Feb 4, 2020

Last time: gradient descent $x_{k+1} = x_k - \alpha_k \underbrace{\nabla f(x_k)}_{g_k}$

error analysis: $e_k = x_k - x_*$ $\left(f(x) = \frac{1}{2}x^T A x + x^T b + c\right)$

$\|e_{k+1}\| < a\|e_k\|$ $a < 1$ $\|e_{k+1}\| \leq O(a^k)$

Newton: $x_{k+1} = x_k - H_k^{-1} g_k$

Stochastic gradient descent (SGD) $x_{k+1} = x_k - \alpha_k g_{J_k}$

$f(x) = \frac{1}{N}\sum_{i=1}^{N} f_i(x)$ $g(x) = \frac{1}{N}\sum_{i=1}^{N} g_i(x)$

$J \subseteq \{1, \dots, N\}$ $g_{J_k} = \frac{1}{|J|}\sum_{j \in J} g_j(x)$ $\mathbb{E}(g_{J_k}) = g_k$

Should we expect SGD to converge?

$$x_{k+1} = x_k - \alpha_k (g_k + u_k)$$ ← error/noise

$$f(x) = \frac{1}{2} x^T A x + x^T b + c \qquad g(x) = Ax + b$$

$$b = -Ax^*$$

$$x_{k+1} - x_* = x_k - x_* - \alpha (A x_k + b + u_k)$$

$$e_{k+1} = e_k - \alpha (A(x_k - x_*)) - \alpha u_k$$

$$=$$

$$e_1 = (I - \alpha A) e_0 - \alpha u_0$$

$$e_2 = \qquad e_0$$

control as before

$$A = \begin{bmatrix} a_1^T \\ \vdots \\ a_m^T \end{bmatrix} \qquad {}_m\overset{k}{\begin{bmatrix} L \end{bmatrix}} \qquad L_{ij} = \begin{cases} 1 & \text{if point } i \text{ in } C_j \\ 0 & \text{o/w} \end{cases}$$

$$R = {}_k\overset{n}{\begin{bmatrix} r_1^T \\ \vdots \\ r_k^T \end{bmatrix}} \qquad A \approx LR$$

$$e_i^T A = a_i^T \approx e_i^T (LR) = (e_i^T L) R = r_j^T \quad j = C(i)$$

$[0 \cdots 0\ 1\ 0 \cdots 0]$

$$\|A - LR\|_F^2$$

Alternatin min

(2)

(1) Fix L

$$R = \min_S \|A - LS\|_F$$

R = means