Online Convex Optimization

Yoav Freund

February 17, 2020

Material follows Chapters 1,2,3 of “Online Convex Optimization” / Elad Hazan.

Outline

Online Linear Optimization

- Instance: $(\mathbf{x}_t, y_t) \in \mathbb{R}^d \times \mathbb{R}$
- Predictor: $\mathbf{w}_t \in \mathbb{R}^d$
- Loss $\ell(\mathbf{w} \cdot \mathbf{x}, y)$ (online regression = square loss)
- Regret: $\mathbf{R}_t(\mathbf{u}) = \sum_{i=1}^{t} \left[\ell(\mathbf{w}_t \cdot \mathbf{x}_t, y_t) - \ell(\mathbf{u} \cdot \mathbf{x}_t, y_t)\right]$

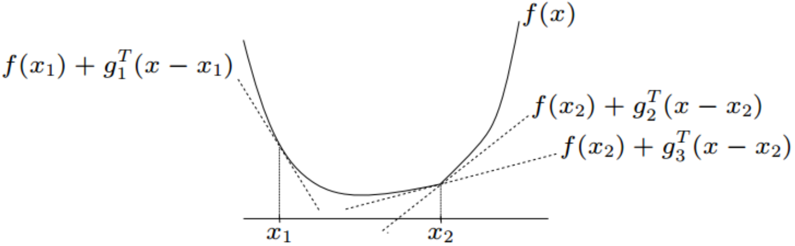
Online convex Optimization

- Optimizer chooses point:$\mathbf{x}_t \in \mathbb{R}^d$
- Adversary chooses convex loss function $f_t : \mathbb{R}^d \to \mathbb{R}$
- optimizer chooses Loss $f_t(\mathbf{x}_t)$
- Regret: $\mathbf{R}_T = \sup_{f_1,\ldots,f_T} \left[\sum_{t=1}^{T} f_t(\mathbf{x}_t) - \min_{\mathbf{u}} \sum_{t=1}^{T} f_t(\mathbf{u})\right]$

Standard convex optimization

- not online convex optimization (CO) has been studied much longer than Online convex optimization (OCO)
- OCO bounds use measures of convexity from CO.
- f is a given convex function.
- $\mathcal{K}$ is a convex set.
- Goal: find $\mathbf{x}^* = \text{argmin}_{\mathbf{x}\in\mathcal{K}} f(\mathbf{x})$
- Method: gradient descent.
- rate of convergence: rate at which $h_t = f(\mathbf{x}_t) - f(\mathbf{x}^*)$ decreases.

The sub-gradient

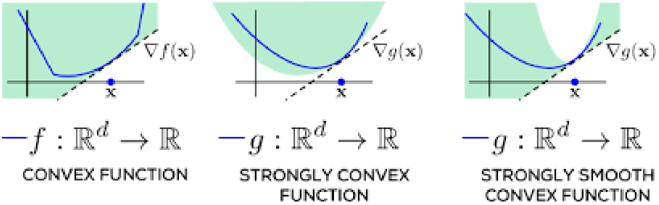


- $f : \mathbb{R}^d \to \mathbb{R}$ is convex.
- $\nabla f(\mathbf{x})$ is the set of vectors $\mathbf{g} \in \mathbb{R}^d$ such the $\forall \mathbf{y}, f(\mathbf{y}) \geq f(\mathbf{x}) + \mathbf{g} \cdot (\mathbf{y} - \mathbf{x})$
- If f is differentiable at x, then $\nabla f(\mathbf{x})$ has only one element.
- Otherwise $\nabla f(\mathbf{x})$ is a continuously infinite set.

Basic Gradient descent

1. **input:** f, T initial point $\mathbf{x}_1 \in \mathcal{K}$, sequence of step sizes $\{\eta_t\}$
2. For $t = 1, \ldots, T$ do:
3. Update: $\mathbf{y}_{t+1} = \mathbf{x}_t - \eta_t \nabla f(\mathbf{x}_t)$
4. Project: $\mathbf{x}_{t+1} = \Pi_{\mathcal{K}}(\mathbf{y}_{t+1})$
5. End For
6. Return $\mathbf{x}_{t+1}$

Degrees of convexity



- $f(x)$ is convex if $\forall \mathbf{x}, \mathbf{y}, f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x}) \cdot (\mathbf{y} - \mathbf{x})$
- $f(x)$ is α-strongly convex if
  $\forall \mathbf{x}, \mathbf{y}, f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x}) \cdot (\mathbf{y} - \mathbf{x}) + \frac{\alpha}{2}\|\mathbf{x} - \mathbf{y}\|^2$
- $f(x)$ is β-smooth if
  $\forall \mathbf{x}, \mathbf{y}, f(\mathbf{y}) \leq f(\mathbf{x}) + \nabla f(\mathbf{x}) \cdot (\mathbf{y} - \mathbf{x}) + \frac{\beta}{2}\|\mathbf{x} - \mathbf{y}\|^2$

Conditioning number

- If *f* is both α-strongly convex and β-smooth we say it is γ-well conditioned where: $\gamma = \frac{\alpha}{\beta} \leq 1$
- What is the meaning of $\gamma = 1$?

Basic Bound on Basic Gradient Descent

- If f is γ-well conditioned.
- and $\eta_t = \frac{1}{\beta}$
- $h_{t+1} \leq h_1 \exp\left(\frac{\gamma t}{4}\right)$

Reduction to smooth, not strongly convex functions

Reduction to non-smooth, strongy convex functions

Reduction to convex functions

Online Gradient Descent

$f_t(\mathbf{x}_t)$ instead of $f(\mathbf{x}_t)$