CSCI 659 Lecture 3

Spring 2026

Instructor: Haipeng Luo

1 Adaptive Algorithms and Optimistic FTRL

While being able to deal with adversarial environments is one key advantage of online learning algorithms, they might be overly conservative exactly for the same reason. Indeed, even if the algorithm is minimax optimal, all we know is that *in the worst case*, it behaves optimally, but it is totally possible that in some benign cases, it performs worse than what one should hope for. Ideally, we want an algorithm that is robust in the worst case, but at the same time able to adapt to easier instances automatically with better than worst-case regret.

There are many different ways to define "easy" instances, each of which might lead to quite different algorithm design. In this lecture, we focus on one particular setup where at the beginning of each round, the learner has access to some prediction about what might happen for this round before making her decision. Formally, we consider the following variant of the OCO setup (recall that assuming linear loss functions is without loss of generality): at each round $t = 1, \ldots, T$,

1. the learner first obtains $m_t \in \mathbb{R}^d$, a predictor for the loss vector ℓ_t of this round,
2. the learner then decides her action $w_t \in \Omega$;
3. the environment decides the loss vector ℓ_t;
4. the learner suffers loss $\langle w_t, \ell_t \rangle$ and observes ℓ_t.

For now, you should think of m_t as some external information. Instead of assuming that it is a good predictor, our goal is to design an algorithm whose regret naturally depends on how accurate these predictors are: the more accurate they are, the smaller the regret. Later, we will discuss how to come up with these predictors on our own (so no extra external information is needed).

It turns out that a small change to the FTRL framework is enough to incorporate this extra predictor and lead to strong guarantees.[1] The resulting algorithm, called *Optimistic or Predictive* FTRL [Rakhlin and Sridharan, 2013], simply adds m_t to the current cumulative losses:

$$w_t = \operatorname*{argmin}_{w \in \Omega} \left\langle w, m_t + \sum_{s<t} \ell_s \right\rangle + \frac{1}{\eta}\psi(w)$$

where as before ψ is a regularizer and η is a learning rate. The algorithm is "optimistic" in the sense that it believes m_t is close to ℓ_t and thus directly includes it when calculating the cumulative loss. Note that if m_t is indeed ℓ_t, then the above strategy is exactly the $(t+1)$-th action of vanilla FTRL, which has constant regret based on the Be-the-Leader lemma discussed last week. In fact, these iterates from the vanilla FTRL will play a crucial role in the analysis, and we now denote them by $w'_1, w'_2, \ldots$, so that

$$w'_{t+1} = \operatorname*{argmin}_{w \in \Omega} \left\langle w, \sum_{s \le t} \ell_s \right\rangle + \frac{1}{\eta}\psi(w).$$

[1] Similar modifications apply to FTPL as well. For simplicity, we focus on FTRL in this lecture.

Extending the techniques from last lecture, we prove the following intermediate regret bound (you should compare this proof with that of Lemma 3 in Lecture 2).

Lemma 1. *The optimistic FTRL strategy ensures for any $u \in \Omega$:*

$$\sum_{t=1}^{T} \langle w_t - u, \ell_t \rangle \leq \frac{\psi(u) - \psi(w'_1)}{\eta} + \sum_{t=1}^{T} \langle w_t - w'_{t+1}, \ell_t - m_t \rangle - \sum_{t=1}^{T} \frac{D_\psi(w_t, w'_t) + D_\psi(w'_{t+1}, w_t)}{\eta}.$$

Proof. Define

$$\Phi_{t-1} = \min_{w \in \Omega} \left\langle w, m_t + \sum_{s<t} \ell_s \right\rangle + \frac{1}{\eta}\psi(w) = \left\langle w_t, m_t + \sum_{s<t} \ell_s \right\rangle + \frac{\psi(w_t)}{\eta} \quad \text{and}$$

$$\Phi'_t = \min_{w \in \Omega} \left\langle w, \sum_{s \leq t} \ell_s \right\rangle + \frac{1}{\eta}\psi(w) = \left\langle w'_{t+1}, \sum_{s \leq t} \ell_s \right\rangle + \frac{\psi(w'_{t+1})}{\eta}.$$

Then by Lemma 2 of Lecture 2, we have

$$\Phi'_{t-1} - \Phi_{t-1} \leq -\langle w_t, m_t \rangle - \frac{1}{\eta} D_\psi(w_t, w'_t)$$

and

$$\Phi_{t-1} - \Phi'_t \leq -\langle w'_{t+1}, \ell_t - m_t \rangle - \frac{1}{\eta} D_\psi(w'_{t+1}, w_t).$$

Adding up both and rearranging shows

$$\langle w_t, \ell_t \rangle \leq \Phi'_t - \Phi'_{t-1} + \langle w_t - w'_{t+1}, \ell_t - m_t \rangle - \frac{D_\psi(w_t, w'_t) + D_\psi(w'_{t+1}, w_t)}{\eta}.$$

Finally, summing over t, telescoping, and realizing $\Phi'_T \leq \left\langle u, \sum_{t \leq T} \ell_t \right\rangle + \frac{\psi(u)}{\eta}$ and $\Phi'_0 = \frac{\psi(w'_1)}{\eta}$ finishes the proof. □

An important distinction between this regret bound and that of vanilla FTRL is the improvement in the stability term, which now depends on the difference between ℓ_t and its predicted value m_t. Indeed, this is even better illustrated when we use a strongly convex regularizer and obtain the following final regret bound.

Theorem 1. *Suppose that the regularizer ψ is strongly convex with respect to some norm $\|\cdot\|$ and has range $B_\psi = \max_{u \in \Omega} \psi(u) - \min_{u \in \Omega} \psi(u)$. Then Optimistic FTRL ensures*

$$\mathcal{R}_T \leq \frac{B_\psi}{\eta} + \eta \sum_{t=1}^{T} \|\ell_t - m_t\|_\star^2 - \frac{1}{4\eta} \sum_{t=2}^{T} \|w_t - w_{t-1}\|^2.$$

Proof. It suffices to handle the stability term and the negative terms from the bound of Lemma 1. Using Lemma 4 of Lecture 2, we immediately obtain $\|w_t - w'_{t+1}\| \leq \eta \|\ell_t - m_t\|_\star$, and thus the stability term $\langle w_t - w'_{t+1}, \ell_t - m_t \rangle$ is further bounded by $\eta \|w_t - w'_{t+1}\| \|\ell_t - m_t\|_\star \leq \eta \|\ell_t - m_t\|_\star^2$. Next, for the term $\sum_{t=1}^{T} D_\psi(w_t, w'_t) + D_\psi(w'_{t+1}, w_t)$, we first drop $D_\psi(w_1, w'_1)$ and $D_\psi(w'_{T+1}, w_T)$ (both nonnegative), and then shift the index to arrive at a lower bound:

$$\sum_{t=2}^{T} D_\psi(w_t, w'_t) + D_\psi(w'_t, w_{t-1}),$$

which can be further lower bounded as

$$\frac{1}{2} \sum_{t=2}^{T} \|w_t - w'_t\|^2 + \|w'_t - w_{t-1}\|^2 \qquad \text{(by strong convexity)}$$

$$\geq \frac{1}{4} \sum_{t=2}^{T} \left(\|w_t - w'_t\| + \|w'_t - w_{t-1}\|\right)^2 \qquad \left(a^2 + b^2 \geq \tfrac{(a+b)^2}{2}\right)$$

$$\geq \frac{1}{4} \sum_{t=2}^{T} \|w_t - w_{t-1}\|^2. \qquad \text{(triangle inequality)}$$

Combining these with Lemma 1 finishes the proof. □

For a moment, assume that we can set the learning rate as $\eta = \sqrt{B_\psi / \sum_{t=1}^T \|\ell_t - m_t\|_\star^2}$. Then ignoring the last negative term, we obtain a regret bound of order $\mathcal{O}\left(\sqrt{B_\psi \sum_{t=1}^T \|\ell_t - m_t\|_\star^2}\right)$, showing that the performance of the algorithm indeed depends on how accurate the predictors m_t are. The prediction error $\sum_{t=1}^T \|\ell_t - m_t\|_\star^2$ thus serves as a measure of difficulty for this problem. For example, if the predictors are perfect with zero error (that is, $m_t = \ell_t$), then we simply have zero regret. As a less extreme example, if the predictors are somewhat accurate such that $\sum_{t=1}^T \|\ell_t - m_t\|_\star^2 = o(T)$, then the regret bound becomes $o(\sqrt{T})$ (ignoring the dependence on B_ψ). On the other hand, even if the predictors are all inaccurate with $\Theta(T)$ error (this is under a mild condition that $\|\ell_t\|_\star$ is bounded, so $\|m_t\|_\star$ is naturally bounded as well), the regret bound simply recovers the $\mathcal{O}(\sqrt{T})$ worst-case regret of vanilla FTRL. In other words, optimistic FTRL is never worse than vanilla FTRL, but could perform much better as long as we are not in the worst-case scenario.

The aforementioned choice of the learning rate η requires knowing the prediction error ahead of time, which is rather unrealistic. However, this can be easily resolved. The simplest approach is the doubling trick (which you have seen in HW1): make a guess on the value of the prediction error (starting from 1 for example), optimally tune the algorithm based on this guess, and once the actual error exceeds the guess, double the guess and restart the algorithm. This leads to the same regret bound up to a constant factor for the exact same reason as the case you need to solve in HW1.

While this doubling trick is a very convenient and general way to resolve such tuning issues, it might not be a practical one. Indeed, restarting the algorithm means completely discarding the previously collected data, which often sounds unreasonable (if not absurd) to practitioners. Fortunately, most of the time such tuning issues can also be resolved by a *time-varying* learning rate that follows the form of the optimal tuning but with only the observed data. For example, in our case, using learning rate $\eta_t = \sqrt{B_\psi / \sum_{s<t} \|\ell_s - m_s\|_\star^2}$ at time t would also lead to the same regret bound. We omit the details and leave this as an exercise for interested readers.

Finally, the negative term $-\frac{1}{4\eta}\sum_{t=2}^T \|w_t - w_{t-1}\|^2$ in the regret bound, which will be used to derive something important in the next lecture, might seem a bit mysterious — it seemingly indicates that the less stable the algorithm, the smaller the regret, exactly the opposite of what we showed before. One explanation is: keep in mind that what Theorem 1 shows is an *upper bound* of the regret, and it might not truly characterizes the behavior of the actual regret itself. In fact, if one takes a step back and focuses on the term $\langle w_t - w'_{t+1}, \ell_t - m_t \rangle - \frac{1}{\eta} D_\psi(w'_{t+1}, w_t)$ in Lemma 1, which is bounded by $\|w_t - w'_{t+1}\| \|\ell_t - m_t\|_\star - \frac{1}{2\eta}\|w_t - w'_{t+1}\|^2$, then by seeing this as a quadratic of $\|w_t - w'_{t+1}\|$ whose range is in $[0, \eta \|\ell_t - m_t\|_\star]$ (based on our earlier analysis), we conclude that this upper bound is again strictly increasing in $\|w_t - w'_{t+1}\|$, thus in favor of stable algorithms and consistent with our earlier intuition.

1.1 Constructing the Predictors

So far we have treated m_t as some external information. Can we instead construct m_t ourselves without using any extra information? Intuitively, if we believe that the past loss vectors $\ell_1, \ldots, \ell_{t-1}$ are predictive for the current loss vector ℓ_t, then we should be able to come up with a reasonable predictor m_t based on $\ell_1, \ldots, \ell_{t-1}$. Indeed, we can treat this as another machine learning problem such as time series forecasting (a salient example would be predicting the next word based on the first few words of a sentence, what an LLM essentially does), which is often solved by autoregressive models. In fact, we can also treat this as *another OCO problem*, where the learner is required to predict m_t at each round, with loss function $\|\ell_t - m_t\|_\star^2$.

Here, we only spell out perhaps the simplest method: use the most recent loss vector as the predictor, that is, $m_t = \ell_{t-1}$ (with arbitrary ℓ_0). This leads to a bound of order $\mathcal{O}\left(\sqrt{B_\psi \sum_{t=1}^T \|\ell_t - \ell_{t-1}\|_\star^2}\right)$, which depends on the variation of the loss sequence (often called the *path length*). Thus, as long as the environment is not changing rapidly, the algorithm is able to exploit this fact and achieves better performance. In the next lecture, we will see a concrete example of such slowly changing environments.

Another remark is that the path length $\sum_{t=1}^T \|\ell_t - \ell_{t-1}\|_\star^2$ is always bounded by

$$\sum_{t=1}^T \|\ell_t - \mu + \mu - \ell_{t-1}\|_\star^2 \le 2\sum_{t=1}^T \left(\|\ell_t - \mu\|_\star^2 + \|\mu - \ell_{t-1}\|_\star^2\right) = \mathcal{O}\left(\sum_{t=1}^T \|\ell_t - \mu\|_\star^2\right)$$

for any $\mu \in \mathbb{R}^d$. In particular, we can pick the μ that minimizes $\sum_{t=1}^T \|\ell_t - \mu\|_\star^2$, showing that our time-varying predictors are overall never worse than a fixed predictor. For example, when $\|\cdot\|$ is the L_2 norm, then the optimal μ is simply the average loss $\frac{1}{T}\sum_{t=1}^T \ell_t$, and the bound becomes the variance of the loss vectors. Note that there are clearly cases where the variance is much worse than the path length though (try to come up with one yourself).

2 Refinement for the Expert Problem

For the rest of the lecture, we focus on the expert problem (switching the notation from $w \in \Omega$ to $p \in \Delta(N)$ again). Applying Optimistic FTRL with the negative entropy regularizer, which corresponds to the following strategy called *Optimistic Hedge*:

$$p_t(i) \propto \exp\left(-\eta\left(m_t(i) + \sum_{s<t} \ell_s(i)\right)\right),$$

we obtain a regret bound of order $\mathcal{O}\left(\sqrt{\sum_{t=1}^T \|\ell_t - m_t\|_\infty \ln N}\right)$ (with the optimal tuning). This bound depends on the worst prediction error among all experts at each time (due to the infinity norm). Instead, is there a way to depend on, for example, only the prediction error for the optimal expert?

To address this question, we first point out that the same algorithm above in fact enjoys a more refined regret bound that depends on some weighted average of the prediction errors (weighted by p_t precisely), instead of the worst case one. This is done by analyzing the stability term more carefully, without directly invoking strong convexity. The result is summarized in the following theorem, and is in fact a generalization of the bound we proved in Theorem 2 of Lecture 1.

Theorem 2. *As long as $\eta|\ell_t(i) - m_t(i)| \le 1$ holds for all t and i, Optimistic Hedge ensures*

$$\mathcal{R}_T \le \frac{\ln N}{\eta} + \eta \sum_{t=1}^T \sum_{i=1}^N p_t(i)(\ell_t(i) - m_t(i))^2.$$

Proof. Using Lemma 1, we have

$$\mathcal{R}_T \le \frac{\ln N}{\eta} + \sum_{t=1}^T \left(\langle p_t - p'_{t+1}, \ell_t - m_t\rangle - \frac{1}{\eta} D_\psi(p'_{t+1}, p_t)\right).$$

Note that we have kept one Bregman divergence term here (but dropped the other one $D_\psi(p_t, p'_t)$), which is crucial for the proof. Specifically, we will show $\langle p_t - p'_{t+1}, \ell_t - m_t\rangle - \frac{1}{\eta}D_\psi(p'_{t+1}, p_t) \le \eta\sum_{i=1}^N p_t(i)(\ell_t(i) - m_t(i))^2$, which then finishes the proof. Indeed, by definitions we have

$$p'_{t+1}(i) \propto \exp\Big(-\eta \sum\nolimits_{s\le t} \ell_s(i)\Big) \propto p_t(i)\exp\left(-\eta(\ell_t(i) - m_t(i))\right).$$

Then, noticing that the Bregman divergence $D_\psi(p,q)$ with respect to the negative entropy is simply the KL-divergence $\mathrm{KL}(p,q) = \sum_{i=1}^N p(i)\ln\frac{p(i)}{q(i)}$, we have

$$\begin{aligned}
-\frac{1}{\eta}D_\psi(p'_{t+1}, p_t) &= \frac{1}{\eta}\sum_{i=1}^N p'_{t+1}(i)\ln\frac{p_t(i)}{p'_{t+1}(i)} \\
&= \frac{1}{\eta}\sum_{i=1}^N p'_{t+1}(i)\ln\frac{\sum_{j=1}^N p_t(j)\exp\left(-\eta(\ell_t(j) - m_t(j))\right)}{\exp\left(-\eta(\ell_t(i) - m_t(i))\right)} \\
&= \langle p'_{t+1}, \ell_t - m_t\rangle + \frac{1}{\eta}\ln\left(\sum_{j=1}^N p_t(j)\exp\left(-\eta(\ell_t(j) - m_t(j))\right)\right).
\end{aligned}$$

The rest of the proof simply repeats what we did in Lecture 1:

$$\begin{aligned}
&\frac{1}{\eta}\ln\left(\sum_{j=1}^N p_t(j)\exp\left(-\eta(\ell_t(j)-m_t(j))\right)\right)\\
&\le \frac{1}{\eta}\ln\left(\sum_{j=1}^N p_t(j)\left(1-\eta(\ell_t(j)-m_t(j))+\eta^2(\ell_t(j)-m_t(j))^2\right)\right)\\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (e^{-y}\le 1-y+y^2,\ \forall y\ge -1)\\
&= \frac{1}{\eta}\ln\left(1-\eta\left\langle p_t,\ell_t-m_t\right\rangle+\eta^2\sum_{j=1}^N p_t(j)(\ell_t(j)-m_t(j))^2\right)\\
&\le -\left\langle p_t,\ell_t-m_t\right\rangle+\eta\sum_{j=1}^N p_t(j)(\ell_t(j)-m_t(j))^2, \qquad (\ln(1+y)\le y)
\end{aligned}$$

where in the first inequality we use the condition $\eta|\ell_t(j)-m_t(j)|\le 1$ of the theorem. □

With this refined regret bound, we make the following modification to the algorithm, which further enables us to derive a bound only in terms of the prediction error of the best expert:

$$p_t(i)\propto\exp\left(-\eta\left(m_t(i)+\sum_{s<t}(\ell_s(i)+c_s(i))\right)\right), \tag{1}$$

where $c_s(i)=4\eta(\ell_s(i)-m_s(i))^2$ is sometimes referred to as a loss *correction term*. Compared to Optimistic Hedge, this new algorithm adds the correction term to the loss of each expert at each time, further penalizing those experts with a large prediction error. The regret of this algorithm against any expert thus naturally depends on the prediction error for this expert, as shown below.

Theorem 3. *As long as $\eta|\ell_t(i)-m_t(i)|\le\frac{1}{4}$ holds for all t and i, Algorithm (1) ensures that for any expert $j\in[N]$, the regret against this expert is bounded as*

$$\sum_{t=1}^T\left\langle p_t-e_j,\ell_t\right\rangle\le\frac{\ln N}{\eta}+4\eta\sum_{t=1}^T(\ell_t(j)-m_t(j))^2$$

(where e_i is the standard basis vector with the i-th coordinate being 1).

Proof. First observe that Algorithm (1) is exactly feeding Optimistic Hedge with the loss vector ℓ_t+c_t (instead of just ℓ_t). This motivates us to apply Theorem 1 with ℓ_t replaced by ℓ_t+c_t. To do so, we first verify that the condition of Theorem 1 is indeed satisfied:

$$\begin{aligned}
\eta|\ell_t(i)+c_t(i)-m_t(i)| &= \eta|\ell_t(i)-m_t(i)+4\eta(\ell_s(i)-m_s(i))^2| && \text{(defintion of } c_t(i)\text{)}\\
&\le \eta|\ell_t(i)-m_t(i)|(1+4\eta|\ell_t(i)-m_t(i)|)\\
&\le 2\eta|\ell_t(i)-m_t(i)| && (2)\\
&\le \tfrac{1}{2},
\end{aligned}$$

where both of the last two inequalities use the condition $\eta|\ell_t(i)-m_t(i)|\le\frac{1}{4}$. Thus, we can indeed apply Theorem 1 and obtain for any expert $j\in[N]$:

$$\begin{aligned}
\sum_{t=1}^T\left\langle p_t-e_j,\ell_t+c_t\right\rangle &\le \frac{\ln N}{\eta}+\eta\sum_{t=1}^T\sum_{i=1}^N p_t(i)(\ell_t(i)+c_t(i)-m_t(i))^2\\
&\le \frac{\ln N}{\eta}+4\eta\sum_{t=1}^T\sum_{i=1}^N p_t(i)(\ell_t(i)-m_t(i))^2 && \text{(by Eq. (2))}\\
&= \frac{\ln N}{\eta}+\sum_{t=1}^T\left\langle p_t,c_t\right\rangle, && \text{(definition of } c_t\text{)}
\end{aligned}$$

which then completes the proof after moving the term $\sum_{t=1}^T\left\langle p_t-e_j,c_t\right\rangle$ to the right. □

Let $j^\star \in \operatorname{argmin}_j \sum_{t=1}^T \ell_t(j)$ be an optimal expert. Then with the learning rate η set to $\min\left\{1/4, \sqrt{\ln N / \sum_{t=1}^T (\ell_t(j^\star) - m_t(j^\star))^2}\right\}$ (we prevent this from going above $\frac{1}{4}$ to make sure that the condition $\eta|\ell_t(i) - m_t(i)| \le \frac{1}{4}$ is always satisfied for losses/predictors in $[0,1]$), we obtain the bound $\mathcal{R}_T = \mathcal{O}\left(\sqrt{\sum_{t=1}^T (\ell_t(j^\star) - m_t(j^\star))^2 \ln N} + \ln N\right)$, which only depends on the prediction error of the optimal expert. However, an obvious issue is that this tuning of the learning rate depends on the knowledge of the best expert $j^\star$, which we clearly do not have ahead of time. Note that this issue is very different from the tuning issue discussed earlier in Section 1, which as mentioned can be easily solved via a time-varying learning rate. Here, the difficulty is that each expert requires a different tuning. A natural attempt would be to let the learning rate vary not only over time, but also across the experts, leading to the following algorithm

$$p_t(i) \propto \exp\left(-\eta_t(i)\left(m_t(i) + \sum_{s<t}(\ell_s(i) + c_s(i))\right)\right), \tag{3}$$

where we replace the fixed learning rate η in Eq. (1) with

$$\eta_t(i) = \min\left\{\frac{1}{4}, \sqrt{\frac{\ln N}{\sum_{s<t}(\ell_t(i) - m_t(i))^2}}\right\}. \tag{4}$$

Unfortunately, no one has been able to prove (or disprove) that this works, and this tuning issue had been referred to as the *impossible tuning issue*.

Somewhat surprisingly, a work by Chen et al. [2021] makes this "impossible tuning" possible, and the solution is very close to the natural idea above. Recall that Algorithm (1) can be written as the following Optimistic FTRL form:

$$p_t = \operatorname*{argmin}_{p \in \Delta(N)} \left\langle p, m_t + \sum_{s<t}(\ell_s + c_s)\right\rangle + \Psi(p), \tag{5}$$

where $\Psi(p) = \frac{1}{\eta}\sum_{i=1}^N p(i) \ln p(i)$ is the entropy regularizer with a fixed learning rate. The solution of [Chen et al., 2021] is simply to replace Ψ with

$$\Psi_t(p) = \sum_{i=1}^N \frac{1}{\eta_t(i)} p(i) \ln p(i)$$

where $\eta_t(i)$ is defined in the same way as Eq. (4).[2] The first important thing to notice here is that this is *not* the same algorithm as Eq. (3), even though they look very closely related. Indeed, by writing down the Lagrangian and setting the gradient to zero, we see that the algorithm of [Chen et al., 2021] is in fact the following:

$$p_t(i) = \exp\left(-\eta_t(i)\left(\lambda + m_t(i) + \sum_{s<t}(\ell_s(i) + c_s(i))\right) - 1\right), \tag{6}$$

for some Lagrangian multiplier λ that can be efficiently found via binary search based on the fact $\sum_i p_t(i) = 1$. Note that if $\eta_t(i)$ does not vary over different i's, then the above is indeed the same as Eq. (3) since the factor $\exp(-\eta_t(i)\lambda - 1)$ simply becomes the constant previously hidden in the $\propto$ sign. But for general $\eta_t(i)$ that varies over i, such as Eq. (4), there is no closed-form for λ, and the two algorithms are thus indeed not equivalent.

Chen et al. [2021] prove the following guarantee of this algorithm: for *any* expert j,

$$\sum_{t=1}^T \langle p_t - e_j, \ell_t\rangle = \mathcal{O}\left(\sqrt{\sum_{t=1}^T (\ell_t(j) - m_t(j))^2 \ln(NT)} + \ln(NT)\right). \tag{7}$$

Other than the small extra $\ln T$ factor (which we will hide using the $\widetilde{\mathcal{O}}(\cdot)$ notation in the remaining discussion), this guarantee is even stronger than our earlier goal $\mathcal{R}_T = \mathcal{O}\left(\sqrt{\sum_{t=1}^T (\ell_t(j^\star) - m_t(j^\star))^2 \ln N} + \ln N\right)$ (think about why the former could be *strictly* stronger).

[2] There are other minor modifications of the algorithm that we omit here; see [Chen et al., 2021] if interested.

2.1 Implications

Why do we bother to spend so much effort to obtain regret bound (7), which in fact might not even be better than the version stated in Theorem 2 with the weighted error? It turns out that there are indeed many nice applications of Eq. (7). We explore some of these in the rest of this lecture and will also see more in the future.

Application 1 First, consider not using any predictors at all, that is, $m_t = \mathbf{0}$ (the all-zero vector). Then Eq. (7) implies $\mathcal{R}_T = \widetilde{\mathcal{O}}\left(\sqrt{\sum_{t=1}^T \ell_t(j^\star)^2 \ln N} + \ln N\right) = \widetilde{\mathcal{O}}(\sqrt{L^\star \ln N} + \ln N)$ where $L^\star = \sum_{t=1}^T \ell_t(j^\star)$ is the loss of the best expert. Therefore, the better the optimal expert, the faster our algorithm converges to such a good expert. This is often called a "small-loss" regret bound, which itself has further strong implications for other problems.

In the next two applications, we make use of a useful trick: for each t, if $m_t(i)$ is the same for all expert i, then by examining either Eq. (5) or Eq. (6), one sees that the distribution p_t is in fact *independent* of m_t (but any future p_s with $s > t$ depends on m_t through c_t of course). This small observation implies that m_t can even "cheat" and depend on ℓ_t, *unknown* at the beginning of round t, as long as it has identical coordinates and becomes available at the end of round t.

Application 2 The first such "cheating" predictor we consider is $m_t(i) = \ell_t(1)$ for all i. Applying Eq. (7), we know that the regret of the algorithm against Expert 1 is only $\widetilde{\mathcal{O}}(\ln N)$, while the regret against all other experts is still at most $\widetilde{\mathcal{O}}(\sqrt{T \ln N})$. This is useful in for example the following scenario: a company has been using a decent online learning algorithm $\mathcal{A}$ for a while, and one day as a researcher of the company you propose a new algorithm $\mathcal{B}$ that you think might perform even better. What the company can do now is to treat these two algorithms as experts, and combine them using an expert algorithm. However, since this is related to the profit of the company and we have no prior knowledge about how $\mathcal{B}$ might actually perform in practice, to be conservative we want our overall performance to be at least not too far away from the baseline $\mathcal{A}$. Then the guarantee mentioned above exactly fits our need here: let $\mathcal{A}$ be Expert 1 so that the loss of the final algorithm is no worse than $\mathcal{A}$ by only $\widetilde{\mathcal{O}}(1)$, but at the same time is also comparable to $\mathcal{B}$ if it is indeed a much better algorithm.

Question 1. *How is this approach compared to the common practice of doing $\mathcal{A}/\mathcal{B}$ testing?*

Question 2. *Can we go one step further and use the following cheating predictor: $m_t(i) = \ell_t(j^\star)$ for all i, where $j^\star$ is the optimal expert, so that our regret against $j^\star$ is always $\widetilde{\mathcal{O}}(\ln N)$?*

Application 3 The next "cheating" predictor we consider is $m_t(i) = \langle p_t, \ell_t \rangle$ for all i. Eq. (7) implies that the regret against any expert j is bounded as

$$\sum_{t=1}^T \langle p_t - e_j, \ell_t \rangle = \widetilde{\mathcal{O}}\left(\sqrt{\sum_{t=1}^T \langle p_t - e_j, \ell_t \rangle^2 \ln N} + \ln N\right), \tag{8}$$

scaling with the cumulative squared instantaneous regret against the same expert j. While the meaning of this bound is less intuitive, it in fact has many further strong implications. The first one being that it recovers the aforementioned "small-loss" bounds.

Theorem 4. *If an algorithm ensures Eq. (8), then it also ensures*

$$\sum_{t=1}^T \langle p_t - e_j, \ell_t \rangle = \widetilde{\mathcal{O}}\left(\sqrt{\sum_{t=1}^T \ell_t(j) \ln N} + \ln N\right),$$

Proof. For conciseness, define $R_T(j) = \sum_{t=1}^T \langle p_t - e_j, \ell_t \rangle$ and $L_T(j) = \sum_{t=1}^T \ell_t(j)$. With the fact $\langle p_t - e_j, \ell_t \rangle^2 \le |\langle p_t - e_j, \ell_t \rangle| \le \langle p_t + e_j, \ell_t \rangle = \langle p_t - e_j + 2e_j, \ell_t \rangle$, Eq. (8) implies:

$$R_T(j) = \widetilde{\mathcal{O}}\left(\sqrt{(R_T(j) + 2L_T(j)) \ln N} + \ln N\right) = \widetilde{\mathcal{O}}\left(\sqrt{R_T(j) \ln N} + \sqrt{L_T(j) \ln N} + \ln N\right),$$

a recursive guarantee on $R_T(j)$. Further using the fact that $x \le b\sqrt{x} + c$ implies $x \le b^2 + 2c$ (verify it yourself) then proves the theorem. □

Even more surprisingly, the same bound also implies almost constant regret in a stochastic setting.

Theorem 5. *Suppose that there exists a good expert $j^\star$ and a constant gap $\Delta \in (0, 1]$ such that $\mathbb{E}_t[\ell_t(i) - \ell_t(j^\star)] \geq \Delta$ for all t and $i \neq j^\star$, where $\mathbb{E}_t$ denotes the conditional expectation given all randomness before round t. If an algorithm ensures Eq. (8), then its regret against $j^\star$ is at most $\widetilde{\mathcal{O}}\left(\frac{\ln N}{\Delta}\right)$.*

Proof. Let $R_T(j^\star) = \sum_{t=1}^T \langle p_t - e_{j^\star}, \ell_t\rangle$ be the regret against expert $j^\star$. By the condition we have

$$\mathbb{E}_t[\langle p_t - e_{j^\star}, \ell_t\rangle] = \mathbb{E}_t\left[\sum_{i\neq j^\star} p_t(i)(\ell_t(i) - \ell_t(j^\star))\right] \geq \Delta \sum_{i\neq j^\star} p_t(i) = \Delta(1 - p_t(j^\star)),$$

and thus $\mathbb{E}[R_T(j^\star)] \geq \Delta B$ where we define $B = \mathbb{E}\left[\sum_{t=1}^T (1 - p_t(j^\star))\right]$. On the other hand,

$$\begin{aligned}\mathbb{E}\left[\sum_{t=1}^T \langle p_t - e_{j^\star}, \ell_t\rangle^2\right] &\leq \mathbb{E}\left[\sum_{t=1}^T \left|\sum_{i=1}^N p_t(i)(\ell_t(i) - \ell_t(j^\star))\right|\right] \\ &\leq \mathbb{E}\left[\sum_{t=1}^T \sum_{i=1}^N p_t(i)|\ell_t(i) - \ell_t(j^\star)|\right] \leq \mathbb{E}\left[\sum_{t=1}^T \sum_{i\neq j^\star} p_t(i)\right] = B.\end{aligned}$$

Combining these with Eq. (8) implies

$$\begin{aligned}\Delta B \leq \mathbb{E}\left[R_T(j^\star)\right] &= \widetilde{\mathcal{O}}\left(\mathbb{E}\left[\sqrt{\sum_{t=1}^T \langle p_t - e_{j^\star}, \ell_t\rangle^2 \ln N}\right] + \ln N\right) \\ &= \widetilde{\mathcal{O}}\left(\sqrt{\mathbb{E}\left[\sum_{t=1}^T \langle p_t - e_{j^\star}, \ell_t\rangle^2\right] \ln N} + \ln N\right) \qquad \text{(Jensen's inequality)} \\ &\leq \widetilde{\mathcal{O}}\left(\sqrt{B \ln N} + \ln N\right). \qquad (9)\end{aligned}$$

Applying $x \leq b\sqrt{x} + c \Rightarrow x \leq b^2 + 2c$ again shows $B = \widetilde{\mathcal{O}}(\ln N/\Delta^2)$. Plugging this back to Eq. (9) finishes the proof. □

We emphasize that the last two theorems hold simultaneously as long as an algorithm satisfies Eq. (8) (the detail of the algorithm is irrelevant), illustrating strong adaptivity of the algorithm. In the future, we will see one more important consequence of having regret bound (8).

Application 4 In the last application, we take $m_t(i) = m'_t(i) + \langle p_t, \ell_t - m'_t\rangle$ for another arbitrary predictor $m'_t \in [0, 1]^N$ (that is available at the beginning of round t). For a similar reason as the last two applications, even though the term $\langle p_t, \ell_t - m'_t\rangle$ is unknown before the end of round t, the algorithm is still valid since this part is the same across all i. In fact, Application 3 is just a special case with $m'_t = \mathbf{0}$. Eq. (7) now implies that the regret against any expert j is bounded as

$$\sum_{t=1}^T \langle p_t - e_j, \ell_t\rangle = \widetilde{\mathcal{O}}\left(\sqrt{\sum_{t=1}^T \langle p_t - e_j, \ell_t - m'_t\rangle^2 \ln N + \ln N}\right), \qquad (10)$$

which is further bounded by $\widetilde{\mathcal{O}}\left(\sqrt{\sum_{t=1}^T \|\ell_t - m'_t\|_\infty^2 \ln N + \ln N}\right)$, a bound similar to the one we saw earlier for Optimistic Hedge (without using correction terms). However, the benefit here is that simultaneously, this bound (10) also enjoys the same fast rate consequence stated in Theorem 5.

Theorem 6. *In the same environment as stated in Theorem 5, if an algorithm ensures Eq. (10), then its regret against $j^\star$ is at most $\widetilde{\mathcal{O}}\left(\frac{\ln N}{\Delta}\right)$.*

The proof is almost identical to that for Theorem 5, and we leave it as an exercise.

We conclude by pointing out that there are in fact many more applications of Eq. (7), including those for the general OCO problem (obtained by combining different OCO algorithms via an expert algorithm satisfying Eq. (7)). See [Chen et al., 2021] for details.

References

Liyu Chen, Haipeng Luo, and Chen-Yu Wei. Impossible tuning made possible: A new expert algorithm and its applications. In *Conference on Learning Theory*, pages 1216–1259. PMLR, 2021.

Sasha Rakhlin and Karthik Sridharan. Optimization, learning, and games with predictable sequences. In *Advances in Neural Information Processing Systems 26*, 2013.