Online Linear & Convex Optimization

Logistics:

- Short Quiz on Thursday.
- If you cannot make it,
  email us (Fred + Haifeng + Aditya)

$$y_t = \langle w^*, x_t \rangle + \xi_t$$

mean-zero noise

$t=1$ to T

Nature gives x_t.

We predict $\hat{y}_t$.

Nature shows us y_t.

$$\text{Regret} = \sum_{t=1}^{T} (y_t - \hat{y}_t)^2$$

$$\mathbb{E}\,\text{Regret} = O(\sqrt{T})$$

based on analysis of $\|w^* - w_t\|^2$

online gradient descent

$$w_0 = 0$$

for

$$w_t = w_{t-1} - \eta \nabla (y_t - \langle w_{t-1}, x_t \rangle)^2$$

Online Convex Optimization Game

For $t=1$ to T:

I play a vector $w_t \in \mathbb{R}^d$ (or K)

Nature shows us a convex differentiable loss f_t

Suffer loss $f_t(w_t)$.

$$\text{Regret} = \sum_{t=1}^{T} f_t(w_t) - \inf_{w \in \mathbb{R}^d} \sum_{t=1}^{T} f_t(w)$$

OGD: ~~$w_0 = 0$~~ $w_1 = 0$

$$w_{t+1} = w_t - \eta \nabla f_t(w_t)$$

("Agnostic") Online linear regression

For $t=1$ to T

Nature shows me x_t

I play w_t

Nature shows me y_t $\longleftrightarrow$ Know f_t

I suffer $\left(y_t - \langle w_t, x_t \rangle\right)^2 = f_t(w_t)$

$$\text{Regret} = \sum \left(y_t - \langle w_t, x_t \rangle\right)^2 - \inf_w \sum \left(y_t - \langle w, x_t \rangle\right)^2$$

OGD $\quad w_{t+1} = w_t - \eta \nabla f_t(w_t)$

Online Convex Optimization via OGD.

Goal: Prove $O(\sqrt{T})$ regret

size of x_t, etc

z_t in book

Step 0: wlog, assume f_t are linear, $f_t(w) = \langle w, \underbrace{\nabla f_t}_{z_t} \rangle$

Observation 1: $w_0 = 0$, $w_{t+1} = w_t - \eta \nabla f_t(w_t)$

Claim: $$w_{t+1} = \underset{w \in \mathbb{R}^d}{\operatorname{argmin}} \underbrace{\sum_{s=1}^{t} f_s(w) + \frac{1}{2\eta}\|w\|^2}_{F_t(w)}$$

$\nabla F_{t-1}(w_t) = 0.$

$\Updownarrow$

$w_t = \operatorname{argmin} F_{t-1}(w)$

$$F_t(w) = F_{t-1}(w) + f_t(w)$$

$$\nabla F_t(w) = \nabla F_{t-1}(w) + \nabla f_t(w)$$

so $$\nabla F_t(w_t) = \underbrace{\nabla F_{t-1}(w_t)}_{0} + \nabla f_t(w_t)$$

$$= \nabla f_t(w_t)$$

$$w_{t+1} = w_t - \eta \nabla f_t(w_t) = w_t - \eta \nabla F_t(w_t)$$

<u>Fact:</u> $$F_{t-1}(w) = F_{t-1}(w_t) + \frac{1}{2\eta}\|w - w_t\|^2$$

by Taylor expansion of F_{t-1} at w_t

$\nabla F_{t-1}(w_t) = 0$

So

$$F_t(w) = F_{t-1}(w_t) + \frac{1}{2\eta}\|w - w_t\|^2 + f_t(w)$$

$$\min_w F_t(w) = \min_w \left[F_{t-1}(w_t) + f_t(w) + \frac{1}{2\eta}\|w - w_t\|^2 \right]$$

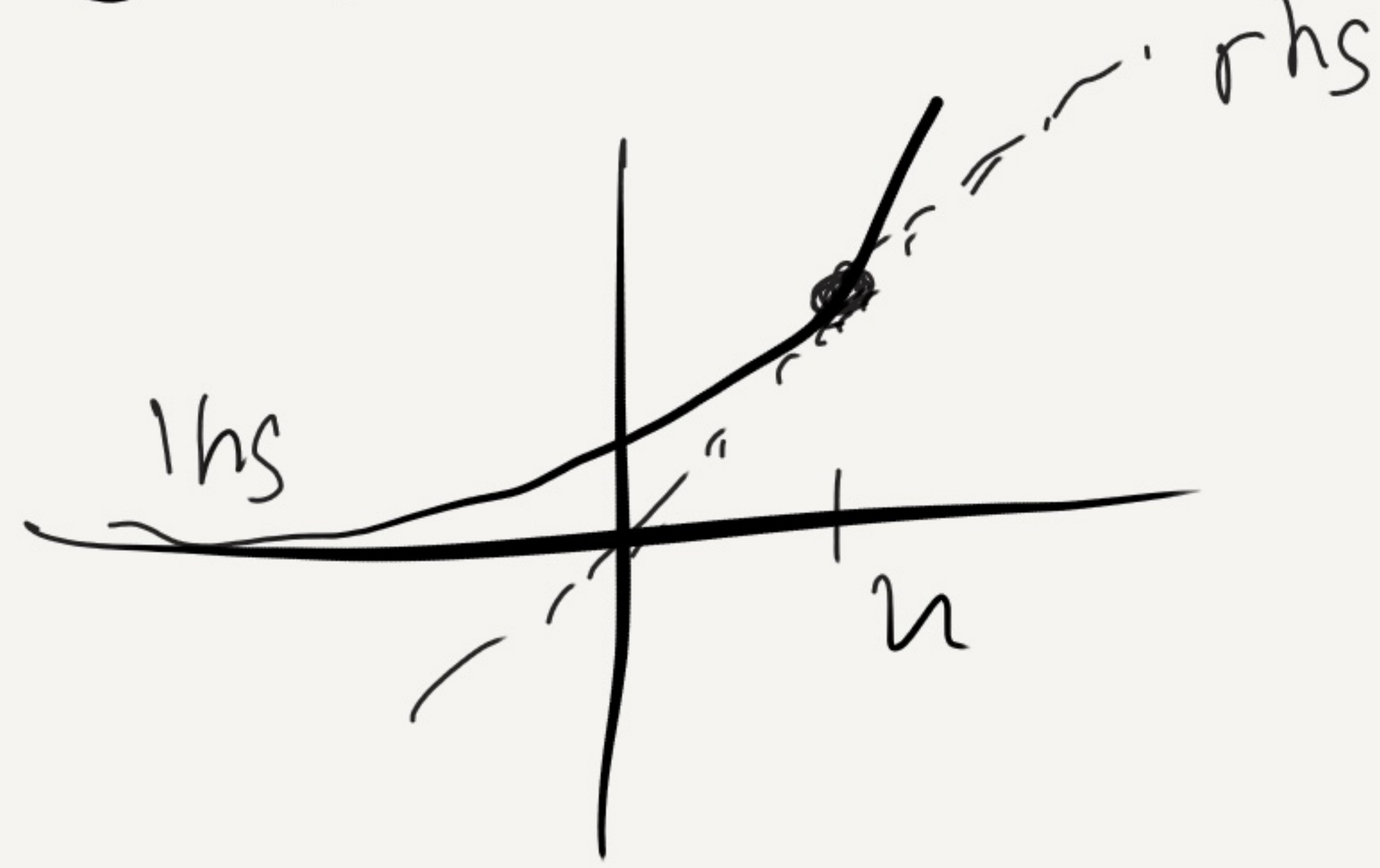
Set $\nabla = 0 \Longleftrightarrow$ optimizer is $\underset{\substack{\parallel \\ w_{t+1}}}{w_t} - \eta \nabla f_t(w_t)$ $\square$

Goal: $\forall w$

$$\underbrace{\sum_{t=1}^{T} f_t(w)}_{\text{Loss of } w} \ge \underbrace{\sum_{t=1}^{T} f_t(w_t)}_{\text{Our total loss}} - O(\sqrt{T})$$

(Idea: Use $\forall w \; f(w) \ge f(u) + \langle \nabla f(u), w - u \rangle$

$\min_{w} f(w) \le f(u)$



)

Lets prove it! (Note: skip from here to next slide)

Claim:

$$\sum_{t=1}^{T} f_t(w) \geq \sum_{t=1}^{T}\left(f_t(w_t) + \langle \nabla f_t, w - w_t \rangle\right)$$

$$= \sum_{t=1}^{T}\left(f_t(w_t) + \langle \nabla f_t, w_{t+1} - w_t \rangle\right)$$

$$+ \sum_{t=1}^{T}\left(\langle \nabla f_t, w - w_{t+1} \rangle\right)$$

Recall: $\nabla f_t(w_{t+1}) = 0.$

$$f_t(w) = f_t(w_{t+1}) + \langle \nabla f_t, w - w_{t+1} \rangle$$

Key claim #2

$$\underbrace{\frac{1}{2\eta}\|w_1\|^2}_{0} + \sum_{t=1}^{T} f_t(w_{t+1}) \le \overbrace{\sum_{t=1}^{T} f_t(w) + \frac{1}{2\eta}\|w\|^2}^{F_t(w)} \quad \forall w$$

Pf by induction: Base case $0 \le \frac{1}{2\eta}\|w\|^2$

assume $T-1$ prove T

Know by IH: $$\frac{1}{2\eta}\|w_1\|^2 + \sum_{t=1}^{T-1} f_t(w_{t+1}) \le \underbrace{\sum_{t=1}^{T-1} f_t(w_{T+1}) + \frac{1}{2\eta}\|w_{T+1}\|^2}$$

$$w_{T+1} = \arg\min F_t(w)$$

$$F_t(w_{T+1}) - f_t(w_{T+1})$$

$$\le F_T(w) - f_t(w_{T+1}) \quad \square$$

$$\sum_{t=1}^{T} f_t(w_t) = \sum_{t=1}^{T} \left(f_t(w_t) - f_t(w_{t+1}) \right)$$

$$+ \sum_{t=1}^{T} f_t(w_{t+1})$$

$\leq \eta G^2 T$

$$\leq \sum_{t=1}^{T} \left(f_t(w_t) - f_t(w_{t+1}) \right)$$

$$+ \sum_{t=1}^{T} f_t(w) + \frac{1}{2\eta} \|w\|^2$$

$\forall w$

by Claim 2

(so if $\eta = \frac{1}{\sqrt{T}}$, regret $\leq G^2\sqrt{T} + \|w\|^2\sqrt{T}/2$)

vs w

$$f_{t+1}(w_t) = f_{t+1}(w_{t+1}) + \langle \nabla f_{t+1}, w_t - w_{t+1} \rangle$$

$$f_{t+1}(w_t) - f_{t+1}(w_{t+1}) \le \langle \nabla f_{t+1}, w_t - w_{t+1} \rangle$$

$$\le \underbrace{\|\nabla f_{t+1}\|}_{\le G} \, \eta \underbrace{\|\nabla f_t\|}_{G}$$