
  - 3:   Observe  $f_t(x_t)$  decided by the adversary
  - 4:    $x_{t+1} = \arg \min_{x \in \mathcal{X}} d_\psi(x, x_t) + \eta_t \langle \nabla f_t(x_t), x - x_t \rangle$
  - 5: **end for**
- 

Theorem 2.3.1 can be used to prove a regret bound for the general OMD algorithm.

**Theorem 2.3.2.** (*Regret Bound for Online Mirror Descent*). Together with the assumptions of Theorem 2.3.1 and  $\eta_t \geq 0$  be a decreasing sequence of learning rates, then we have

$$R_T \leq \max_{t \leq T} \frac{d_\psi(x, x_t)}{\eta_T} + \frac{1}{2\alpha} \sum_{t=1}^T \eta_t \|\nabla f_t(x_t)\|_*^2$$

By choosing  $\eta_t = \frac{D\sqrt{\alpha}}{\sqrt{\sum_{t=1}^T \|\nabla f_t(x_t)\|_*^2}}$ , where  $D = \max_{t \leq T} d_\psi(x, x_t)$ , we have a bound for the OMD algorithm of

$$R_T \leq \frac{2D}{\sqrt{\alpha}} \sqrt{\sum_{t=1}^T \|\nabla f_t(x_t)\|_*^2}$$

If the gradient under the dual norm is bounded by  $\|\nabla f_t(x_t)\|_* \leq G \forall t \leq T$  then we have that  $R_T \leq \frac{2DG}{\sqrt{\alpha}} \sqrt{T}$  which is sublinear in  $T$ .

The OMD algorithm is a general technique to exploit the geometric convexity of the problem and gives rise to Hannan Consistent strategies in the case of uncountable convex decision spaces. By specializing the loss function and the Bregman divergences we can generate many algorithms that are state of the art in the Online Convex optimization problem, and achieve better theoretical guarantees than the general analysis we saw for the OMD algorithm.

### 2.3.2 Geometric Convexity and Why OMD works

The reason for OMD to work is not that we are following the gradient, that points to the minimum of the function (the gradient of a loss function

better title

does not point to the minimum in general). In reality the reason why OMD and other first order methods are effective is because of the convexity of the loss function and because of the following inequality for the instantaneous regret of convex loss functions:

$$f_t(x_t) - f_t(x) \leq \langle \nabla f_t(x_t), x_t - x \rangle ; \quad (2.19)$$

and so to minimize the left hand side of Equation (2.19) we can minimize the right hand side of Equation (2.19). Minimizing strictly a linear approximation of the instantaneous regret is not ideal since the environment is adversarial. Instead we minimize the linear approximation together with a regularization term which is given by the Bregman divergence  $d_\psi$ .

### 2.3.3 From Online Learning to Statistical Learning

Now we explore the connection between the Online Optimization framework and classical concepts of classical Statistical Learning techniques. More concretely we can prove and design a whole class of algorithm that are Agnostically PAC Learnable with Online Learning Techniques. Classical statistical learning theory deals with examples (or observations) and models of the phenomena. Then it uses the model to predict the future observations [Bousquet et al., 2003]. Quite informally one could say that we are trying to infer concept from examples. A concept is a map  $\mathcal{C} : \mathcal{X} \rightarrow \mathcal{Y}$ , where  $\mathcal{X}$  is the domain space and  $\mathcal{Y}$  is the set of labels for the examples. We then observe a sample from an unknown distribution  $\mathcal{D}$  such that  $(x, y) \sim \mathcal{D}$ . What we need to achieve is to learn a mapping  $y : \mathcal{X} \rightarrow \mathcal{Y}$  such that the error under the distribution  $\mathcal{D}$  is small. The loss function needed to define this error is not specific to the problem and can be decided by the user, this is called generalization error and, for a loss function  $l : \mathcal{Y} \times \mathcal{Y} \rightarrow \mathbb{R}$ , it is defined as:

$$e(h) = \mathbb{E}_{(x,y) \sim \mathcal{D}}[l(h(x), y)] \quad (2.20)$$

The goal for an algorithm  $\mathcal{A}$  is to produce a hypothesis  $h$  with small generalization error. It is generally difficult to generalize well and how difficult is clarified by the following theorem called the *No free lunch theorem* [Mitchell et al., 1997]. This restriction gives raise to the concept of Probably Approximately Correct (PAC) learnability.

**Definition 2.3.4.** (PAC learnable). An hypothesis class  $\mathcal{H}$  is PAC learnable w.r.t. the loss  $l$  if there exists a learner  $\mathcal{A}$  that given a sample  $S_N$  of examples learns an hypothesis  $h \in \mathcal{H}$  s.t. for all  $\epsilon, \delta$  there exists  $N_{\epsilon, \delta}$  such that for any distribution  $\mathcal{D}$  we have a generalization error  $\mathbb{P}[e(h) < \epsilon] \geq 1 - \delta$

add example  
for the function  
 $f(x) = \max(x_1^2 + (x_2 + 1)^2, x_2^2 + (x_1 + 1)^2)$

Fai tutti i periodi così  
ma NON si può fare.  
O metti ; o . oppure  
devi usare 'which'

Usually we also require that the algorithm  $\mathcal{A}$  learns the concept  $h$  in polynomial time w.r.t. the parameter of the problem.

An example of such learning problems could be the classification of spam emails. In this case  $\mathcal{X}$  is the vectorial representation of the text and  $\mathcal{Y} = \{0, 1\}$ , indicating whether or not the email is a spam or not. If we choose as a model a linear classifier then the hypothesis space is  $\mathcal{H} = \{h = \mathbb{I}[\langle x, w \rangle \geq 1/2]\}$  and the loss could be chosen as  $l(y_1, y_2) = |y_1 - y_2|$ .

PAC learnability is intuitively requiring that there exists an hypothesis  $h \in \mathcal{H}$  with near zero generalization error, otherwise the class  $\mathcal{H}$  is not PAC learnable, otherwise the class  $\mathcal{H}$  is not PAC learnable. But we can weaken the concept of PAC learnability by addressing directly this issue.

**Definition 2.3.5.** (*PAC agnostic learnable*). Given the same definitions of Definition 2.3.4, an hypothesis class  $\mathcal{H}$  is PAC agnostic learnable if we have a generalization error  $\mathbb{P}[e(h) < \inf_{\tilde{h} \in \mathcal{H}} e(\tilde{h}) + \epsilon] \geq 1 - \delta$

Which hypothesis spaces  $\mathcal{H}$  are PAC learnable (agnostically or not) is an open and complex issue, but the case for convex hypothesis class  $\mathcal{H} \subset \mathcal{R}$  can be solved by Online Learning techniques, showing the versatility of the methods. Moreover approach to prove such theorem gives a constructive methodology to solve agnostic PAC learnable problems.

**Theorem 2.3.3.** For every hypothesis class  $\mathcal{H}$  and bounded loss function  $l : \mathcal{Y} \times \mathcal{Y} \rightarrow \mathbb{R}$ , for which does exists a low regret algorithm  $\mathcal{A}$ , then the problem is agnostic PAC learnable. In particular these conditions are satisfied if the hypothesis space  $\mathcal{H}$  and the loss function  $l$  are convex.

*Proof.* (Sketch). Initialize the learner with the hypothesis  $h_0 = \mathcal{H}$ . For every iteration  $t \leq T$ : observe a sample  $(x_t, y_t) \sim \mathcal{D}$  and a loss function  $l_t := l(h_t(x_t), y_t)$ . Then update the hypothesis  $h_{t+1} = \mathcal{A}(l_1, \dots, l_t)$ .

At  $t = T$  return  $\bar{h} = \frac{1}{T} \sum_{t=1}^T h_t \in \mathcal{H}$ .

The proof then continues by defining the random variable  $X_T^{(1)} = \sum_{t=1}^T e(h_t) - l(h_t(x_t), y_t)$ . This is a martingale and  $\mathbb{E}[X_T^{(1)}] = 0$ . Moreover  $|X_T^{(1)} - X_{T-1}^{(1)}| < K$  since the loss function  $f$  is bounded. We can normalize the losses so that  $K = 1$ , and then apply the Azuma martingale inequality  $\mathbb{P}[X_T^{(1)} > c] \leq e^{-\frac{c^2}{2T}}$ .

For an appropriate choice of  $c$  we get

$$\mathbb{P} \left[ \frac{1}{T} \left[ \sum_{t=1}^T e(h_t) - l(h_t(x_t), y_t) \right] > \sqrt{\frac{2 \log(\delta/2)}{T}} \right] \leq \delta/2 \quad (2.21)$$

defining  $h^* = \arg \inf_{h \in \mathcal{H}} e(h)$  and  $X_T^{(2)} = \sum_{t=1}^T e(h^*) - l(h^*(x_t), y_t)$  we can obtain

$$\mathbb{P} \left[ \frac{1}{T} \left( \sum_{t=1}^T e(h^*) - l(h^*(x_t), y_t) \right) < -\sqrt{\frac{2 \log(\delta/2)}{T}} \right] \leq \delta/2 \quad (2.22)$$

By the definition of regret  $y_T$  we obtain

$$\frac{1}{T} \sum_{t=1}^T e(h_t) - e(h^*) = y_T/T + X_T^{(1)} - X_T^{(2)} \quad (2.23)$$

and from inequalities (2.21), (2.22) and from Equation (2.23) we have:

$$\mathbb{P} \left[ \frac{1}{T} \sum_{t=1}^T e(h_t) - e(h^*) > \frac{y_T}{T} + 2\sqrt{\frac{2 \log(\delta/2)}{T}} \right] \leq \delta \quad (2.24)$$

Now simply thanks to the linearity of the error operator  $e : \mathcal{H} \rightarrow \mathbb{R}$  we have that

$$\mathbb{P} \left[ e(\bar{h}) < e(h^*) + y_T/T + 2\sqrt{\frac{2 \log(\delta/2)}{T}} \right] \leq 1 - \delta$$

and since  $y_T/T \rightarrow 0$  we can find  $\tilde{T}$  large enough such that the thesis is verified.  $\square$

This result has been presented since it is useful to prove the general behavior of Hannan **C**onsistent strategies in environments driven by a stationary distribution.

## Chapter 3

# Information, Prediction and Investing

In Chapter 2 we described at a high level the framework for Online Learning in Adversarial environment. Now we draw its connections with predictions and investments. It <sup>surely</sup> seems counter-intuitive to speak about predictions in an adversarial framework, since we are used to think about predictions only of stochastic processes, but the way to think about predictions in adversarial environments is to think about probability assignment and empirical frequencies. The roots of this formulation are to be traced back to the Bell Laboratories in the '50s from works of Kelly [Kelly Jr, 2011], linking sequential betting and concept from information theory [Cover and Thomas, 2012]. This connection is of paramount importance to understand sequential investing as an instance of sequential decision problem. We <sup>will</sup> first draw the parallelism between probability assignment over discrete events and Online Learning, and then extend the discussion to sequential investments.

### 3.1 Probability assignment

The decision space  $\mathcal{D}$  in the case of finite  $N$  possible bets is the  $\Delta_{N-1} \subset \mathbb{R}^N$  probability simplex while the outcome  $\mathcal{Y}$  space is the set  $\{1, \dots, N\}$ , representing the winning bet at each turn. The loss function  $f(x, y)$  should have these natural properties: low when  $x_y \sim 1$  and high when  $x_y \sim 0$ , where  $x_y$  is the probability assigned to the outcome  $y$ . The inverse log-likelihood seems a reasonable proposal, simply because the multiplicative additive property of the logarithm, but has also a deeper connection to information that we will discuss later on:

dimmi cosa stai facendo

è un po vago

**Definition 3.1.1.** (*Self Information Loss*). In the sequential probability

assignment problem the loss function  $f(x, y)$ ,  $x \in \Delta_{N-1}$  and  $y \in [1, \dots, N]$  is defined as

$$f(x, y) = -\log \left( x^{(y)} \right)$$

where  $x^{(y)}$  is the probability assigned to outcome  $y \in \mathcal{Y}$ .

In the case of simulable experts, the prediction  $x_t$  of the agent is a function of the history of outcomes  $y^{t-1} := \{y_1, y_2, \dots, y_{t-1}\}$ .

The cumulative loss for the agent  $\mathcal{A}$  is then given by

$$L_T = \sum_{t=1}^T f(x_t, y_t) \quad (3.1)$$

and can be interpreted as the log-likelihood assigned to the outcome sequence  $y^T$  since

$$L_T = \sum_{t=1}^T f(x_t, y_t) = -\log \left( \prod_{t=1}^T x_t^{(y_t)} \right) \quad (3.2)$$

where we can interpret  $\prod_{t=1}^T x_t^{(y_t)}$  as the probability assigned to the entire outcome sequence  $y^T$ . This is already very similar to the compression-entropy rate one encounters in a classical lossless encoder, such as the arithmetic encoder [Langdon, 1984]. We will explore the connections to information theory later on in the chapter.

ref esatta al ch

Similarly we can define the loss for an expert  $e \in \mathcal{E}$  as

$$L_{T,e} = \sum_{t=1}^T f(x_{t,e}, y_t) = -\log \left( \prod_{t=1}^T x_{t,e}^{(y_t)} \right) \quad (3.3)$$

and the regret for each expert  $e \in \mathcal{E}$  is defined as

$$R_{T,e} = L_T - L_{T,e} = \log \left( \prod_{t=1}^T x_{t,e}^{(y_t)} / \prod_{t=1}^T x_t^{(y_t)} \right) \quad (3.4)$$

and the regret w.r.t. a generic class  $\mathcal{E}$  of experts is defined as

$$R_T = \sup_{e \in \mathcal{E}} \log \left( \prod_{t=1}^T x_{t,e}^{(y_t)} / \prod_{t=1}^T x_t^{(y_t)} \right) \quad (3.5)$$

where the class of experts  $\mathcal{E}$  can be finite or uncountable.



### 3.1.1 Connection to Information Theory

The link between sequential predictions and information theory has been observed in [Kelly Jr, 2011], and connects the concept of sequential betting (or predictions) and entropy.

Kelly put himself in a setting where the bettor has to predict the outcomes of binary events, given private information from an *information channel* prone to errors, the binary events pays double for a correct prediction and zero for an incorrect one. The input bits of the information channel are the correct outcomes of the binary sequential event, but they reach the end of the private channel with probability  $p$  of being correct and  $q = 1 - p$  of being wrong. Clearly the optimal strategy with  $p = 1$  is to bet everything on each turn reaching a final wealth, after  $T$  rounds, of  $V_T = 2^T$ . In case  $p < 1$  it is not clear which strategy is the best to follow, this is clearly related and still under philosophical debate as the St. Petersburg paradox [Samuelson, 1977]. Kelly proposed to maximize the grow rate of the wealth by investing a constant fraction of the current wealth. The growth rate  $G$  of the wealth  $V_T$  is defined as

$$G = \lim_{T \rightarrow +\infty} \frac{1}{T} \log_2(V_T)$$

Calling  $l \in [0, 1]$  the fraction of the wealth invested in the bet we have a capital after  $T$  rounds of

$$V_T = (1 + l)^W (1 - l)^{T-W}$$

and the associated growth rate is

$$G = p \log_2(1 + l) + q \log_2(1 - l)$$

which is maximized for  $l = p - q$  giving  $G_{\max} = 1 + p \log_2(p) + q \log_2(q)$  which is the rate of transmission for the communication channel, *i.e.* the number of bits transferred for unit of time. This is the trivial case and can be extended to arbitrary odds and number of bets.

The equivalent formulation in Online Learning can be obtained by observing that  $\mathcal{D} = \Delta_0$  and that we are betting a fraction  $l_t$  on the event being 0 and a fraction  $1 - l_t$  on the the outcome being 1. In that case the wealth at time  $t$  will be  $V_t = V_{t-1} l_t^{1-y_t} (1 - l_t)^{y_t}$  and hence:

$$\log(V_T) = \sum_{t=1}^T \log(l_t(y_t - 1) + (1 - l_t)y_t) \quad (3.6)$$

sembra che  
manchi una  
connessione in  
questa frase

which is equivalent to defining the cumulative loss

$$L_T = -\log(V_T) = \sum_{t=1}^T -\log(l_t(1 - y_t) + (1 - l_t)y_t)$$

which is equivalent to the loss defined in Equation (3.2).

By defining the growth rate at  $T$  as  $G_T = \frac{1}{T} \log_2(V_T)$  we can observe that  $L_T = TG_T \log(2)$  and so a learner  $\mathcal{A}$  that obtains sub-linear regret  $R_T/T \rightarrow 0$ , where the expert class is composed of constant experts for which  $l_t = \text{const}$ , is equivalent to obtaining a growth rate  $G_T \rightarrow G_{\max}$ .

This draws the connection to information rate as defined by Shannon in terms of information bits and growth rate of a betting strategy, and the fact that an Hannan Consistent strategy is able to converge to the highest growth rate.

### 3.1.2 Horse Races

In this section we will see how sequential investment is equivalent to the problem of sequential betting discussed in the previous section. In the previous chapter we saw ~~that~~ how to formalize sequential betting in the simple case of doubling odds and binary outcomes into the Online Learning formulation. Now we will extend the model to account variable odds and multiple bets, and how this is connected to investing.

Let us model horse races as a sequential betting process, in which we have  $N$  horses each paying a payoff of  $o_{t,i} \forall i \in 1, \dots, N$ . The agent  $\mathcal{A}$  splits its wealth into  $N$  separate betting by choosing an element of the simplex  $\Delta_{N-1}$ .

The wealth of the agent  $\mathcal{A}$  at time  $t$  will be the  $V_t = V_{t-1} \langle \mathbf{x}_t, \mathbf{y}_t \rangle$ , where  $\mathbf{y}_t = o_{y_t} \mathbf{e}_{y_t} \in \mathbb{R}^N$ , *i.e.* the basis vector with 1 as the  $y_t \in 1, \dots, N$  component, which represents the winning horse for the turn, and  $o_{y_t}$  is the payout of the bet at time  $t$ , on the  $t_y$  horse winning. As we did in the previous section we can apply  $-\log(\cdot)$  so that we can embed the problem into an Online Learning framework. By defining

$$L_T = -\log(V_T) = -\log(V_{T-1}) - \log(\langle \mathbf{x}_T, \mathbf{y}_T \rangle),$$

that implies

$$L_T = \sum_{t=1}^T -\log(\langle \mathbf{x}_t, \mathbf{y}_t \rangle) \tag{3.7}$$

we obtain exactly the same formulation presented at the beginning of the chapter. Moreover, we note that the regret  $R_T$  does not depend on the value of the payout  $o_{y_t}$ .

We saw in Section 2.1.4 that Theorem 2.1.3 assures that we have a sublinear regret  $R_T = \mathcal{O}(\log T)$  in case that the expert class  $\mathcal{E}$  is being generated by the convex hull of finite basic experts  $\mathcal{E}_N$ , which in this case can be taken as the  $N$  experts always predicting  $\mathbf{x}_{t,j} = \mathbf{e}_j, \forall j \in 1, \dots, N$ . The convex hull generated by  $\mathcal{E}_N$  is then composed by experts predicting a constant element of the simplex  $\mathbf{x}_{t,e} = \mathbf{x}_e \in \Delta_{N-1}$ .

Theorem 2.1.3 is stating that we can obtain asymptotic wealth equivalent to the one obtained by the best expert in hindsight, for all sequences of outcomes.

A very similar formulation can be obtained for the case of sequential investments. In the case of horse races we have just one winner for each day, while in the case of stock investing we have a different payout for each stock. In the following section we will present how to model sequential decision problems in the Online Learning formulation.

## 3.2 From Horse Races to Online Portfolio Optimization

We can formulate the portfolio allocation as a sequential betting problem. Let us imagine that there are no real life issues associated with trading costs and liquidity (they will be discussed in the following chapters). Then the best strategy would be to invest at each round  $t$  the entire capital on one single stock, knowing that will be the best performance stock at round  $t$ . But assuming an adversarial environment we have to randomize our allocation, or equivalently distribute our wealth accordingly to our randomization probabilities as in Equation (3.7).

agg ref

F:???

### 3.2.1 The Online Portfolio Optimization Model

The model consists in a sequential wealth allocation in  $N \in \mathbb{N}$  stocks for discrete rounds  $t \in \{1, \dots, T\}$ , where  $T$  is the investment horizon. Note that the set of times is arbitrary, and could also be non-homogeneous—in practice in the Online Portfolio Optimization case, it is usually thought to be in days. The evolution of the stock  $i \in 1, \dots, N$  prices at time  $t$ ,  $P_{t,i}$ , define the price relatives  $r_{i,t} = \frac{P_{i,t+1}}{P_{i,t}}$ , and we can define the price relative vector at time  $t$  as  $\mathbf{r}_t = (r_{1,t}, \dots, r_{N,t}) \in \mathbb{R}^N$ .

An investor dividing at round  $t$  its wealth  $W_t$  into a fraction  $\mathbf{x}_t \in \Delta_{N-1}$  for each asset will get a wealth  $W_{t+1} = W_t \langle \mathbf{x}_t, \mathbf{r}_t \rangle$  at round  $t + 1$ . As in

Section 3.1.1 we can define the growth rate

$$G_T = \log(W_T) = \sum_{t=1}^T \log(\langle \mathbf{x}_t, \mathbf{r}_t \rangle)$$

As in the case of binary outcomes, *i.e.* horse races, we can redefine the problem in an Online Learning framework, by defining the loss  $f(\mathbf{x}, \mathbf{y}) = -\log(\langle \mathbf{x}_t, \mathbf{y}_t \rangle)$  and a cumulative loss as

$$L_T = -G_T = \sum_{t=1}^T -\log(\langle \mathbf{x}_t, \mathbf{y}_t \rangle)$$

Exactly as in the previous Section, the expert class is generated by the convex hull of the base class  $\mathcal{E}_N$ , composed by the experts always betting on the win of the same horse  $i \in 1, \dots, N$ , or, equivalently, allocating all the portfolio on the same asset, at every round. The convex hull of the class is the class of experts  $\mathcal{E}$ , so that at every turn  $t$  the expert is allocating all the wealth in a specific element  $\mathbf{x} \in \Delta_{N-1}$ . In the Online Portfolio literature this class of allocations is called *Constant Rebalancing Portfolio (CRP)*.

As in every adversarial environment, we have to compare our losses with the best expert in the expert class, through the concept of regret:

$$R_T = L_T - \inf_{e \in \mathcal{E}} L_{T,e} \quad (3.8)$$

$$= \sum_{t=1}^T -\log(\langle \mathbf{x}_t, \mathbf{y}_t \rangle) - \inf_{\mathbf{x} \in \Delta_{N-1}} \sum_{t=1}^T -\log(\langle \mathbf{x}, \mathbf{y}_t \rangle) \quad (3.9)$$

The CRP attaining the minimum loss

$$\mathbf{x}^* = \inf_{\mathbf{x} \in \Delta_{N-1}} \sum_{t=1}^T -\log(\langle \mathbf{x}, \mathbf{y}_t \rangle)$$

is called *Best Constant Rebalancing Portfolio (BCRP)*.

As we shall see in the next Section, Constant Rebalancing Portfolios (CRP) are a very powerful class of strategies and being competitive (in terms of sublinear regret) with respect to this class assures good theoretical guarantees.

### 3.2.2 Effectiveness of Constant Rebalancing Portfolios

The CRP is a strategy that each round  $t$  redistributes its wealth into the same distribution  $\mathbf{x} \in \Delta_{N-1}$ . As we saw in the previous Section this strategies can be identified as the ones generated by expert class  $\mathcal{E}$  defined previously. The Buy and Hold (BAH) is a strategies that buys an allocation

cite

aggiungere algo  
(non necessario)  
per OPO

at the start of the investment period and hold on to it to the end of the investment horizon  $T$ . The wealth of an BHA strategy can be written as  $W_T = \langle \mathbf{x}, \prod_{t=1}^T \mathbf{r}_t \rangle$ .

A simple example can illustrate the effectiveness of the CRP strategies, and the inherently difference that exists between the Modern Portfolio Theory and the Online Portfolio Optimization techniques. Imagine to have two stocks, and the adversary can choose the value of the price relatives in the set:  $r_{1,t}, r_{2,t} \in \{\frac{3}{5}, \frac{8}{5}\}$ . Imagine that the adversary picks a price relative in the set  $\{\frac{3}{5}, \frac{8}{5}\}$  with equal probability. Every BHA allocation is exponentially decaying  $\mathbb{E}[W_T] = \langle \mathbf{x}, (\frac{24}{25}, \frac{24}{25}) \rangle = \frac{24}{25}$  and hence has decaying growth rate  $G_T < 0$ . Conversely, the equally allocated CRP  $\mathbf{x} = (\frac{1}{2}, \frac{1}{2})$  has positive growth rate and exponentially increasing wealth:  $\mathbb{E}[W_T] = (11/10)^T$  and  $G_T = T \log(11/10) > 0$ .

Historically, this example has been called Shannon Demon [Poundstone, 2010] and being compared to the Maxwell's Demon, since as in the thermodynamics case, the Shannon's Demon is generating wealth (energy in the case on Maxwell) from nothing since both stocks are martingales, and oppositors to the Capital Growth Theory, used this argument to invalidate this ideas. In reality there is nothing strange about this example, and it is just one of the many techniques that exploits the existence of volatility and convert it into wealth, as theoretically does a delta-hedged option in the Black and Scholes model [Black and Scholes, 1973].



## Chapter 4

# Algorithms for the Online Portfolio Optimization Problem

In this section we will review the state of the art algorithms for the Online Portfolio Optimization problem and discuss their theoretical guarantees.

### 4.1 Universal Portfolios

### 4.2 Exponential Gradient

### 4.3 Online Newton Step





# Bibliography

- [Agarwal et al., 2010] Agarwal, A., Bartlett, P., and Dama, M. (2010). Optimal allocation strategies for the dark pool problem. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, pages 9–16.
- [Banerjee et al., 2005] Banerjee, A., Merugu, S., Dhillon, I. S., and Ghosh, J. (2005). Clustering with bregman divergences. *Journal of machine learning research*, 6(Oct):1705–1749.
- [Belmega et al., 2018] Belmega, E. V., Mertikopoulos, P., Negrel, R., and Sanguinetti, L. (2018). Online convex optimization and no-regret learning: Algorithms, guarantees and applications. *arXiv preprint arXiv:1804.04529*.
- [Black and Scholes, 1973] Black, F. and Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of political economy*, 81(3):637–654.
- [Blackwell et al., 1956] Blackwell, D. et al. (1956). An analog of the minimax theorem for vector payoffs. *Pacific Journal of Mathematics*, 6(1):1–8.
- [Bousquet et al., 2003] Bousquet, O., Boucheron, S., and Lugosi, G. (2003). Introduction to statistical learning theory. In *Summer School on Machine Learning*, pages 169–207. Springer.
- [Cesa-Bianchi and Lugosi, 2003] Cesa-Bianchi, N. and Lugosi, G. (2003). Potential-based algorithms in on-line prediction and game theory. *Machine Learning*, 51(3):239–261.
- [Cesa-Bianchi and Lugosi, 2006] Cesa-Bianchi, N. and Lugosi, G. (2006). *Prediction, learning, and games*. Cambridge university press.
- [Cover, 1966] Cover, T. M. (1966). Behavior of sequential predictors of binary sequences. Technical report, STANFORD UNIV CALIF STANFORD ELECTRONICS LABS.

- [Cover and Thomas, 2012] Cover, T. M. and Thomas, J. A. (2012). *Elements of information theory*. John Wiley & Sons.
- [Hakansson et al., 1995] Hakansson, N. H., Ziemba, W. T., et al. (1995). Capital growth theory. *Handbooks in operations research and management science*, 9:65–86.
- [Hannan, 1957] Hannan, J. (1957). Approximation to bayes risk in repeated play. *Contributions to the Theory of Games*, 3:97–139.
- [Hart and Mas-Colell, 2001] Hart, S. and Mas-Colell, A. (2001). A general class of adaptive strategies. *Journal of Economic Theory*, 98(1):26–54.
- [Jiang et al., 2016] Jiang, Y., Yuan, J., and Zeng, M. (2016). A Game Theoretic Study of Enterprise Mergers and Acquisitions: The Case of RJR Nabisco Being Acquired by KKR. *Business and Management Studies*, 2(2):21–33.
- [Kelly Jr, 2011] Kelly Jr, J. L. (2011). A new interpretation of information rate. In *The Kelly Capital Growth Investment Criterion: Theory and Practice*, pages 25–34. World Scientific.
- [Langdon, 1984] Langdon, G. G. (1984). An introduction to arithmetic coding. *IBM Journal of Research and Development*, 28(2):135–149.
- [Mitchell et al., 1997] Mitchell, T. M. et al. (1997). Machine learning.
- [Poundstone, 2010] Poundstone, W. (2010). *Fortune’s Formula: The untold story of the scientific betting system that beat the casinos and wall street*. Hill and Wang.
- [Samuelson, 1977] Samuelson, P. A. (1977). St. petersburg paradoxes: Defanged, dissected, and historically described. *Journal of Economic Literature*, 15(1):24–55.
- [Shalev-Shwartz et al., 2012] Shalev-Shwartz, S. et al. (2012). Online learning and online convex optimization. *Foundations and Trends® in Machine Learning*, 4(2):107–194.
- [Srebro et al., 2011] Srebro, N., Sridharan, K., and Tewari, A. (2011). On the universality of online mirror descent. In *Advances in neural information processing systems*, pages 2645–2653.
- [Sutton et al., ] Sutton, R. S. et al. *Introduction to reinforcement learning*, volume 135.

- [Van Erven and Harremos, 2014] Van Erven, T. and Harremos, P. (2014). Rényi divergence and kullback-leibler divergence. *IEEE Transactions on Information Theory*, 60(7):3797–3820.
- [Zinkevich, 2003] Zinkevich, M. (2003). Online convex programming and generalized infinitesimal gradient ascent. In *Proceedings of the 20th international conference on machine learning (icml-03)*, pages 928–936.