CSE599s, Spring 2014, Online Learning	Lecture 16 - 05/22/2014
Online Convex Optimization with Bandit Feedback	
Lecturer: Brendan McMahan	*Scribe: Durmus Karatay*

1 The Game

We will analyze a general online convex optmimization problem, where we have a convex set $\mathcal{W}$ and a sequence of loss functions $f_t : \mathcal{W} \to \mathbb{R}$. 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 η, α, δ where $\alpha \in [0, 1]$.
$v_1 = 0 \in \mathcal{W}$
for $t = 1, 2, ..., T$
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 $f_t(w_t)$
$v_{t+1} = \prod_{(1-\alpha)\mathcal{W}}(v_t - \eta \hat{g}_t)$, 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.

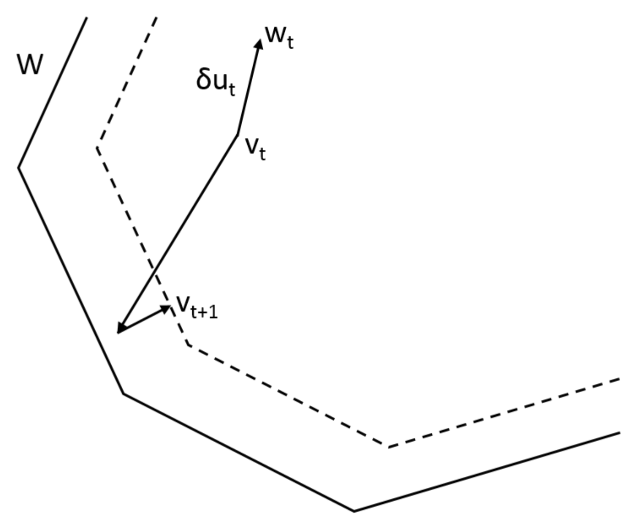

Figure 1: Graphical Interpretation of the Bandit Gradient Descent Game

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:

$$\mathbf{E}\left[\text{Regret}\right] \leq \mathcal{O}(T^{\frac{3}{4}})$$

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 choosen uniformly at random, $\delta > 0$,*

$$\nabla \hat{f}(x) = \mathbf{E}\,[\frac{d}{\delta} f(v + \delta u) u].$$

Proof. When d =1, $u \in [-1, +1]$,

$$\begin{aligned} \mathbf{E}\,[\frac{1}{\delta} f(v + \delta u) u] &= \frac{1}{2}\frac{f(v+\delta)}{\delta} - \frac{1}{2}\frac{f(v-\delta)}{\delta}, && (1)\\ &\cong f'(v). && (2) \end{aligned}$$

It follows that

$$\begin{aligned} \hat{f}(x) &= \mathbf{E}\,[f(v + \delta u)], && (3)\\ \nabla \hat{f}(x) &= \mathbf{E}\,[\frac{d}{\delta} f(v + \delta u) u], \text{ where } u : ||u||^2 < 1. && (4) \end{aligned}$$

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

3.2 Expected Gradient Descent

Regret bound for online gradient descent is:

$$\begin{aligned} \sum_{t=1}^{T} f_t(w_t) - \min_{u \in W} \sum_{t=1}^{T} f_t(u) &\le \frac{B^2}{\eta} + \eta \frac{G^2 T}{2}, \text{ where } \eta = BG\sqrt{T}, && (5)\\ \sum_{t=1}^{T} f_t(w_t) - \min_{u \in W} \sum_{t=1}^{T} f_t(u) &\le BG\sqrt{T}. && (6) \end{aligned}$$

Lemma 2. *In the randomized version, every round we will get $\hat{g}_t = \frac{d}{\delta} f_t(w_t) u_t$. $\mathbf{E}\,[g_t] = \nabla f(v_t)$ and $||g_t|| < G$. For optimum η:*

$$\mathbf{E}\,[\sum_{t=1}^{T} f_t(w_t)] - \min_{u \in W} \sum_{t=1}^{T} f_t(u) \le BG\sqrt{T}.$$

Proof.

$$\begin{aligned} h_t(w) &= f_t(w) + w(\hat{g}_t - \nabla f_t(w_t)), && (7)\\ \nabla_w h_t(w)|_{w=w_t} &= \nabla f_t(w_t) + \hat{g}_t - \nabla f_t(w_t), && (8)\\ &= \hat{g}_t. && (9) \end{aligned}$$

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