Online learning through the lens of potential descent

Kamron Saniee (kns2110) and Navid Ardeshir (na2844)

April 2020

1.0 Introduction

We discuss theoretical topics at the interface of online learning and repeated games. In the first two sections we motivate the Blackwell Approachability Theorem and its generalization by potential-based approachability. We then shed light on Vovk's mixability condition for loss functions and applications of regret bounds to auction pricing. We connect these ideas to the subject of Correlated Equilibria in game theory by proving existence of a calibrated forecaster in the appendix. An overarching principle throughout our discussion is given by the minimization of potential functions which measure the forecaster's divergence from her goal. Much of the material is inspired by [PLG] as well as the papers [MSC, ABH]. We hope this report can serve as a focused introduction to ideas that we found insightful in this area.

1.1 Minimax Theorem and Blackwell Approachability

One interesting connection between learning theory and game theory is that the classical Minimax Theorem of von Neumann can be used to show the existence of a boosting algorithm. In game-theoretic language, this theorem guarantees the row player a mixed strategy for the zero-sum game with loss matrix A which limits her loss – equivalently, the column player's payoff – to the minimax loss (the "value of the game").

Theorem 1. *(Minimax Theorem, von Neumann (1929)) Let $A \in \mathbf{R}^{n\times m}$ and let $\triangle(S)$ denote the set of probability distributions on a set S. Then we have the following:*

$$\min_{p\in\triangle([n])} \max_{h\in[m]} p^\top A e_h = \max_{w\in\triangle([m])} \min_{i\in[n]} e_i^\top A w$$

Under the assumption of "no-regret" learning (defined below), a slight generalization of the minimax theorem can be shown as follows. The following argument appears in [ABH]. Consider a repeated game in which the row player (resp. column player) makes the decision $x_t \in \mathcal{X}$ ($y_t \in \mathcal{Y}$) at round t, and define the *regret* incurred by the row player as

$$\text{Regret}_T^{\mathsf{row}} = \sum_{t=1}^{T} x_t^\top A y_t - \min_x x^\top \sum_{t=1}^{T} A y_t.$$

The regret of the column player is similarly

$$\text{Regret}_T^{\mathsf{col}} = \sum_{t=1}^{T} x_t^\top (-A) y_t - \min_y \sum_{t=1}^{T} x_t^\top (-A) y = \max_y \sum_{t=1}^{T} x_t^\top A y - \sum_{t=1}^{T} x_t^\top A y_t.$$

Suppose that both players are able to make decisions such that their regrets are $o(T)$ (such strategies are termed *no-regret*). Since the sets $\mathcal{X}, \mathcal{Y}$ are convex and compact (they are probability simplices on the players actions), they are closed under taking averages, and extrema of continuous functions on $\mathcal{X} \times \mathcal{Y}$ are achieved. It follows that

$$\frac{1}{T}\sum_{t=1}^{T} x_t^\top A y_t = \frac{\text{Regret}_T^{\mathsf{row}}}{T} + \min_x x^\top A(\frac{1}{T}\sum_{t=1}^{T} y_t) \le o(1) + \max_y \min_x x^\top A y$$

and likewise

$$\frac{1}{T}\sum_{t=1}^{T} x_t^\top A y_t = \max_y (\frac{1}{T}\sum_{t=1}^{T} x_t)^\top A y - \frac{\text{Regret}_T^{\mathsf{col}}}{T} \ge \min_x \max_y x^\top A y - o(1)$$

which gives $\max_y \min_x x^\top A y \ge \min_x \max_y x^\top A y$. Also,

$$\begin{aligned} \min_x x^\top A y \le x^\top A y \quad \forall x \in \mathcal{X}, y \in \mathcal{Y} &\implies \max_y \min_x x^\top A y \le \max_y x^\top A y \quad \forall x \in \mathcal{X} \\ &\implies \max_y \min_x x^\top A y \le \min_x \max_y x^\top A y \end{aligned}$$

so $\max_y \min_x x^\top A y = \min_x \max_y x^\top A y$. Thus, the classical minimax theorem can be recovered from the existence of no-regret strategies for repeated games.

The significance of the minimax theorem historically motivated the question of what can be said when the losses are vectors in $\mathbf{R}^d$ instead of scalar. This can model the situation of multiple loss types, as well as the N-experts problem (Example 2 below). In the work of Blackwell [B], an especially useful condition was identified as whether or not the row player can guarantee that the euclidean distance of her average loss vector to a desired subset of $\mathbf{R}^d$ converges to zero. Accordingly, we have the following definition:

Definition 1. *$S \in \mathbf{R}^d$ is called* approachable *[PLG], §7.7 if a strategy exists for the row player such that, for any sequence of plays by the column player, we have*

$$\lim_{T \longrightarrow \infty} d(\frac{1}{T}\sum_{t=1}^{T} \ell(I_t, J_t), S) = 0 \quad a.s.$$

As mentioned in Blackwell's original paper, the notion of approachability was motivated by the following property of repeated zero-sum games. We will show that no matter how the column player plays, the row player's minimax strategy guarantees that the average loss asymptotically approaches to the set $(-\infty, V]$ almost surely (Ex. 7.7 in [PLG]), where V is the value of the game.

Theorem 2. *In a repeated zero-sum game with scalar loss matrix A, the set $(-\infty, v]$ is approachable if and only if $v \geq V$.*

Proof. Let I_t, J_t denote the realized row and column player actions at time t, and $\mathbf{E}_t[\cdot]$ denote the expectation conditional on the row player's past plays $I_1, \ldots, I_{t-1}$ and on J_t. Suppose the row player chooses her action according to the minimax strategy x every round, and let $y_t \in \triangle([m])$ be the column player's mixed strategy at time t. Then by the minimax theorem,

$$\mathbf{E}_t \ell(I_t, J_t) = x^\top A e_{J_t} \leq \max_y x^\top A y =: V$$

and by the law of iterated expectation

$$\mathbf{E}\ell(I_t, J_t) = \mathbf{E}\mathbf{E}_t \ell(I_t, J_t)$$

which implies $\ell(I_t, J_t) - \mathbf{E}_t \ell(I_t, J_t)$ is a martingale difference sequence bounded by $\max_{i,j} a_{ij} - \min_{i,j} a_{ij} =: c$. By the Azuma-Hoeffding inequality, we have for any $\epsilon > 0$

$$\mathbf{P}\left(\frac{1}{T}\sum_{t=1}^{T} \ell(I_t, J_t) - \frac{1}{T}\sum_{t=1}^{T} \mathbf{E}_t \ell(I_t, J_t) > \epsilon\right) \leq e^{-2T\epsilon^2/c^2}$$

Since $\sum_{T=1}^{\infty} e^{-2T\epsilon^2/c^2} < \infty$, the Borel-Cantelli lemma implies that

$$\limsup \frac{1}{T}\sum_{t=1}^{T} \ell(I_t, J_t) - \frac{1}{T}\sum_{t=1}^{T} \mathbf{E}_t \ell(I_t, J_t) \leq \epsilon \quad a.s.$$

and thus

$$\limsup \frac{1}{T}\sum_{t=1}^{T} \ell(I_t, J_t) - V \leq \epsilon \quad a.s.$$

By applying the same argument from the perspective of the row player (conditioning on the column player's past, and bounding the left tail), the column player can guarantee that

$$\liminf \frac{1}{T}\sum_{t=1}^{T} \ell(I_t, J_t) \geq V - \epsilon \quad a.s.$$

□

Blackwell's approachability theorem provides necessary and sufficient conditions for approachability of convex sets in the case of vector-valued losses:

Theorem 3. *(Blackwell's approachability theorem). A convex set S is approachable if and only if every halfspace $H \supset S$ is approachable.*

Note that approachability of limit points is equivalent to approachability of S, so that S may be assumed closed and convex without loss of generality. The proof of this result essentially reduces to the case of scalar losses: indeed, by Theorem 2, a halfspace $H = \{u \in \mathbf{R}^d : \langle u, a\rangle \leq v\}$ parameterized by $a \in \mathbf{S}^{d-1}$ is approachable if and only if the value of the game with (scalar) loss matrix $\langle \ell(i,j), a\rangle$ is at most v. This, in turn, guarantees the existence of a mixed strategy which keeps the loss within the halfspace. Iterated orthogonal projections of the average loss vector onto S can then be used to construct a sequence of halfspaces containing S, and this way one obtains an algorithm for approaching S [cf. PLG, Thm. 7.5 for the proof].

Example 1. Suppose that the loss matrix is circulant, i.e., $\ell(i,j) = \ell(i+1 \mod n, j+1 \mod m)$, where n divides m. Then S is approachable if and only if S contains the "center of mass" $\frac{1}{nm}\sum_{i,j}\ell(i,j) \in \mathbf{R}^d$. By playing a uniform mixed strategy at every round, each player can ensure the sequence of losses is *i.i.d.* and uniform on the set of loss vectors, which ensures the average loss converges almost surely to the center of mass by the strong law of large numbers.

Example 2. (*N-Experts*) While the notion of vector-valued loss may at first seem unintuitive, the N-experts setting provides a case of special interest. In this case the "column player" corresponds to "nature" which selects an outcome $y_t \in [m]$ at time t. The forecaster chooses an expert I_t and incurs scalar loss $\ell(I_t, y_t)$, while the experts incur losses $\ell(k, y_t)$, $k = 1, \ldots, N$. By forming the vector-valued loss matrix with $(i,j)^{th}$ entry $(\ell(i,j) - \ell(1,j), \ldots, \ell(i,j) - \ell(N,j))$, the existence of a *Hannan consistent* (or no-regret) strategy is converted to the approachability of the nonpositive orthant $\mathbf{R}^N_- = \{u : u_1, \ldots, u_N \leq 0\}$, since the average loss in this game is the average regret vector

$$\frac{1}{t}\mathsf{R}_t = \frac{1}{t}\sum_{s=1}^{t}\left(\ell(I_s, y_s) - \ell(1, y_s), \ldots, \ell(I_s, y_s) - \ell(N, j)\right).$$

Note that the nonpositive orthant is particularly well structured for approachability, since it suffices to consider halfspaces passing through 0 whose normal vectors a lie in the positive orthant (any other halfspaces containing $\mathbf{R}^N_-$ are approachable if these halfspaces are). Also, while the proof of the approachability theorem does not in general suggest an explicit mixed strategy at time t (its existence is merely a consequence of the minimax theorem), in the case of orthant approachability, it can be calculated in simple form in terms of the normal vector of the corresponding halfplane divided by the sum of its components, as shown in the next section.

1.2 Potential Based Approachability

Hart and Mas Colell [MSC] used the same algorithmic idea in Blackwell's proof of the approachability theorem to come up with a class of strategies for approaching a convex set. As mentioned, the original proof relies on projection onto the target set, which is well defined and computationally feasible under the convexity assumption. Therefore, it is natural to ask what happens if we change the notion of "distance" from the current state to S. This would result in different projections onto the set, and thus, new strategies to approach a given set. The idea is to define a *potential function* on the space $\mathbf{R}^d$ and try to move in the directions that decrease the potential. This is analogous to an energy functional minimized by a physical system. Let $\Phi : \mathbf{R}^d \mapsto \mathbf{R}_+$ be a twice differentiable convex function which plays the role of potential and indicates how close we are from the equilibrium state. Since the goal is to reach the set S, we need to impose the condition that Φ vanish on S and take positive values outside of S. For instance, $\Phi(x) = \inf_{y \in S} ||y - x||_2$ satisfies these conditions and $\nabla\Phi(x)$ corresponds to the orthogonal projection of x onto S used in Blackwell's original proof. Now the main idea is to use strategies which keep the loss vector oriented towards S and then track progress of the potential function at every stage instead of $d(\frac{1}{T}\sum_{t=1}^{T} \ell(I_t, J_t), S)$, which is the quantity of interest. Recall that by the argument given in section 1.1, we have the following:

Corollary 1. A halfspace $H = \{u \in \mathbf{R}^d : \langle u, a\rangle \leq v\}$ is approachable if and only if there exists a mixed strategy $p \in \triangle([n])$ for the row player such that:

$$\max_{j \in [m]} \langle \ell(p, j), a\rangle \leq v$$

Definition 2. *(Bregman's Projection.) For every $x \notin S$ we can define a unique projection point on S:*

$$\Pi_S(x) := \arg\min_{y \in S} \Phi(y) - \Phi(x) - \langle \nabla\Phi(x), y - x\rangle = \arg\max_{y \in S} \langle y, \nabla\Phi(x)\rangle$$

Note that this projection is not necessarily based on a norm since the argument inside does not satisfy the triangle inequality; however, it has the properties that we need to replicate the algorithm in Blackwell's original proof.

Definition 3. *(Support of Convex Sets.) For every $x \in \mathbf{R}^d$ and convex set S we may define $u(x) := \max_{y \in S} \langle y, x\rangle$ to be the* support *function of S. Furthermore, we have $\nabla u(x) = \arg\max_{y \in S} \langle y, x\rangle$. It is also referred to as the conjugate function of the indicator of S.*

Let us consider hyperplane passing through $\Pi_S(x)$ with normal vector $\nabla\Phi(x)$. Since S is approachable, the existence of the following potential strategy $p \in \triangle(n)$ for every point $x \notin S$ is guaranteed by the Corollary 1.:

$$\max_{j \in [m]} \langle \ell(p, j), \nabla\Phi(x)\rangle \leq u(\nabla\Phi(x)) \qquad (\star)$$

It is worth mentioning that explicitly constructing such strategies p is problem dependent, but in some problems of interest, such as N-Experts and Calibration, the strategy can be easily calculated in terms of the gradient of Φ. We may define the current state $\mathbf{x}_t \in \mathbf{R}^d$ as the following:

$$\mathbf{x}_t := \frac{1}{t}\sum_{s=1}^{t} \mathbf{E}[\ell(I_s, J_s)|I_{1:s-1}, J_s] = \frac{1}{t}\sum_{s=1}^{t} \ell(p_s, J_s)$$

Where p_t is a strategy that satisfies $(\star)$ at $\mathbf{x}_{t-1}$. Looking at the evolution vector $\mathbf{x}_{t+1} - \mathbf{x}_t = \frac{1}{t}(\ell(p_t, J_t) - \mathbf{x}_t)$ which is shown in red in the figure below, one can easily observe that $(\star)$ implies this is a descending direction regardless of the opponent's action, which leads the average loss closer to the target set. Therefore, moving infinitesimally along this direction should decrease the potential.

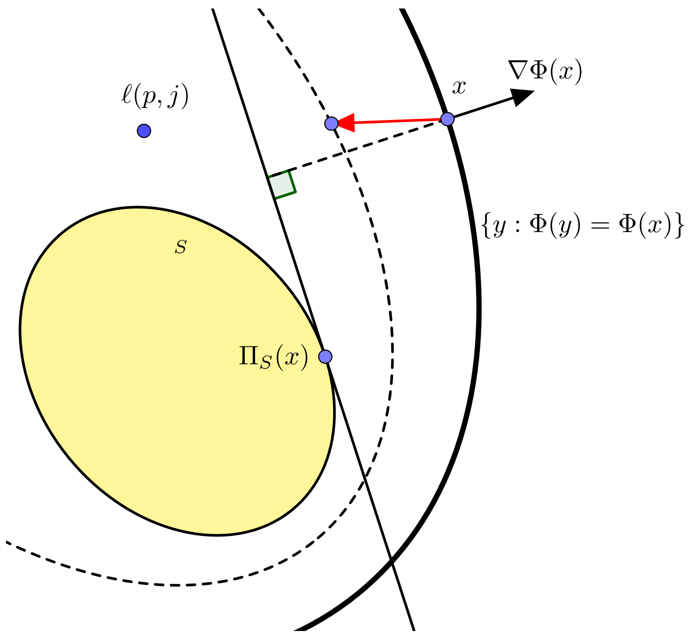

Figure 1: As an old saying "A picture is worth a thousand words"!

Theorem 4. *Let $S \subseteq \mathbf{R}^d$ be a closed convex set which is approachable. Let $\Phi : \mathbf{R}^d \mapsto \mathbf{R}_+$ be a twice differentiable convex function vanishing on S and positive outside of S. Then if the row player follows a potential based strategy which satisfies:*

$$\max_{j\in[m]} \langle \ell(p_t, j), \nabla\Phi(x_t)\rangle \leq u(\nabla\Phi(x_t))$$

Then regardless of the column player's actions:

$$\Phi(\frac{1}{T}\sum_{t=1}^{T} \ell(I_t, J_t)) = O(\frac{\log(T)}{T})$$

In particular, we have $d(\frac{1}{T}\sum_{t=1}^{T} \ell(I_t, J_t), S) \longrightarrow 0$ a.s.

The proof is basic and just uses standard convex analysis which can be found in [PLG] §7.8, [MSC], but relies heavily on the $\frac{1}{t}$ scaling of the cumulative loss. As a direct consequence of this powerful theorem, one can form a class of Hannan consistent strategies. To that end, one needs to prove approachability of the nonpositive orthant, or equivalently, approachability of halfspaces with normal vector in $\mathbf{R}^d_+$. Recall that $\underline{\ell}(i,j) := (\ell(i,j) - \ell(1,j), \ldots, \ell(i,j) - \ell(N,j))$ represents the loss vector in this context. By choosing p to be proportional to the normal vector $\nabla_k \Phi$, we can satisfy the condition in Corollary 1:

$$\begin{aligned}\langle \underline{\ell}(p,j), \nabla\Phi\rangle &= \sum_{i=1}^{N}\sum_{k=1}^{N} p_i(\ell(i,j) - \ell(k,j))\nabla_k\Phi \\ &= (\sum_{k=1}^{N} \nabla_k\Phi)(\sum_{i=1}^{N} p_i\ell(i,j)) - \sum_{k=1}^{N} \nabla_k\Phi\ell(k,j) = 0\end{aligned}$$

Note that Φ can be any convex function which is zero only on $\mathbf{R}^d_-$. Consequently, the gradient is an element of the positive orthant and therefore becomes a legitimate probability vector by the correct normalization.

1.3 Mixability of Loss Functions

A key principle in potential-based forecasting is that the value of the potential function at time t should not escape too far from the initial potential $\Phi(0)$. In the case of the exponential potential $\Phi_\eta(\mathsf{R}_t) = \frac{1}{\eta}\log\sum_i e^{\eta \mathsf{R}_{t,i}}$, preventing the potential from ever increasing ensures a regret bound uniform in the number of rounds t: we have

$$\max_i \mathsf{R}_{t,i} \le \Phi_\eta(\mathsf{R}_t) \le \Phi_\eta(0) = \frac{\log N}{\eta}$$

The property of nonincreasing potential can be seen to lead to the class of so-called *mixable* loss functions (and its subclass of *exp-concave* losses) as follows. Consider again the $N-$experts setting, and let

$$q_{i,t-1} = \frac{e^{-\eta L_{t-1,i}}}{\sum_j e^{-\eta L_{t-1,j}}},\ i = 1, \ldots, N$$

be weights given by the exponential potential, $L_{t,i} = \sum_{s\le t} \ell_{s,i}$ be the cumulative loss of the i^{th} expert at time t, $\mathsf{R}_{t,i} = L_t - L_{t,i}$ be the forecaster's regret with respect to the i^{th} expert, and r_t be the vector of instantaneous regrets $\ell_t - \ell_{t,i}$.

We have

$$\begin{aligned}
& \Phi_\eta(\mathsf{R}_{t-1}) \geq \Phi_\eta(\mathsf{R}_{t-1} + \mathsf{r}_t) \\
\iff \quad & \sum_i e^{\eta(L_{t-1} - L_{t-1,i})} \geq \sum_i e^{\eta(L_{t-1} - L_{t-1,i} + \ell_t - \ell_{t,i})} \\
\iff \quad & e^{-\eta \ell_t} \geq \sum_i e^{-\eta \ell_{t,i}} \frac{e^{-\eta L_{t-1,i}}}{\sum_j e^{-\eta L_{t-1,j}}} \\
\iff \quad & \ell_t \leq -\frac{1}{\eta} \log \sum_i e^{-\eta \ell_{t,i}} q_{i,t-1}.
\end{aligned}$$

The *exp-concave* losses are defined by concavity of $\exp(-\eta\ell(\cdot, y))$; this immediately ensures the above inequalities hold. From this angle, the notion of a *mixability curve* introduced by Vovk in [V] specifies general conditions under which this property, and consequently, uniform bounds, hold. Formally,

Definition 4. *The* mixability curve *of a loss function* $\ell : \mathcal{X} \times \mathcal{Y} \to \mathbf{R}$ *is defined as*

$$\begin{aligned}
\mu(\eta) := \inf \Big\{ \; & c > 0 : \forall N, \forall (q_1, \ldots, q_N) \in \triangle([N]) \quad \text{(the weights on each expert)}, \\
& \forall (I_1, \ldots, I_N) \in \mathcal{X} \qquad\qquad \text{(the expert predictions)}, \\
& \exists \; p : \text{no matter the outcome } y \in \mathcal{Y}, \\
& \ell(p, y) \leq -\frac{c}{\eta} \log \sum_i e^{-\eta \ell(I_i, y)} q_i \; \Big\}
\end{aligned}$$

Theorem 5. *Suppose the decision space* $\mathcal{X}$ *is compact, the loss function* ℓ *is continuous in its first argument,* $\exists\, I : \ell(I, y) < \infty \; \forall\, y$, *and* $\nexists\, I : \ell(I, y) = 0 \; \forall\, y$. *Then* $\mu(\eta) \geq 1$.

The proof of this result is in Vovk's paper and follows by contradiction: supposing $\exists\, c < 1$ satisfying the property in Definition 1, one can construct a sequence of decisions $(I^{(j)})_{j=1}^{\infty}$ such that $\lim \ell(I^{(j)}, y) = 0 \; \forall\, y$. By compactness, the limit point exists in $\mathcal{X}$, and by continuity of ℓ the limit point violates the last assumption. In terms of the parameterization above, it is also shown in [V] that $\mu(\eta)$ is a continuous and increasing function of η. It follows from the definition of mixability that we have

$$L_t \leq \mu(\eta) \min_i L_{t,i} + \frac{\mu(\eta)}{\eta} \log N \quad \forall t \geq 1$$

and so the largest η such that $\mu(\eta) = 1$ (if it exists for the ℓ in question) gives the best bound on the forecaster's loss/regret. It is also shown in the main theorem of [V] that the mixability curve $\eta > 0 \mapsto (\eta, \mu(\eta))$ in fact determines the best achievable constants in bounds of the above form. Examples of mixable losses are given in [PLG] §3.6 and are thus of special interest to the analysis of sharp regret bounds.

1.4 Auctions

An interesting application of the expert advice setting is in designing nearly optimal auctions. Myerson ([GTA], §14.9) characterised the single-item auction which optimizes the auctioneer's revenue, denoted $\text{OPT} \coloneqq \max_{p\in[0,1]} p(1-F(p))$. Each bidder's valuation for the item is drawn from the distribution F, which is known to the auctioneer (a practically questionable setting). Although one might think that achieving this optimal revenue is hopeless without enough information about bidder values, one can in fact learn this distribution sequentially through holding an auxiliary auction repeatedly and by changing the allocation and payment rule in each round in a clever manner so that the auctioneer has vanishing regret. In other words, the per-round revenue of the auctioneer will eventually be as though they knew the underlying distribution in hindsight and acted optimally. The main advantage of the latter auction is its robustness where there are no underlying statistical assumptions on the bidders' valuations. The idea is, instead of revealing the bids simultaneously, to hold private auctions with each bidder individually (note that this is only feasible for goods with infinite supply) and sell the item to the current bidder at that price if their bid was higher than our proposed price and then adjust our price for the next round based on the new obtained information. This is called the *Take-It-Or-Leave-It* mechanism. Usually, auction designers are constrained to design truthful auctions, meaning that bidders only bid their true valuation of the item, so that the auctioneer can obtain useful information to maximize their profit. Indeed, the proposed auction is truthful since no one has an incentive to buy an item higher than their valuation and they don't want to lose the chance to buy the item when the price is lower than their valuation. Formally:

Definition 5. *In a Take-It-Or-Leave-It online auction with n bidders, we perform the following at round t:*

- *The auctioneer computes a price $p_t \in [0,1]$, which can depend on past information denoted $\mathcal{F}_{t-1}$,*
- *The auctioneer and bidder at round t reveal their price and bid simultaneously. The item is sold at price p_t if the bid v_t is higher than p_t.*

The main idea to come up with pricing strategies is to discretize the set of prices and treat each one as an expert. Let us take $\mathbf{P} \coloneqq \{\frac{1}{N}, \frac{2}{N}, ..., 1\} \subseteq [0,1]$ to represent the set of our experts. In other words, expert i always suggests to use price $\frac{i}{N}$ for the item. Now the auctioneer revenue at round t is defined as $g(p_t, v_t) \coloneqq p_t \mathbf{1}\{v_t \geq p_t\}$ (note that this gain is not convex in its first argument) and the i^{th} expert obtains a gain equal to $r_{t,i} = \frac{i}{N}\mathbf{1}\{v_t \geq \frac{i}{N}\}$. This choice of gain is natural since we are trying to deduce something about the revenue of the auction. Now we can obtain regret bounds by adopting a polynomial weighting algorithm. The pricing strategy can be implicitly formulated by the following:

$$p_t = \frac{\sum_{i=1}^N w_i^t \frac{i}{N}}{\sum_{i=1}^N w_i^t}, \quad w_i^t = w_i^{t-1}(1 - \eta r_{t,i})$$

Theorem 6. *If the auctioneer sells the item in each round based on polynomial weighting pricing, then the auctioneer has no regret. In particular, if bidder valuations* $(\mathbf{v}_i)_{i\leq n}$ *are drawn independently from an underlying distribution* $F(\cdot)$ *then the average per round revenue will converge to the optimal revenue. Moreover, the rate of convergence can be characterized as follows:* $\sum_{t=1}^{n} g(p_t, \mathbf{v_t}) \geq n \max_{p\in[0,1]} p(1-F(p)) - O_p(\sqrt{n\log(n)})$

The ideas of this proof were gathered from [PM]. Based on regret bounds for polynomial weighting strategies [PLG] §4 for arbitrary bounded losses, we have:

$$\begin{aligned}\sum_{t=1}^{n} g(p_t, \mathbf{v_t}) &\geq \max_{p\in\mathbf{P}} \sum_{i=1}^{n} g(p, \mathbf{v_t}) - n\eta - \frac{\log(N)}{\eta} \\ &\geq \max_{p\in\mathbf{P}} p|\{i : \mathbf{v_i} \geq p\}| - n\eta - \frac{\log(N)}{\eta} \\ &\geq \sup_{p\in[0,1]} p|\{i : \mathbf{v_i} \geq p\}| - \frac{n}{N} - n\eta - \frac{\log(N)}{\eta}\end{aligned}$$

Where the last inequality comes from the fact that if we reduce the optimal price down to $\lfloor Np \rfloor / N$, then the price will be perturbed by at most $\frac{1}{N}$, and so the total revenue will be reduced by at most $\frac{n}{N}$. By choosing N, η optimally, we get the best lower bound one can get:

$$\sum_{t=1}^{n} g(p_t, \mathbf{v_t}) \geq \sup_{p\in[0,1]} p|\{i : \mathbf{v_i} \geq p\}| - O(\sqrt{n\log(n)})$$

Note that this regret bound holds for all valuation sequences (in particular, without statistical assumptions). Now, by combining the Rademacher Complexity bound for the class of indicators $\{p \mapsto \mathbf{1}\{v \geq p\}, v \in [0,1]\}$ having VC dimension 1 and Azuma-Hoeffding, we have the following with probability at least $1-\delta$:

$$\sup_{p\in[0,1]} |\sum_{i=1}^{n} \mathbf{1}\{\mathbf{v_i} \geq p\} - n(1-F(p))| \leq 2\sqrt{2n\log(2n)} + \sqrt{n\log(\frac{2}{\delta})}$$

To reiterate, the power of this algorithm stems from the information acquired by observing bidders' values for the item in each round so we can calculate what would have been our revenue had the auctioneer chosen different prices. This is equivalent to saying we observe the incurred loss for each individual expert. Similar results can be achieved in more limited feedback settings. The following problem arises in the context of dynamic pricing where instead of performing an auction, there is a price tag for the item at each round and customers buy the item in the same sequential manner with the slight difference that they won't reveal their values. In this setting the only information seller can obtain at each round is whether that internal value was higher or lower than the current price tag, which is analogous to the Multi-Armed Bandit problem, since only the loss of the current expert in use is observed.

1.5 Discussion

The idea for this report came from Professor Hsu's suggestion of the paper [V]. Subsequently, we found the book [PLG] and related papers [ABH, MSC]. For lack of time and space, we could not include material that interested us on the subjects of network routing, data compression, and computational issues that arise in implementation of online forecasting methods with large space complexity.

References

[PLG] Cesa-Bianchi, N. and Lugosi, G., 2006. Prediction, learning, and games. Cambridge University Press.

[MSC] Hart, S. and Mas-Colell, A., 1999. A general class of adaptive strategies. Journal of Economic Theory, 98(1), (pp. 26-54)

[ABH] Abernethy, J., Bartlett, P.L. and Hazan, E., 2011, December. Blackwell approachability and no-regret learning are equivalent. In Proceedings of the 24th Annual Conference on Learning Theory (pp. 27-46).

[V] Vovk, V., 1998. A game of prediction with expert advice. Journal of Computer and System Sciences, 56(2), pp.153-173.

[GTA] Karlin, A.R. and Peres, Y., 2017. Game theory, alive (Vol. 101). American Mathematical Soc.

[B] D. Blackwell, 1956. An analog of the minimax theorem for vector payoffs. Pacific Journal of Mathematics, 6(1), pp. 1–8.

[PM] Aaron Roth. 2020. Profit Maximization in Online Learning

Appendix: From Correlated Equilibria to Calibration

Although one can easily show for a repeated game that if both players adopt some Hannan consistent strategy, then the product empirical distribution of players actions denoted by $\hat{p}_t \times \hat{q}_t$ where $\hat{p}_{t,k} = \frac{1}{t}\sum_{i=1}^{t} \mathbf{1}_{\{\mathbf{I}_t = k\}}$ will converge to the set of Nash equilibria $p \times q$ almost surely, the empirical joint will not necessarily converge to that set. In fact, it will converge to a set called Hannan set [PLG] §7.4 which contains all the Correlated equilibria. This raises the question of whether the Hannan set is the smallest that we can jointly converge to. The answer is no, and by introducing a stronger notion of regret called internal regret we prove existence of strategies for players to jointly converge to the set of correlated equilibria contingent on controlling internal cumulative regret (see [PLG] §7.4). Internal Regret for the row player is defined as follows:

$$\hat{r}_{(i,i'),t} = \mathbf{1}_{\{I_t=i\}}(\ell(i,J_t) - \ell(i^{'},J_t))$$

In other words, cumulative internal regret $\hat{R}_{(i,i'),T} = \frac{1}{T}\sum_{t=1}^{T}\hat{r}_{(i,i'),t}$ measures the difference of what could have been the row player's loss had she played action $i^{'}$ every time he played i. Recall that correlated strategies in a 2 player game corresponds to the situation where every player obtains a suggested action from a vector (I,J) drawn from the joint distribution $\mathbf{P}$ and then decides which action to play. Equilibrium holds when no player has an incentive to deviate from the suggested action. Moreover, if row player knew $\mathbf{P}$ then $\mathbf{P}(J|I)$ would be the best guess for her to predict the column player's action. Therefore, $\mathbf{P}$ is a Correlated equilibrium if and only if action i is the best response to $\mathbf{P}(.|I=i)$. The goal is to ensure the empirical joint distribution converges to a Correlated Equilibrium but the issue is the limiting distribution is not known to the row player in advance. The remedy is to estimate it! In order to do this players should predict mixed strategies which their opponents are going to use in the next round and play the best action as if the opponent were using that predicted strategy. Let us denote the row player's prediction of the opponent's strategy at round t to be $\hat{q}_t$. Then the estimate of the joint distribution from the row player's point of view is:

$$\hat{\mathbf{P}}_T(i,j) = \frac{1}{T}\sum_{t=1}^{T}\mathbf{1}_{\{\hat{q}_t\in\hat{B}_i\}}\mathbf{1}_{\{J_t=j\}}$$

where $\hat{B}_i$ is the set of column player's mixed strategies q to which best respond for row player is to play action i. By the preceding argument if we let $\underline{J}_t = e_{J_t} \in \mathbf{R}^m$ then the following motivates the row player to use a calibrated forecast for estimating the opponent's strategy:

$$\hat{\mathbf{P}}_T(.|j) = \frac{\sum_{t=1}^{T}\underline{J}_t\mathbf{1}_{\{\hat{q}_t\in\hat{B}_i\}}}{\sum_{t=1}^{T}\mathbf{1}_{\{\hat{q}_t\in\hat{B}_i\}}}$$

Theorem 7. *In a 2-player repeated game, if both players play the best response action to a calibrated forecast of their opponent's mixed strategy, then their joint empirical distribution will converge to the set of Correlated Equilibria almost surely.*

See [PLG] §7.6 for a detailed proof. Note that existence of a calibrated forecaster can be seen in view of existence of a vanishing internal cumulative regret strategy (see [PLG] §4.5) which can be constructed by the means of Blackwell approachability. To this end, one may consider the loss matrix:

$$\underline{\ell}(i,j) = [\hat{r}_{(k,s)}]_{k,s} = [\mathbf{1}_{\{k=i\}}(\ell(i,j) - \ell(s,j))]_{k.s} \in \mathbf{R}^{n\times n}$$

In order to control internal cumulative regret

$$\max_{i,i'} \frac{1}{T}\hat{R}_{(i,i'),T} = \max \frac{1}{T}\sum_{t=1}^{T} \underline{\ell}(I_t, J_t)$$

where the second max is over the elements of the matrix argument, one is required to show $S = \mathbf{R}_{-}^{n\times n}$ is approachable. Thus, by the Corollary 1, we need to prove such a strategy p exists for a hyperplane with normal matrix $A \in \mathbf{R}_{+}^{n\times n}$ such that the loss matrix lies in the halfspace which contains S. In fact, a stronger result holds and one can prove a strategy p exists such that the expected loss matrix lies on the hyperplane:

$$\begin{aligned} tr(A^\top \underline{\ell}(p,j)) &= \sum_{i=1}^{n} p_i \sum_{s=1}^{n} A_{i,s}(\ell(i,j) - \ell(s,j)) \\ &= \sum_{i=1}^{n} \ell(i,j)(p_i \sum_{s=1}^{n} A_{i,s}) - \sum_{s=1}^{n} \ell(s,j)(\sum_{i=1}^{n} p_i A_{i,s}) \\ &= \sum_{i=1}^{n} \ell(i,j) \underbrace{(p_i \sum_{s=1}^{n} A_{i,s} - \sum_{i'=1}^{n} p_{i'} A_{i',i})}_{\text{Coefficients}} = 0 \end{aligned}$$

The above holds if all the coefficients for losses become zero. It is sufficient to show a valid solution exists for the linear system $p^\top \tilde{A} = p$ where $\tilde{A}_{i,j} = \frac{A_{i,j}}{\sum_{i'=1}^{n} A_{i',j}}$ is a column stochastic matrix. This is implied by the Perron-Frobenius theorem. As mentioned before, these probability vectors can be calculated explicitly and it only depends on the normal vector in that round.