GDI **SUMMARY** of GRADIENT DESCENT COMPLEXITY Dec 6, 2016

Note the difference!

Strict Convexity $f(\theta x+(1-\theta)y) < \theta f(x)+(1-\theta)f(y)$

Strong Convexity $\exists m > 0$ s.t. $\nabla^2 f(x) \succeq mI$ for all $x \in S$.

e^{-x} is strictly convex but not strongly convex.

Big Picture Results for Gradient Descent from last lecture, assuming f strongly convex.

Exact line search At least as good as $t = \frac{1}{M}$,

which gives

$$f(x^{(k)}) - p^* \leq \left(1-\frac{1}{K}\right)^k \left(f(x^{(0)})-p^*\right)$$

where $K = M/m$

$$M = \sup_{x\in S} \lambda_{max}(\nabla^2 f(x))$$

$$m = \inf_{x\in S} \lambda_{min}(\nabla^2 f(x))$$

$\Rightarrow$ # iters for accy ε is $O\left(\frac{1}{\log \varepsilon}\right)$

BUT don't know M, m in practice.

Hence: BACKTRACKING LINE SEARCH

$\alpha \in (0, 1/2)$ ARMIJO COND.

$\beta \in (0, 1)$ BACKTRACK PARAM.

When $\beta = \frac{1}{2}$, $M \geq 1/2$, get

$$f(x^{(k)}) - p^* \leq \left(1-\frac{\alpha}{K}\right)^k \left(f(x^{(0)})-p^*\right).$$

GD2

Gradient Descent, cont'd.

Nesterov (Thm 2.1.15) also gives another complexity result for

$$t = \frac{2}{m+M}$$

namely

$$\|x^{(k)} - x^*\| \le \left(\frac{1-1/K}{1+1/K}\right)^k \|x^{(0)} - x^*\|$$

which leads to

$$f^{(k)} - p^* \le K \cdot \left(\frac{1-1/K}{1+1/K}\right)^{2k} \left(f^{(0)} - p^*\right)$$

uses Thm 2.1.8.

where x^* = minimizer, $p^* = f(x^*)$.

(WORSE for small k if K large, better as $k \to \infty$)

Both BV and Nesterov also give results for the non-(strongly convex) case – these are much weaker.