* <u>Other basic descent methods:</u>

There are other descent methods for which you can guarantee:

$$f(x_{k+1}) \le f(x_k) - \frac{\beta}{2} \|\nabla f(x_k)\|_2^2 \quad \text{for some } \beta > 0.$$

* Examples:

1) Preconditioned methods:

$\underline{x_{k+1} = x_k - \alpha S_k \nabla f(x_k)}$, where S_k is a PD matrix w/ eigenvalues in $[\mu_1, \mu_2]$

$0 < \mu_1 < \mu_2 < \infty$.

From Lemma 2.2:

$$f(x_{k+1}) \le f(x_k) + \langle \nabla f(x_k), x_{k+1} - x_k \rangle + \frac{L}{2} \|x_{k+1} - x_k\|_2^2$$

$$= f(x_k) - \alpha \underbrace{\langle S_k \nabla f(x_k), \nabla f(x_k) \rangle}_{\ge \mu_1 \|\nabla f(x_k)\|_2^2}$$

$$+ \frac{L}{2} \alpha^2 \underbrace{\|S_k \nabla f(x_k)\|_2^2}_{\le \mu_2^2 \|\nabla f(x_k)\|_2^2}$$

$$\le f(x_k) - (\underbrace{\alpha \mu_1 - \frac{L}{2} \mu_2^2 \alpha^2}_{>0 \text{ for suff. small } \alpha}) \|\nabla f(x_k)\|_2^2.$$

Newton's method uses $S_k = (\nabla^2 f(x_k))^{-1}$; need $\nabla^2 f(x_k)$ to have positive evals for this work

2) Gauss-Southwell (greedy coordinate descent)

$$x_{k+1} = x_k - \alpha \underbrace{\nabla_{i_k} f(x_k) e_{i_k}}_{-p_k}, \quad e_{i_k} = [0, 0, \dots, \underbrace{1}_{i_k \text{ position}}, \dots, 0]$$

* $$\boxed{i_k = \arg\max_{1 \le i \le n} |\nabla_i f(x_k)|}$$

$$\|p_k\|_2 \geq \frac{1}{d} \|\nabla f(x_k)\|_2$$

3) Randomized coordinate descent (HW #2)

4) Stochastic gradient descent, where

$$p_k = -g(x_k, \xi_k) \quad , \quad \mathbb{E}_{\xi_k}[g(x_k, \xi_k)] = \nabla f(x_k)$$

$\xi_k \downarrow$ i.i.d. r.v.

$$x_{k+1} = x_k + \alpha p_k ,$$

under certain assumptions.

* Convergence of basic descent methods:

* Assume:

$$f(x_{k+1}) \leq f(x_k) - \frac{\alpha}{2} \|\nabla f(x_k)\|_2^2 .$$

1 nonconvex case:

$$\min_{0 \leq i \leq k} \|\nabla f(x_i)\|_2 \leq \sqrt{\frac{2(f(x_0) - f_*)}{\alpha(k+1)}} ,$$

where $f(x) \geq f_* > -\infty$, $\forall x$.

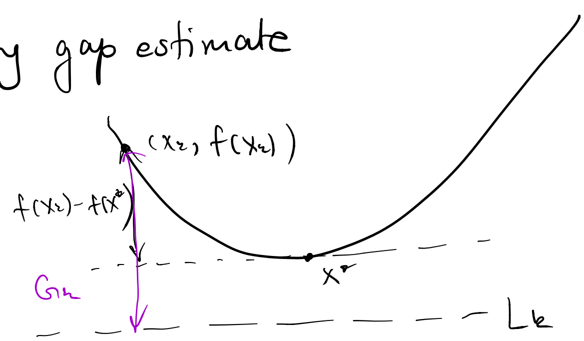
2 Convex f.

$$\boxed{\begin{array}{l}\text{convexity:} \\ \forall x: \quad f(x^*) \geq f(x) + \langle \nabla f(x), x^* - x \rangle\end{array}}$$

$f(x_{k+1}) - f(x^*) \leq G_k$ optimality gap estimate

$f(x^*) \geq L_k$ lower bound

Let A_k be a strictly increasing positive numbers.



Goal:

$$A_k G_k - A_{k-1} G_{k-1} \le \boxed{E_k} \quad \text{"error"}$$

$$\Rightarrow \quad A_k G_k \le A_0 G_0 + \sum_{i=1}^{k} E_i$$

$$f(x_{k+1}) - f(x^*) \le G_k \le \underbrace{\frac{\overbrace{A_0 G_0}^{\text{bounded}}}{A_k} + \frac{\sum_{i=1}^{k} E_i}{A_k}}_{} \quad *$$

want it to grow slowly compared to A_k

Lower bound on $f(x^*)$:

$\forall i$:

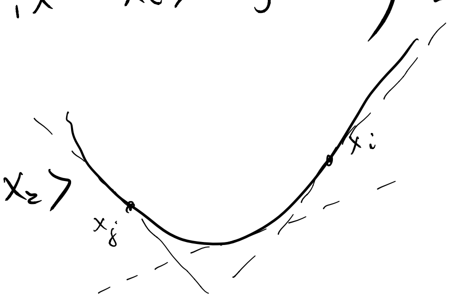
$$f(x^*) \ge f(x_i) + \langle \nabla f(x_i), x^* - x_i \rangle$$

Let $\{a_i\}_{i \ge 0}$ be a sequence of positive numbers s.t. $A_k = \sum_{i=0}^{k} a_i$ so that $\frac{1}{A_k} \sum_{i=0}^{k} a_i = 1$.

$$\underbrace{\sum_{i=0}^{k} a_i}_{A_k} f(x^*) \ge \sum_{i=0}^{k} a_i \Big(\langle \nabla f(x_i), x^* - x_i \rangle + f(x_i) \Big)$$

$$f(x^*) \ge \frac{1}{A_k} \sum_{i=0}^{k} a_i \Big(\langle \nabla f(x_i), x^* - x_i \rangle + f(x_i) \Big) =: L_k$$

$$A_k L_k - A_{k-1} L_{k-1} = a_k f(x_k) + a_k \langle \nabla f(x_k), x^* - x_k \rangle$$



$$G_k = f(x_{k+1}) - L_k$$

$$A_k G_k - A_{k-1} G_{k-1} = A_k f(x_{k+1}) - A_{k-1} f(x_k) - a_k f(x_k) - a_k \langle \nabla f(x_k), x^* - x_k \rangle$$

$$\overset{A_k = A_{k-1} + a_k}{=} A_k \big(f(x_{k+1}) - f(x_k) \big) - a_k \langle \nabla f(x_k), x^* - x_k \rangle$$

"Descent Lemma"

$$\le -\frac{A_k \alpha}{2}\|\nabla f(x_k)\|_2^2$$

$$\underbrace{- a_k \langle \nabla f(x_k), x^* - x_k \rangle}$$

C.S.

$$\le -\frac{A_k \alpha}{2}\|\nabla f(x_k)\|_2^2 + a_k \|\nabla f(x_k)\|_2 \cdot \|x^* - x_k\|_2$$

Useful inequality: $-\frac{p^2}{2} + pq \le \frac{q^2}{2}$

$p = \sqrt{\alpha A_k}\,\|\nabla f(x_k)\|_2$, $q = \frac{a_k}{\sqrt{\alpha A_k}}\|x^* - x_k\|_2$

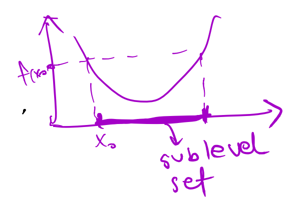
$$A_k G_k - A_{k-1} G_{k-1} \le \frac{a_k^2}{2\alpha A_k}\|x^* - x_k\|_2^2 = E_k$$

Ex. $A_0 G_0 \le \frac{a_0^2}{2\alpha A_0}\|x^* - x_0\|_2^2$

$$\Rightarrow f(x_{k+1}) - f(x^*) \le G_k \le \frac{1}{A_k}\sum_{i=0}^{k} \frac{a_i^2}{2\alpha A_i}\|x^* - x_k\|_2^2$$

Define $R = \max\{\|x^* - x\|_2 : f(x) \le f(x_0)\}$

sublevel set: $\{x : f(x) \le f(x_0)\}$



$$\Rightarrow f(x_{k+1}) - f(x^*) \le \frac{R^2}{2\alpha A_k}\sum_{i=0}^{k}\frac{a_i^2}{A_i}$$

How should we choose a_i ($A_i = \sum_{j=0}^{i} a_j$)?

$a_i \propto i^p$, $p > 0$, for k large enough $A_k \propto \frac{k^{p+1}}{p+1}$.

(for integer p, Faulhaber's formula)

One particular choice

$a_i = \frac{i+1}{2}$; $A_i = \frac{(i+1)(i+2)}{4}$

$\frac{a_i^2}{A_i} \le 1$

$$\Rightarrow \quad f(x_{k+1}) - f(x^*) \leq \frac{2R^2}{\alpha(k+2)} .$$