1 Recap

1.1 Online Convex optimization framework

In this section we review the general framework for online convex optimization framework.
In this framework, we first make a prediction from feasible set and then receive a convex function and calculate the loss we suffered as the result of our prediction.
Rigorously speaking, consider a convex feasible set $S$, and convex (loss) functions $f_t, t \in \{1, 2, \ldots\}$.
The framework will be as following:

**input:** A convex set $S$

for $t = 1, 2, \ldots$
predict a vector $w_t \in S$ receive a convex loss function $f_t : S \rightarrow \mathbb{R}$ suffer loss $f_t(w_t)$

1.2 Regret Analysis

Regret of $\{w_1, w_2, \ldots, w_T\}$ w.r.t a point $u \in S$ is defined as following:

$$\text{Reg}_T(u) = \sum_{t=1}^{T} f_t(w_t) - f_t(u)$$  (1)
Additionally, regret w.r.t a set of competing points $U$ is defined as following:

$$\text{Reg}_T(U) = \max_{u \in U} \text{Reg}_T(u)$$  \hspace{1cm} (2)

### 1.3 Follow-The-Leader (FTL)

One approach for proposing a prediction is using the Follow-The-Leader (FTL) approach. In this approach, at time $t$, we choose a prediction that would have the best performance in the whole past time if chosen, i.e.,

$$w_t = \arg \min_{w \in S} \sum_{i=1}^{t-1} f_i(w), \quad \forall t$$  \hspace{1cm} (3)

**Lemma:** For FTL algorithm, we have the following upper bound for the regret:

$$\text{Reg}_T(u) \leq \sum_{t=1}^{T} (f_t(w_t) - f_t(w_{t+1})), \quad \forall u \in S$$  \hspace{1cm} (4)

### 1.4 Follow-the-Regularized-Leader (FTRL)

To improve the performance of FTL in case of jump in the prediction, the cost function of the FTL is regularized as following:

$$w_t = \arg \min_{w \in S} \sum_{i=1}^{t-1} f_i(w) + R(w), \quad \forall t$$  \hspace{1cm} (5)

which is called Follow-the-regularized-Leader (FTRL). **Lemma:** For FTRL algorithm, we have the following upper bound for the regret:

$$\text{Reg}_T(u) \leq R(u) - R(w_1) + \sum_{t=1}^{T} (f_t(w_t) - f_t(w_{t+1})), \quad \forall u \in S$$  \hspace{1cm} (6)


\section{FTRL in General Case}

Consider the FTRL algorithm defined as following:

\[
\mathbf{w}_t = \arg \min_{\mathbf{w} \in S} \sum_{i=1}^{t-1} f_i(\mathbf{w}) + R(\mathbf{w}), \quad \forall t \tag{7}
\]

where each \(f_i(\mathbf{w})\) is convex and \(R(\mathbf{w})\) is strongly convex.

\textbf{Definition:} \(f : S \to \mathbb{R}\) is \(\sigma\)-strongly convex w.r.t the norm \(||\cdot||\), if

\[
\forall \mathbf{u}, \mathbf{w} \in S, \forall \mathbf{z} \in \partial f(\mathbf{w}), \quad f(\mathbf{u}) \geq f(\mathbf{w}) + \langle \mathbf{z}, \mathbf{u} - \mathbf{w} \rangle + \frac{\sigma}{2} ||\mathbf{u} - \mathbf{w}||^2 \tag{8}
\]

As an example, \(R(\mathbf{w}) = \frac{1}{2}||\mathbf{w}||^2\) is 1-strongly convex w.r.t \(\ell_2\) norm.

Or,

\[R(\mathbf{w}) = \sum_{i=1}^d w[i] \log(w[i])\] is 1-strongly convex w.r.t \(\ell_1\) on \(S = \{\mathbf{w} : w > 0, ||\mathbf{w}||_1 \leq 1\}\).

From previous section, we know that:

\[\text{Reg}_T(\mathbf{u}) \leq R(\mathbf{u}) - R(\mathbf{w}_1) + \sum_{t=0}^{T} (f_t(\mathbf{w}_t) - f_t(\mathbf{w}_{t+1})), \quad \forall \mathbf{u} \in S \tag{9}\]

We need to bound the summation \(\sum_{t=1}^{T} (f_t(\mathbf{w}_t) - f_t(\mathbf{w}_{t+1}))\).

Define \(F_t(\mathbf{w}) = \sum_{i=1}^{t-1} f_i(\mathbf{w}) + R(\mathbf{w})\)

Based on the updating rule of FTRL we have:

\[F_t(\mathbf{w}_{t+1}) \geq F_t(\mathbf{w}_t) + \frac{\sigma}{2} ||\mathbf{w}_t - \mathbf{w}_{t+1}||^2\]

\[F_{t+1}(\mathbf{w}_t) \geq F_{t+1}(\mathbf{w}_t + 1_t) + \frac{\sigma}{2} ||\mathbf{w}_t - \mathbf{w}_{t+1}||^2\]

by summation and sorting we will have:
\[ \sigma^2 \| \mathbf{w}_t - \mathbf{w}_{t+1} \|^2 \leq f_t(\mathbf{w}_t) - f_t(\mathbf{w}_{t+1}) \leq L \| \mathbf{w}_t - \mathbf{w}_{t+1} \| \to \| \mathbf{w}_t - \mathbf{w}_{t+1} \| \leq \frac{L}{\sigma} \to \]

As the result, we will have:

\[ f_t(\mathbf{w}_t) - f_t(\mathbf{w}_{t+1}) \leq \frac{L^2}{\sigma} \rightarrow \text{Reg}_T(\mathbf{u}) \leq R(\mathbf{u}) - \min_{\mathbf{v} \in S} R(\mathbf{v}) 6T \frac{L^2}{\sigma} \quad (10) \]

**Theorem:** Consider the following conditions:

1) \( f_t \) is convex for all \( t \).
2) \( f_t \) is \( L \)-Lipschitz w.r.t some norm \( \| \cdot \| \).
3) \( R(.) \) is \( \sigma \)-strongly convex w.r.t norm \( \| \cdot \| \).

Then,

\[ \text{Reg}_T(\mathbf{u}) \leq R(\mathbf{u}) - \min_{\mathbf{v} \in S} R(\mathbf{v}) 6T \frac{L^2}{\sigma} \quad (11) \]

**Example 1:**
Consider \( S = \{ \mathbf{w} : \| \mathbf{w} \|_2 \leq B \} \), \( R(\mathbf{w}) = \frac{1}{\eta} \| \mathbf{w} \|_2^2 \), \( \eta = \frac{B}{L\sqrt{2T}} \rightarrow \text{Reg}_T(S) \leq BL\sqrt{(2T)} \)

**Example 2:**
Consider \( S = \{ \mathbf{w} : \| \mathbf{w} \|_1 \leq 1, \mathbf{w} \geq 0 \} \), \( R(\mathbf{w}) = \frac{1}{\eta} \sum_i w[i] \log w[i], \eta = \frac{\sqrt{\log(d)}}{L\sqrt{T}} \rightarrow \text{Reg}_T(S) \leq 2L \sqrt{(T\log(d))} \)