**CSE599s, Spring 2012, Online Learning** Lecture 5 - 04/10/2012

# Online Subgradient Descent (OGD)

*Lecturer: Ofer Dekel* *Scribe: Alex Cheng*

## 1 Subgradient

**Definition 1.** *subgradient $g \in \mathbb{R}^n$ of a convex function $f : \mathcal{W} \rightarrow \mathbb{R}$ at a point $\hat{w} \in \mathcal{W}$ s.t.*
$\forall w \in \mathcal{W} f(w) \geq f(\hat{w}) + (w - \hat{w}) \cdot g$

**Definition 2.** *subdifferential at $\hat{w} \in \mathcal{W}$ is the set of all subgrads of $f$ at $\hat{w}$ : $\partial f(\hat{w})$*

**Lemma 3.** $\sum_{t=1}^{T} f_t(w_t) + f_t(w^*) \leq \sum_{t=1}^{T} w_t \cdot g_t - w^* \cdot g_t$ $\quad (g_t \in \partial f_t(w_t))$

## 2 FTRL

FTRL has 3 weaknesses

1. dep on $T$ (finite horizon) $\rightarrow$ doubling
2. linear funcs $\rightarrow$ subgrads $\rightarrow$ all convex G-Lipschitz func
3. $\mathcal{W} = \mathbb{R}^n$

## 3 Online Subgradient Descent (OGD)

with infinite horizen and projections onto the feasible set $\mathcal{W}$

| **OGD** (Zinkevich '03) |
| --- |
| $w_1 = (0, \ldots, 0) \in \mathbb{R}^n$ (assume $0 \in \mathcal{W}$)<br>for $f = 1, 2, \ldots$<br>- player plays $w_t$<br>- adversary plays $f_t \in \mathcal{F}$ (set of convex G-Lipshitz func)<br>Update:<br>1) GD: $w'_{t+1} = w_t - \eta_t g_t$ ($g_t$ subgrad $f_t$ at $w_t$, $\eta_t$ is learning rate $\eta_t \geq \eta_{t+1} > 0$)<br>2) Projection: $w_{t+a} = \Pi_w(w'_{t+1}) \equiv \underset{w \in \mathcal{W}}{\text{argmin}} \|w'_{t+1} - w\|_2^2$ (convex optimization) |

**Lemma 4.** *(Kolmogorov) if $w' \in \mathcal{W}, w = \Pi_{\mathcal{W}}(w'), w^* \in \mathcal{W}$ then $(w' - w)\dot{(}w^* - w) \leq 0$*

*Proof.* Consider $\hat{w} \in \mathcal{W}$ s.t. $(w' - \hat{w})\dot{(}w^* - \hat{w}) > 0$ and we will show $\hat{w} \neq \Pi_{\mathcal{W}}(w')$

Define $Z(\lambda) = (1 - \lambda)\hat{w} + \lambda w^*$ , $1 \; \lambda \in [0, 1]$
$= \hat{w} + \lambda(w^* - \hat{w})$

we will show: along this line, there is a point that is closer to $w'$
Note: from convexity $Z(\lambda) \in \mathcal{W}$ for all $\lambda \in [0, 1]$

$$\begin{aligned}\|w' - Z(\lambda)\|^2 &= \|w' - \hat{w} - \lambda(w^* - \hat{w})\|^2 \\ &= \|w' - \hat{w}\|^2 + \lambda^2\|w^* - \hat{w}\|^2 - 2\lambda(w' - \hat{w})(w^* - \hat{w})\end{aligned}$$

let $h(\lambda) = \lambda^2\|w^* - \hat{w}\|^2 - 2\lambda(w' - \hat{w})(w^* - \hat{w})$ the roots are 0 and $\frac{(w'-\hat{w})(w^*-\hat{w})}{\|w^*-\hat{w}\|^2}$
$h(\lambda) < 0$ for $\lambda \in (0, \frac{(w'-\hat{w})(w^*-\hat{w})}{\|w^*-\hat{w}\|^2})$
$\Rightarrow \|w' - Z(\lambda)\|^2 < \|w' - \hat{w}\|^2$
$\Rightarrow$ any $\lambda$ in the range where $h(\lambda) < 0$ produces a point $Z(\lambda) \in \mathcal{W}$ that is closer to $w'$

□

**Theorem 5.** $w' \notin \mathcal{W}, w = \Pi_w(w'), w^* \in \mathcal{W}, \frac{1}{2}\|w' - w\|^2 + \frac{1}{2}\|w - w^*\|^2 \leq \frac{1}{2}\|w^* - w'\|^2$

"reverse triangle inequality"

*Proof.* $\frac{1}{2}\|w'\|^2 + \frac{1}{2}\|w\|^2 - w' \cdot w + \frac{1}{2}\|w\|^2 + \frac{1}{2}\|w^*\|^2 - w \cdot w^* - \frac{1}{2}\|w^*\|^2 - \frac{1}{2}\|w'\|^2 + w^* \cdot w'$
$= w \cdot w - w' \cdot w - w \cdot w^* + w^* \cdot w'$
$= w'(w^* - w) - w(w^* - w) = (w' - w)(w^* - w) \leq 0$ (by lemma) □

## 3.1 Regret Bound

Potential func $\Phi(w, w^*) = \frac{1}{2}\|w - w^*\|^2$
choose $w^* \in \mathcal{W}$

### 3.1.1 Projection

from theorem: $\Phi(w'_{t+1}, w^*) - \Phi(w_{t+1}, w^*) \geq \frac{1}{2}\|w'_{t+1} - w_{t+1}\|^2 \geq 0$
(projection gets you closer to $w^*$)

### 3.1.2 Gradient Descent

$\Phi(w_t, w^*) - \Phi(w'_{t+1}, w^*)$
$= \frac{1}{2}\|w_t - w^*\|^2 - \frac{1}{2}\|w'_{t+1} - w^*\|^2$ definition of $\Phi$
$= \frac{1}{2}\|w_t - w^*\|^2 - \frac{1}{2}\|w_t - w^* - \eta_t g_t\|^2$ definition of GD
$= \frac{1}{2}\|w_t - w^*\|^2 - \frac{1}{2}\|w_t - w^*\|^2 - \frac{1}{2}\eta_t(w_t - w^*) \cdot g_t\|^2$
$= -\frac{1}{2}\eta_t^2\|g_t\|^2 + \eta_t(w_t - w^*) \cdot g_t$

$\iff$

$$\begin{aligned}(w_t - w^*) \cdot g_t &= \frac{1}{\eta_t}\Phi(w_t, w^*) - \Phi(w'_{t+1}, w^*) + \frac{1}{2}\eta_t\|g_t\|^2 \\ &\leq \frac{1}{\eta_t}(\Phi(w_t, w^*) - \Phi(w'_{t+1}, w^*) + \Phi(w'_{t+1}, w^*) - \Phi(w_{t+1}, w^*)) + \frac{1}{2}\eta_t\|g_t\|^2 \text{ (from projection)}\end{aligned}$$

$$\begin{aligned}\sum_{t=1}^T f_t(w_t) + f_t(w^*) &\leq \sum_{t=1}^T w_t \cdot g_t - w^* \cdot g_t \text{ (from lemma)} \\ &\leq sum_{t=1}^T \frac{1}{\eta_t}(\Phi(w_t, w^*) - \Phi(w_{t+1}, w^*)) + \sum_{t=1}^T \frac{1}{2}\eta_t\|g_t\|^2 \\ &= \frac{1}{\eta_1}\Phi(w_1, w^*) - \frac{1}{\eta_T}\Phi(w_{T+1}, w^*) + \sum_{t=2}^T(\frac{1}{\eta_T} - \frac{1}{\eta_{t-1}})(\Phi(w_t, w^*)) + \sum_{t=1}^T \frac{1}{2}\eta_t\|g_t\|^2\end{aligned}$$

| **assume** |
| --- |
| $f_t$ are G-Lipschitz $\Rightarrow \|g_t\|^2 \leq G^2$ |
| $\|w^*\| \leq R\ \forall t\|w_t\| \leq R \Rightarrow \frac{1}{2}\|w_t - w^*\|^2 \leq 2R^2$ |

$$\leq \frac{2}{\eta_1} R^2 - \frac{1}{\eta_T}\Phi(w_{T+1}, w^*) + 2\sum_{t=2}^{T}(\frac{1}{\eta_T} - \frac{1}{\eta_{t-1}})R^2 + \sum_{t=1}^{T}\frac{1}{2}\eta_t^2 G^2$$

Note: $\frac{1}{\eta_T}\Phi(w_{T+1}, w^*) > 0$, so substracting +'ve, not needed for upper bound so we can elimate that term

$$= 2R^2(\frac{1}{\eta_T} + 2\sum_{t=2}^{T}(\frac{1}{\eta_T} - \frac{1}{\eta_{t-1}}) + \frac{G^2}{2}\sum_{t=1}^{T}\eta_t$$
$$= 2R^2\frac{1}{\eta_T} + \frac{G^2}{2}\sum_{t=1}^{T}\eta_t$$

the optimal value is $\eta_t = \eta \cdot \frac{1}{\sqrt{T}}, subsitutingthatin$

$$= \frac{2R^2\sqrt{T}}{\eta} + \frac{G^2}{2}T\eta\frac{1}{\sqrt{T}} \Rightarrow \eta = \frac{2R}{G}$$
$$= RG\sqrt{T} + RG\sqrt{T} = 2RG\sqrt{T}$$

this is the finite horizon choice. for the infinite horizon case (if $T$ is not known), set $\eta_t = \frac{\eta}{\sqrt{t}}$

$$Regret \leq \frac{2R^2}{\eta_T} + \frac{G^2}{2}\sum_{t=1}^{T}\eta_t$$
$$= \frac{2R^2\sqrt{T}}{\eta} + \frac{G^2}{2}\eta\sum_{t=1}^{T}\frac{1}{\sqrt{t}}$$
$$\leq \frac{2R^2\sqrt{T}}{\eta} + \frac{G^2}{2}\eta\sqrt{T}$$

set $\eta = \frac{R\sqrt{2}}{G}$

$$= \sqrt{2}RG\sqrt{T} + RG\sqrt{2}\sqrt{T}$$
$$= 2\sqrt{2}RG\sqrt{T}$$