Notes: L^* bounds

He-Zhe Lin

January 26, 2021

We want to show a kind of regret bound that depends on the cumulative competitor loss.

Lemma 1. Let $V \subseteq \mathbb{R}^d$ a closed non-empty convex set with diameter D, i.e. $\max_{x,y\in V} \|x-y\|^2 \leq D$. Let $\ell_1, \dots, \ell_T$ an arbitrary sequence of convex functions $\ell_t : \mathbb{R}^d \to (-\infty, +\infty]$ subdifferentiable in open sets containing V for all t. Set $\eta_t = \frac{\sqrt{2}D}{2\sqrt{\sum_{i=1}^t \|g_i\|_2^2}}$, then for all u, the following regret bound holds

$$\text{Regret}_T(u) = \sum_{t=1}^T \ell_t(x_t) - \sum_{t=1}^T \ell_t(u) \leq \sqrt{2}D\sqrt{\sum_{t=1}^T \|g_t\|_2^2}$$

If in addition, we know that the loss functions are L-smooth, then we can obtain the L^* bound.

Definition 1 (Smooth Function)**.** Let $f : V \to \mathbb{R}$ differentiable. We say that f is L-smooth w.r.t $\|\cdot\|$ if $\|\nabla f(x) - \nabla f(y)\|_* \leq L\|x-y\|$ for all $x, y \in V$.

Remark 1. In fact, the definition here is the same as ”Lipschitz gradient” in the lecture of optimization by Yen-Huan. Note that a function with lipschitz gradient implies the smoothness of the function.

In the following we will need the property that the square of the dual norm of the gradient at a point is bounded by $2L$ times of the difference of the function value at that point and global minimum. The proof is from Theorem 2.1.5 in *Lectures on Convex Optimization* by Nestrov.

Lemma 2. Let $f : \mathbb{R}^d \to \mathbb{R}$ be L-smooth and bounded from below; then for all $x \in \mathbb{R}^d$, we have

$$\|\nabla f(x)\|_*^2 \leq 2L(f(x) - f(x^*)),$$

where x^* denotes the global minimizer of f.

Proof. By smoothness of f, we have

$$f(x^*) = \min_{y\in\mathbb{R}^d} f(y) \leq \min_{y\in\mathbb{R}^d} f(x) + \langle \nabla f(x), y - x\rangle + \frac{L}{2}\|y-x\|^2$$

Let $\|y - x\| = r$. Given any r, we can find a y such that $\langle \nabla f(x), y-x\rangle = -r\|\nabla f(x)\|_*$. Hence, write the inequality to be

$$f(x^*) \leq \min_{r\geq 0} f(x) - r\|\nabla f(x)\|_* + \frac{L}{2}r^2$$

Completing the square of the objective function, we have

$$f(x^*) \le \min_{r \ge 0} f(x) - \frac{1}{2L}\|\nabla f(x)\|_*^2 + \left(\sqrt{\frac{1}{2L}}\|\nabla f(x)\|_* - \sqrt{\frac{L}{2}}r\right)^2 = f(x) - \frac{1}{2L}\|\nabla f(x)\|_*^2$$

It follows that

$$\|\nabla f(x)\|_*^2 \le 2L(f(x) - f(x^*))$$

□

WLOG, we assume that the loss functions are bounded below by 0. Combining the result from the previous lemmas, we have

$$\text{Regret}_T(u) = \sum_{t=1}^{T} \ell_t(x_t) - \sum_{t=1}^{T} \ell_t(u) \le 2D\sqrt{L\sum_{t=1}^{T} \ell_t(x_t)}$$

Let $\sum_{t=1}^{T} \ell_t(x_t) = F$ and $\sum_{t=1}^{T} \ell_t(u) = F_u$, we have that $F - F_u \le 2D\sqrt{MF_u}$. Taking square of both side, we have

$$F^2 - 2FF_u + F_u^2 \le 4D^2MF_u.$$

Write the inequality as a quadratic function of F, we have

$$F^2 - (2F_u + 4D^2M)F + F_u^2 \le 0$$

Hence we obtain a upper bound of F:

$$F \le \frac{2F_u + 4D^2M + \sqrt{16D^4M^2 + 16F_uD^2M}}{2} \le F_u + 4D^2M + 2D\sqrt{MF_u}.$$

It follows that

$$\text{Regret}_T(u) = \sum_{t=1}^{T} \ell_t(x_t) - \sum_{t=1}^{T} \ell_t(u) \le 4D^2M + 2D\sqrt{M\sum_{t=1}^{T} \ell_t(u)}.$$

The worst case is still $O(\sqrt{T})$; however, if the competitor performs well, then the regret is small. In the best scenario, when the cumulative loss of the competitor is 0, then the regret is simply a constant.