1 The Game

We will analyze a general online convex optimization problem, where we have a convex set \( \mathcal{W} \) and a sequence of loss functions \( f_t: \mathcal{W} \rightarrow \mathbb{R} \). However, the feedback that the player gets is a bandit feedback. The player only sees \( f_t(w_t) \) in each round.

**Bandit Gradient Descent Game**

| choose parameters \( \eta, \alpha, \delta \) where \( \alpha \in [0, 1] \). |
| \( v_1 = 0 \in \mathcal{W} \) |
| for \( t = 1, 2, \ldots, T \) |
| choose unit vector \( u_t \in \mathbb{R}^d \) uniformly at random |
| assign \( w_t = v_t + \delta u_t \in \mathcal{W} \) |
| play \( w_t \), observe \( f_t(w_t) \) |
| \( v_{t+1} = \prod_t (1 - \alpha)\mathcal{W}(v_t - \eta \hat{g}_t) \), where \( \hat{g}_t = \frac{d}{\delta} f_t(w_t)u_t \)

Note: Projection to \((1 - \alpha)\mathcal{W}\) is required to make sure that \( w_t \) stays inside the convex set.

![Figure 1: Graphical Interpretation of the Bandit Gradient Descent Game](image)

2 Regret Bound of Bandit Gradient Descent

The regret bound for online linear optimization is \( \mathcal{O}(\sqrt{T}) \) and the regret bound for the experts algorithm is \( \mathcal{O}(\sqrt{T \log d}) \). The expected regret is calculated for the bandit gradient descent game, since it is uniformly randomized:

\[
\mathbb{E}[\text{Regret}] \leq \mathcal{O}(T^{3/4})
\]
3 Analysis

We will need two tricks to analyze the algorithm: (1) one point gradient estimation and (2) expected gradient-descent.

3.1 One Point Gradient Estimation

Lemma 1. $|u| = 1$ and choose uniformly at random, $\delta > 0$,
\[
\nabla \hat{f}(x) = \mathbb{E}[\frac{d}{\delta}f(v + \delta u)u].
\]

Proof. When $d=1$, $u \in [-1, +1]$,
\[
\mathbb{E}[\frac{1}{\delta}f(v + \delta u)u] = \frac{1}{2} \frac{f(v + \delta)}{\delta} - \frac{1}{2} \frac{f(v - \delta)}{\delta},
\]
\[
\cong f'(v).
\]

It follows that
\[
\hat{f}(x) = \mathbb{E}[f(v + \delta u)],
\]
\[
\nabla \hat{f}(x) = \mathbb{E}[\frac{d}{\delta}f(v + \delta u)u],\text{ where } u : ||u||^2 < 1.
\]

$\hat{f}$ is the smoothed version $f$. Note that $\hat{f}$ is differentiable even though $f$ is not. \qed

3.2 Expected Gradient Descent

Regret bound for online gradient descent is:
\[
\sum_{t=1}^{T} f_t(w_t) - \min_{u \in W} \sum_{t=1}^{T} f_t(u) \leq \frac{B^2}{\eta} + \eta \frac{G^2T}{2}, \text{ where } \eta = BG\sqrt{T},
\]
\[
\sum_{t=1}^{T} f_t(w_t) - \min_{u \in W} \sum_{t=1}^{T} f_t(u) \leq BG\sqrt{T}.
\]

Lemma 2. In the randomized version, every round we will get $\hat{g}_t = \frac{d}{\delta}f_t(w_t)u_t$. $\mathbb{E}[g_t] = \nabla f(v_t)$ and $||g_t|| < G$. For optimum $\eta$:
\[
\mathbb{E} \left[ \sum_{t=1}^{T} f_t(w_t) \right] - \min_{u \in W} \sum_{t=1}^{T} f_t(u) \leq BG\sqrt{T}.
\]

Proof.
\[
h_t(w) = f_t(w) + w(\hat{g}_t - \nabla f_t(w_t)),
\]
\[
\nabla_w h_t(w)_{|w=w_t} = \nabla f_t(w_t) + \hat{g}_t - \nabla f_t(w_t),
\]
\[
= \hat{g}_t.
\]

Online gradient descent on the random function $h_t$ is equal to expected gradient descent on the fixed functions $f_t$. \qed