DSC 211 Introduction to Optimization Winter 2024 — Instructor: Jun-Kun Wang
Scribe: Merlin Chang, Marialena Sfyraki — Feb 22, 2024
Editors/TAs: Merlin Chang, Marialena Sfyraki

Lecture 14: Online Linear Optimization and Regert

1 Online Optimization Recap

Recall the setting for online optimization,

Protocol/Setting

1: **for** $t = 1, 2, \ldots$ **do**
2: Commit a point z_t with its convex decision space $Z \subset R^d$.
3: Receive a loss function $\ell_t(\cdot) : Z \to \mathbb{R}$ and incurs a loss $\ell_t(z_t)$.
4: **end for**

The objective is to devise a strategy of making decision in the decision space Z at each step t where the cumulative regret is minimized:

Definition 1 (Regret with respect to a benchmark z_*)**.**

$$\text{Regret}_T(z_*) = \sum_{t=1}^{T} \ell_t(z_t) - \sum_{t=1}^{T} \ell_t(z_*)$$

where $z_* \in Z$ is any fixed benchmark in Z.

2 Online Linear Optimization

Online linear optimization refers to online optimization problems where the loss function is linear in its first argument.

Protocol/Setting

1: **for** $t = 1, 2, \ldots$ **do**
2: Commit a point z_t with its convex decision space $Z \subset R^d$.
3: Receive a loss function $\ell_t(z) = c_t^\top z_t$ and incurs a loss $\ell_t(z_t)$.
4: **end for**

In this case, the cumulative regret that we want to minimize with respect to a fixed benchmark is defined as

$$\text{Regret}_T(z_*) = \sum_{t=1}^{T} c_t^\top z_t - \sum_{t=1}^{T} c_t^\top z_*$$

Let's consider the following applications for OLO.

2.1 Application 1 - Path Planning

Imagine you wish to send a steam of information over the web through a series of connected network of servers. The cost between any two servers can change at each time step t. We can then formulate OLO with the following assumptions:

- Let A represents the set of paths in the given network of servers with d edges. In this case, $A \subset \{0,1\}^d \subset \mathbb{R}^d$ is our decision space.
- Z be the set of potential weight for the edges of the graph. In other words, any $z \in Z$ is a potential gradient of the loss function at any given time step.
- If we wish to send one packet of information through a path a at time t, then the total cost is simply $a_t^\top z_t$, which correspond to the linear loss $l_t(a) = a_t^\top z_t$, where z_t is the cost of the path at time t.
- The cumulative regret R_{best} corresponds to the total extra cost of sending a steam of information compare to the best/benchmark path in hindsight.

2.2 Application 2 - Prediction with expert advice

Let's consider the decision space is the probability simplex.

$$\Delta_n := \{z \in \mathbb{R}^n : \sum_{i=1}^{n} z[i] = 1, z[i] \geq 0\}.$$

In this case, OLO at each time step has the following interpretation, at each t, plays $z_t \in \Delta_n$, receives a cost vector c_t,

$$\begin{aligned}
\langle z_t, c_t \rangle &= \sum_{i=1}^{n} z_t[i] c_t[i] \\
&= \sum_{i=1}^{n} \mathrm{P}(i_t = i) c_t[i] \\
&= \mathbb{E}_{i_t \sim z_t} \left[c_t[i_t] \right]
\end{aligned}$$

3 OLO Algorithms

3.1 FTRL

Recall that the FTRL algorithm is aimed at solving the stability issue with FTL, by adding a strongly convex regularized term at each around. At round t, play

$$z_t = \arg\min_{z \in Z} \sum_{s=1}^{t-1} \ell_s(z) + R(z),$$

where $R(z) : Z \to \mathbb{R}$ is strongly convex.
Let's consider the following settings, the decision space is still the probability simplex:

$$\Delta_n := \{z \in \mathbb{R}^n : \sum_{i=1}^{n} z[i] = 1, z[i] \geq 0\}.$$

Let $\phi(\cdot)$ be the negative entropy. We have proven this in HW 1, that the negative entropy is 1-strongly convex with respect to $\|\cdot\|_1$ when restricted to the probability simplex.

Lemma 1.

$$\max_{z \in \Delta_n} \phi(z) - \min_{z \in \Delta_n} \phi(z) = \log n.$$

Proof. $\max_{z \in \Delta_n} \phi(z) = 0$, $\min_{z \in \Delta_n} \phi(z)$ is same as maximizing the entropy. One way to think of entropy is a metric that measures how much uncertainty there is in the system. We can see that when the probability distribution is exactly uniform, we have the most amount of uncertainty. Thus, we know that when $z_* = \frac{\mathbf{1}}{\mathbf{n}}$, entropy is maximized. Thus, we can see that

$$\begin{aligned}
\max_{z \in \Delta_n} \phi(z) - \min_{z \in \Delta_n} \phi(z) &= 0 - \phi(z*) \\
&= -\sum \frac{1}{n} \log \frac{1}{n} \\
&= \log n
\end{aligned}$$

□

Lemma 1 provide a way to bound the difference between any two regulaized terms. If we let $R(\cdot) = \eta\phi(\cdot)$, then we have the following bound:

$$R(x) - R(y) \leq \eta \log n$$

for any $x, y \in \Delta_n$.

Recall the FTRL theorem in last lecture: Suppose each $\ell_1(\cdot), \ell_2(\cdot), \dots, \ell_T(\cdot)$ is convex and each $\ell_t(\cdot)$ is L_t-Lipschitz w.r.t. a norm $\|\cdot\|$. Denote L_0 as the constant such that $L_t \le L_0$.

Theorem 1. *FTRL with a η-strongly convex regularizer $R(\cdot)$ w.r.t. $\|\cdot\|$ has*

$$\mathrm{Regret}_T(z_*) \le R(z_*) - R(z_1) + \frac{TL_0^2}{\eta}.$$

Let us apply the theorem to OLO with the decision space being the probability simplex, where $\|\cdot\| \leftarrow \|\cdot\|_1$ and $R(\cdot) \leftarrow \eta\phi(\cdot)$. We have

$$\begin{aligned}\mathrm{Regret}_T(z_*) &\le R(z_*) - R(z_1) + \frac{TL_0^2}{\eta} \\ &\le \eta \log n + \frac{TL_0^2}{\eta} \\ &= 2L_0(\sqrt{T\log n}), \text{ when } \eta = \sqrt{\frac{TL_0^2}{\log n}}.\end{aligned}$$

Hence, the regret depends logarithmically with the number of experts/items (i.e., $\sqrt{\log n}$), which is in contrast to the case of using the squared of the l_1 norm as the regularizer (i.e., $\sqrt{n}$). A similar result was shown when we studied Mirror Descent in Lecture 12.

If we apply online gradient descent to it, the update step in FTRL would be the following:

$$z_t = \arg\min_{z \in \Delta_n} \sum_{s=1}^{t-1} \langle z, c_s \rangle + \frac{1}{\eta} \sum_{i=1}^{n} z_i \log z_i.$$

The update step has a closed form solution, for each $i \in [n]$,

$$z_t[i] = \frac{\exp\left(-\eta \sum_{s=1}^{t-1} c_s[i]\right)}{\sum_{j=1}^{n} \exp\left(-\eta \sum_{s=1}^{t-1} c_s[j]\right)}$$

This can be solved by formulating it as constraint optimization restricting the domain to be Δ_n and using KKT conditions to derive the final solution.

3.2 Online Mirror Descent

Recall in offline setting, mirror descent has the following guarantee:

Theorem 2. *Choose a generating function $\phi(z)$ that is 1-strongly convex w.r.t. $\|\cdot\|$. Then, Mirror Descent has*

$$\sum_{t=1}^{T} f(z_t) - f(z_*) \leq \frac{1}{\eta} D^{\phi}_{z_1}(z_*) + \sum_{t=1}^{T} \frac{\eta}{2} \|g_t\|_*^2,$$

where $g_t \in \partial f(z_t)$ is the sub-gradient of $f(\cdot)$ at z_t and $D^{\phi}_{z_1}(z_*)$ is the initial Bregman Divergence.

In the online setting, we have the following theorem:

Theorem 3. *Choose a generating function $\phi(z)$ that is 1-strongly convex w.r.t. $\|\cdot\|$. Then, Mirror Descent has*

$$\sum_{t=1}^{T} \ell_t(z_t) - \ell_t(z_*) \leq \frac{1}{\eta} D^{\phi}_{z_1}(z_*) + \sum_{t=1}^{T} \frac{\eta}{2} \|g_t\|_*^2,$$

for any benchmark $z_* \in Z$.

where $\ell_t(z)$ is convex but not necessarily differentiable, and $g_t \in \partial \ell_t(z_t)$ is the sub-gradient of $\ell_t(\cdot)$ at z_t.
The analysis and proof follow very similarly to mirror descent in offline setting in Lecture 12, and we just need to let $f(\cdot) \leftarrow \ell_t(\cdot)$ in the proof. You are encouraged to check this.

3.3 Optimistic Mirror Descent

Assume there is a good guess m_t of g_t, at every iteration, we make 2 updates, where first one is the same as mirror descent. The second update rely on the optimistic guess of the sub-gradient m_t.

1: **for** $t = 1, 2, \ldots$ **do**
2: $\quad z_{t-\frac{1}{2}} = \arg\min_{z \in C} \langle g_{t-1}, z \rangle + \frac{1}{\eta} D^{\phi}_{z_{t-\frac{3}{2}}}(z)$.
3: $\quad z_t = \arg\min_{z \in C} \langle m_t, z \rangle + \frac{1}{\eta} D^{\phi}_{z_{t-\frac{1}{2}}}(z)$.
4: **end for**

$$\text{Regret}_T(z_*) \leq \frac{1}{\eta} D^{\phi}_{z_1}(z_*) + \sum_{t=1}^{T} \frac{\eta}{2} \|g_t - m_t\|_*^2.$$

for any benchmark $z_* \in Z$.

From the bound, it the guess m_t is close to g_t, then the regret can be potentially better than $O(\sqrt{T})$ of those non-optimistic variants, e.g., Online Mirror Descent.

On the other hand, even if the guess m_t is a poor estimate of g_t, it can still have $O(\sqrt{T})$ regret as those non-optimistic variants. Specifically, if the size of both m_t and g_t is bounded by G, then the second term of the regret bound above can be upper-bounded as

$$\|g_t - m_t\|_*^2 \leq (\|g_t\|_* + \|m_t\|_*)^2 \leq 4G^2$$

Thus, the regret bound is still $O(\sqrt{T})$ with an appropriately chosen η.

4 Some notions of regret

Recall we discuss that it is impossible to design an algorithm that has an $o(T)$ upper bound of the following metric:

$$\sum_{t=1}^{T} \ell_t(z_t) - \sum_{t=1}^{T} \min_{z \in Z} \ell_t(z). \tag{1}$$

It is emphasized that we observe $\ell_t(\cdot)$ only after committing a point z_t in the decision space Z. For the metric defined on (1), when we do not commit a point that is the argmin of a loss function $\ell_t(\cdot)$ that we are going to receive, we will suffer some loss at each round, and consequently the value of (1) will grow linearly with T. This motivated us to consider "regret" with respect to a fixed benchmark action z_*:

$$\text{Regret}_T(z_*) = \sum_{t=1}^{T} \ell_t(z_t) - \sum_{t=1}^{T} \ell_t(z_*).$$

One question arises is that: Does an algorithm has an $O(\sqrt{T})$ regret imply it also has $O(\sqrt{T})$ cumulative loss?

Consider the following situation with two experts with the following loss sequences.

- Expert 1: $1, 1, 1, 1, 1, \ldots, 1,\ 0, 0, 0, 0, 0, \ldots, 0$
- Expert 2: $0, 0, 0, 0, 0, \ldots, 0,\ 1, 1, 1, 1, 1, \ldots, 1$

We can see that the cumulative loss for our algorithm is the following:

$$\begin{aligned} \sum_{t=1}^{T} \ell_t(z_t) &= \text{Regret}_T(z_*) + \sum_{t=1}^{T} \ell_t(z_*) \\ &\leq O\left(\sqrt{T}\right) + \frac{T}{2}, \text{ either expert as benchmark} \\ &= \Omega(T) \end{aligned}$$

Thus, the lower bound on the cumulative loss is still in linear of T.

We will present two remedies for this issue.
One way to solve this is with interval regret. Consider a time interval with the starting point $s \in [N]$ and the end point $e \in [N]$ such that $1 \le s \le e \le T$. The interval regret with respect to an interval consists of rounds $\{s, s+1, \ldots, e-1, e\}$

Theorem 4 (Interval Regret)**.** *The interval regret for interval I is defined as:*

$$R_I := \sum_{t \in I} \ell_t(z_t) - \min_{z \in Z} \sum_{t \in I} \ell_t(z).$$

If we could design an OCO algorithm that has $R_I = O(\sqrt{I})$ interval regret, then we can achieve sub-linear cumulative loss. To see this, consider the two-expert scenario above, and let I_1 be the first half of the round and interval 2 be the second half. Then the cumulative loss for any interval is

$$\sum_{t \in I} l_t(z_t) = R_I + \min_{z \in Z} \sum_{t \in I} \ell_t(z) = O\left(\sqrt{\frac{T}{2}}\right) + 0$$

Thus, we can see with 2 intervals, the total cumulative loss is on the order of $O\left(\sqrt{\frac{T}{2}}\right)$, which is sub-linear with respect to T.
Remark. There exists an algorithm that guarantees $R_I = O\left(\sqrt{|I| \log T}\right)$ for all intervals I's. This algorithm is insensitive to the particular interval in question.

Another remedy is to consider dynamic regret. In other words, we will allow the benchmark to be dynamic as oppose to a fixed point in the decision space.

Theorem 5 (Dynamic regret)**.** *Consider comparing a sequence of benchmarks $y_1, y_2, \ldots, y_T \in Z$.*

$$R_T(\{y_1, y_2, \ldots, y_T\}) = \sum_{t=1}^{T} \ell_t(z_t) - \sum_{t=1}^{T} \ell_t(y_t).$$

We can see if $y_1 = y_2 = \ldots, y_T = z_*$, then we recover the standard regret

$$\text{Regret}_T(z_*) = \sum_{t=1}^{T} \ell_t(z_t) - \sum_{t=1}^{T} \ell_t(z_*).$$

On the other hand, if $y_t = \arg\min_{y \in Z} \ell_t(z)$, then it becomes

$$\sum_{t=1}^{T} \ell_t(z_t) - \sum_{t=1}^{T} \min_{z} \ell_t(z).$$

Thus, we will try to aim somewhere in between the two above situations. In some sense, We need to express the dynamic regret in terms of the complexity of the benchmark sequence.

We have the following notions for complexity of the benchmark sequence. First, we consider the complexity in terms of how many times the benchmark is swapped in the sequence. Let S denote the number of switches:

$$S := 1 + \sum_{t=2}^{T} \mathbb{1}\{y_t \neq y_{t-1}\}$$

Remark. There exists an algorithm that guarantees the following bound of the dynamic regret:

$$R_T(\{y_1, y_2, \ldots, y_T\}) = \tilde{O}(\sqrt{ST}).$$

In this case, if S is a small constant, namely only a few switches during the sequence, then the dynamic regret is still sub-linear.
Another view for sequence complexity is the variation of the loss sequence, which is defined as

$$V_T = \sum_{t=2}^{T} \max_{z \in Z} |\ell_t(z) - \ell_{t-1}(z)|.$$

which can be understood as the cumulative differences between two consecutive loss functions. Putting all together, we have the following theorem:

Theorem 6. *If an algorithm has $R_I = \tilde{O}(\sqrt{|I|})$ for all intervals I, then it has*

$$R_T(\{z_1^*, z_2^*, \ldots, z_T^*\}) = \tilde{O}(\sqrt{T} + T^{2/3} V_T^{1/3}),$$

where $z_t^* = \arg\min_{z \in Z} \ell_t(z)$.

5 Online to Batch Conversion

Let us consider a scenario that we receive training data in a streaming fashion

$$\underbrace{(x_1, y_1)}_{\text{sample at round 1}}, \underbrace{(x_2, y_2)}_{\text{sample at round 2}}, \ldots, \underbrace{(x_T, y_T)}_{\text{sample at round } T},$$

and we apply an our algorithm to this scenario. On each round t, our OCO algorithm chooses some $z_t \in Z$ and suffers loss $\ell(z_t; x_t, y_t)$.

Theorem 7. *Assume that each* (x_t, y_t) *is generated in an i.i.d manner. Let* $L(z) := \mathbb{E}_{(x,y)}[\ell(z; x, y)]$ *and* $z_* := \arg\min_{z \in Z} L(z)$*. Let* $z_1, z_2, \ldots, z_T$ *be generated from an OCO algorithm. Then,*

$$\mathbb{E}\left[\frac{1}{T}\sum_{t=1}^{T} L(z_t)\right] \leq L(z_*) + \frac{1}{T}\mathbb{E}[\text{Regret}_T(z_*)].$$

Furthermore, if $L(\cdot)$ *is convex, then*

$$\mathbb{E}\left[L\left(\frac{1}{T}\sum_{t=1}^{T} z_t\right)\right] \leq L(z_*) + \frac{1}{T}\mathbb{E}[\text{Regret}_T(z_*)]. \tag{2}$$

Remark: If the underlying OCO algorithm enjoys a sub-linear regret, then the last term on (2) $\frac{1}{T}\mathbb{E}[\text{Regret}_T(z_*)] \to 0$ as $T \to \infty$. As $z_* \in Z$ is the best to minimize the population loss $L(\cdot)$, what (2) implies is that the OCO algorithm is a learning algorithm. Recall we also saw that an OCO algorithm (e.g., Online Mirror Descent) can also be used to solve (offline) optimization problems. This shows a deep interplay between optimization and learning.

5.1 Probability Recap

We need a couple of notions.

- We will consider a filtration $\{H_t\}$. Informally, we can think of the filtration H_t as the history up to round t.
- We will consider also a stochastic process $X_1, X_2, \ldots X_T$ where each X_t is measurable with respect to H_t. Informally, we can think of X_t is a deterministic function of the history H_t; once we have the realization of H_t, we know X_t!

Definition 2. *Martingale* $X_1, X_2, \ldots X_T$ *is a martingale with respect to a filtration* $\{H_t\}$*, if* $\mathbb{E}[|X_t|] < \infty$ *and* $\mathbb{E}[X_t|H_{t-1}] = X_{t-1}$.

Definition 3. *Martingale difference sequence The process* $U_1, U_2, \ldots, U_T$ *is a martingale difference sequence if* $\mathbb{E}[|U_t|] < \infty$ *and* $\mathbb{E}[U_t|H_{t-1}] = 0$.

Implications:

- Clearly, $U_t := X_t - X_{t-1}$ is a martingale difference sequence. Since

$$\mathbb{E}[U_t] = \mathbb{E}[X_t - X_{t-1}|H_{t-1}] = \mathbb{E}[X_t|H_{t-1}] - \mathbb{E}[X_{t-1}|H_{t-1}] \tag{3}$$

$$= X_{t-1} - \mathbb{E}[X_{t-1}|H_{t-1}] \tag{4}$$

$$= X_{t-1} - X_{t-1}, \tag{5}$$

where the second to last equality uses the definition of $\{X_t\}$ is a martingale, and the last one uses that x_{t-1} is measurable w.r.t. H_{t-1}.

- The unconditional expectation $\mathbb{E}[U_t] = 0$. This can be shown by an application of the law of total expectation.

5.2 Proof of Theorem 7

Lemma 2. *Let us define* $U_t := \big(\ell(z_t;(x_t,y_t)) - L(z_t)\big) - \big(\ell(z_*;(x_t,y_t)) - L(z_*)\big)$. *Then, we have*

$$\sum_{t=1}^{T} L(z_t) \leq L(z_*) + \text{Regret}_T(z_*) - \sum_{t=1}^{T} U_t$$

Proof.

$$\sum_{t=1}^{T} \big(L(z_t) - L(z_*)\big) = \sum_{t=1}^{T} \ell\left(z_t;(x_t,y_t)\right) - \ell\left(z_*;(x_t,y_t)\right) - U_t \tag{6}$$

$$\leq \sum_{t=1}^{T} \ell\left(z_t;(x_t,y_t)\right) - \inf_{z\in Z} \ell\left(z;(x_t,y_t)\right) - U_t \tag{7}$$

□

Now we are ready to prove Theorem 7.

Proof. (of Theorem 7) Taking the expectation on both sides of the inequality in Lemma 2, we have

$$\frac{1}{T}\mathbb{E}\left[\sum_{t=1}^{T} L(z_t)\right] \leq \frac{1}{T}L(z_*) + \frac{1}{T}\mathbb{E}\left[\text{Regret}_T(z_*)\right] - \frac{1}{T}\mathbb{E}\left[\sum_{t=1}^{T} U_t\right]. \tag{8}$$

Now observe that $\{U_t\}$ is a martingale difference sequence. Hence, by the property of a martingale difference sequence indicated in the recap (Subsection 5.1), we have $\mathbb{E}\left[\sum_{t=1}^{T} U_t\right] = 0$.

By combining the above results, we have completed the proof.

□

Bibliographic notes

For more details about Online Convex Optimization, see e.g., [1], [2], and [3].

References

[1] Shai Shalev-Shwartz Online Learning and Online Convex Optimization Foundations and Trends in Machine Learning. 2011.

[2] Elad Hazan. Introduction to Online Convex Optimization. Second Edition. The MIT Press. 2022.

[3] Francesco Orabona A Modern Introduction to Online Learning arXiv:1912.13213. 2023