Anatomy of a convergence rate proof.

- start with some quadratic upper or lower bd on $f(y)$ around $f(x)$

  $\uparrow$ current iterate

  Lipschitz grad:
  upper bound on $f(y)$

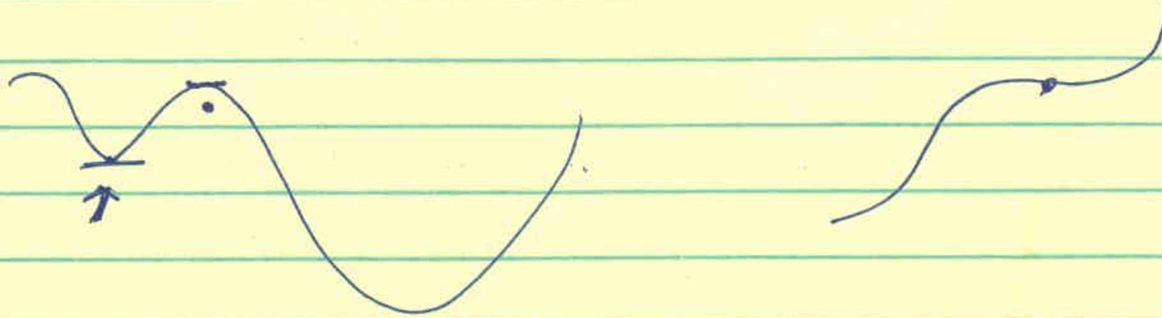
  strong convexity:
  lower bound on $f(y)$

- establish some "sufficient descent" property at $f(x^+)$

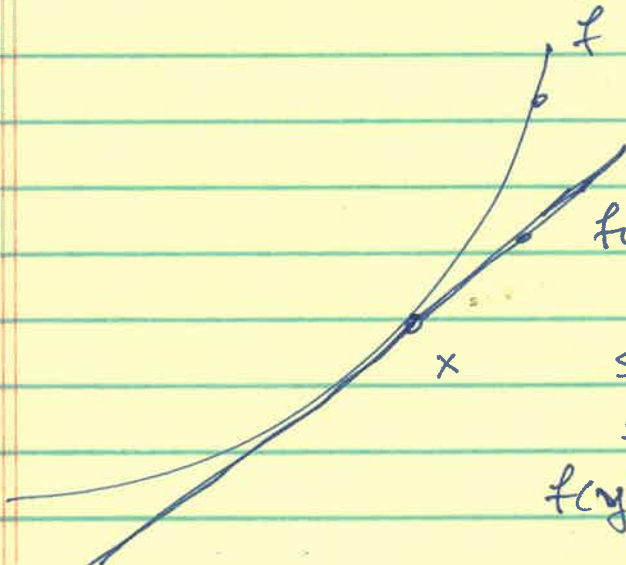
  $\uparrow$ next iterate

eg. with strong descent	eg. in grad descent with $t \le 1/L$
$\|x^+ - x^*\|_2^2 \le$	$f(x^+) \le f(x) - \frac{t}{2}\|\nabla f(x)\|_2^2$

- iterate / recurse this property to get a global statement about $f(x^{(k)})$ or $x^{(k)}$

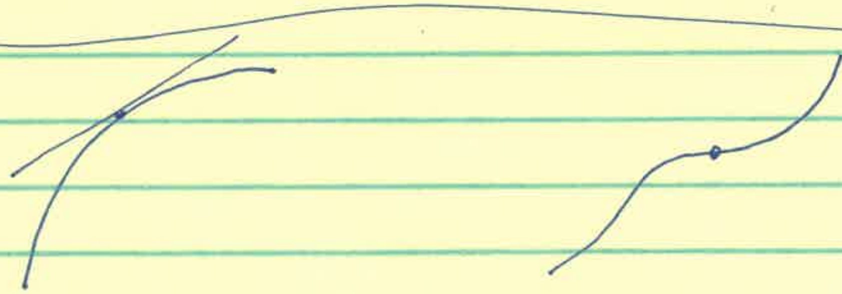


Gradient Boosting : do grad descent
on some loss function
(eg. classification or regression)
where we map gradients to vector of
predictions by
a <u>tree</u>



$f(x) + \nabla f(x)^T (y-x)$

subgradient of f at x is g
st
$f(y) \geq f(x) + g^T (y-x)$ all y.



$f(x) = x^3$

$f(y) \geq f(x) + g^T(y-x)$

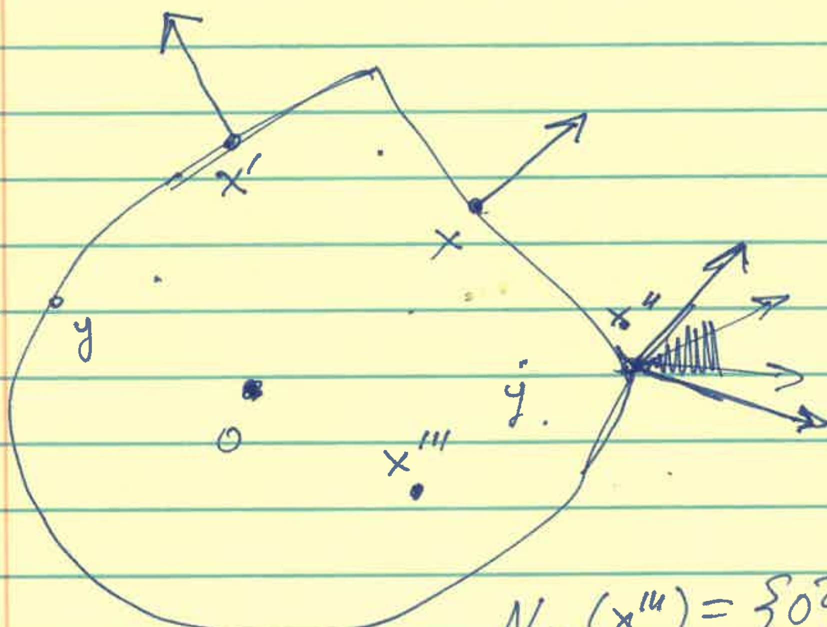
$\|y\|_2 \geq g^T y.$

any g : $\|g\|_2 \leq 1$ works.

$I_C(y) \geq I_C(x) + g^T(y-x)$

$I_C(y) \geq g^T(y-x)$

- $y \in C$ get $0 \geq g^T(y-x)$
  i.e. $g^T x \geq g^T y$
- $y \notin C$ get $\infty \geq g^T(y-x)$ ✓.



$N_C(x''') = \{0\}$

---

x s.t. $f_1(x) = f_2(x)$

$\operatorname{conv}\left(\bigcup_{i:\, f_i(x) = f(x)} \{\nabla_i f(x)\} \right)$ = line segment joining $\nabla_1 f(x) + \nabla_2 f(x)$

$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p}$$

$p \geq 1$. $\quad q$ s.t. $\frac{1}{q} + \frac{1}{p} = 1$.

fact. $\|x\|_p = \max_{y : \|y\|_q \leq 1} y^T x$.

$\quad$ " $f_y(x)$

$S = \{ y : \|y\|_q \leq 1 \}$

$\beta_i = S_\lambda(y_i)$

check: $\quad y_i - \beta_i = \lambda \, \mathrm{sign}(\beta_i) \quad$ if $\beta_i \neq 0$.

$\quad |y_i - \beta_i| \leq \lambda \quad$ if $\beta_i = 0$

- $y_i > \lambda$. $\quad \beta_i = y_i - \lambda > 0$.

  $y_i - \beta_i = \lambda = \lambda \cdot \mathrm{sign}(\beta_i)$ ✓

- $y_i < -\lambda$. $\quad$ similar

- $y_i \in [-\lambda, \lambda]$. $\quad \beta_i = 0$.

  $|y_i| \leq \lambda$ ✓