CS292F Lecture 6 subgradient and prox gradient

Assumption: G-Lipschitz f. $|f(x)-f(y)| \le G\|x-y\|_2$

$$\boxed{\text{Thm. } \lim_{k\to\infty} f(x^{(k)}_{best}) = f^*}$$

when t_k chosen to be $\lim_k \sum t_k = \infty$ $\lim_{k\to\infty} \sum t_k^2 < \infty$

Proof.

$$\|x^{(k)}-x^*\|_2^2 = \|x^{(k-1)} - t_k g^{(k-1)} - x^*\|_2^2$$

$$= \|x^{(k-1)}-x^*\|_2^2 + t_k^2 \|g^{(k-1)}\|^2 \underbrace{- 2t_k (g^{(k-1)})^T (x^{(k-1)}-x^*)}$$

$$= \|x^{(k-1)}-x^*\|^2 + t_k^2\|g^{(k-1)}\|^2 + 2t_k (g^{(k-1)})^T(x^*-x^{(k-1)})$$

$$\le \|x^{(k-1)}-x^*\|^2 + t_k^2\|g^{(k-1)}\|^2 + 2t_k(f^* - f(x^{(k-1)}))$$

$$\le \|x^{(k-1)}-x^*\|^2 + t_k^2 G^2 - 2t_k(-f^* + f(x^{(k-1)}))$$

Alg:

$$x^{(k)} = x^{(k-1)} - t_k g^{(k-1)}$$

Convexity, g is subgradient

$$f(x^*) \ge f(x^{(k-1)}) + \underbrace{(g^{(k-1)})^T (x^* - x^{(k-1)})}$$

By telescoping for k steps, we get

$$\|x^{(k)}-x^*\|^2 \le \|x^{(0)}-x^*\|^2 + \underbrace{\sum_{l=1}^{k} t_l^2}_{<+\infty} G^2 - \underbrace{2\sum_{l} t_l}_{\infty \text{ as } k \text{ gets larger}} (-f^* + f(x^{(l-1)}))$$

$$2\sum_{\ell=1}^{k} t_\ell \left(f(x^{(\ell-1)}) - f^*\right) \le \|x^{(0)} - x^*\|^2 - \|x^{(k)} - x^*\|^2 + \sum_{\ell=1}^{k} t_\ell^2 G^2$$

$$\le$$

$$\left(2\sum_{\ell=1}^{k} t_\ell\right)\left(f(x_{best}^{k}) - f^*\right)$$

$$f(x_{best}^{k}) - f^* \le \frac{\|x^{(0)} - x^*\|^2}{2\sum_{\ell=1}^{k} t_\ell} + \frac{\sum_{\ell=1}^{k} t_\ell^2 G^2}{2\sum_{\ell=1}^{k} t_\ell}$$

Substitute our condition on diminishing t_ℓ

We complete the proof of the theorem. $\square$

Run for K iterations and use a fixed learning rate t

$\|x^{(0)} - x^*\| \le B$

$$f(x_{best}^{(k)}) - f^* \le \frac{1}{2}\left(\frac{B^2}{kt} + \frac{kt^2G^2}{kt}\right)$$

$$= \frac{1}{2}\left(\frac{B^2}{kt} + tG^2\right) \qquad t = \frac{B}{G\sqrt{K}}$$

$$O\left(\frac{1}{\varepsilon^2}\right) \qquad \simeq O\left(\frac{BG}{\sqrt{k}}\right) \qquad \frac{B^2}{kt} = tG^2$$