SGD Ex

July 11, 2022

P4

E1

Def. 17:

$$D_f(x,y) + D_f(y,x) = \langle \nabla f(x) - \nabla f(y), x - y \rangle = \langle \nabla f(y) - \nabla f(x), y - x \rangle \tag{1}$$

$\forall x, y \in \mathbb{R}^d$:

$$\begin{aligned} \mu ||x-y||^2 &\leq 2D_f(x,y), \\ \frac{\mu}{2}||x-y||^2 &\leq D_f(x,y), \\ \frac{\mu}{2}||x-y||^2 &\leq D_f(y,x), \\ D_f(x,y) + \frac{\mu}{2}||x-y||^2 &\leq D_f(x,y) + D_f(y,x), \\ D_f(x,y) + \frac{\mu}{2}||x-y||^2 &\overset{(1)}{\leq} \langle \nabla f(x) - \nabla f(y), x - y \rangle. \end{aligned} \tag{2}$$

E2

$$\begin{aligned} D_f(x,y) + \frac{\mu}{2}||x-y||^2 &\leq \langle \nabla f(x) - \nabla f(y), x - y \rangle, \\ \langle \nabla f(x) - \nabla f(y), x - y \rangle &\geq \underbrace{D_f(x,y)}_{\geq \frac{\mu}{2}||x-y||^2} + \frac{\mu}{2}||x-y||^2, \\ \langle \nabla f(x) - \nabla f(y), x - y \rangle &\geq \frac{\mu}{2}||x-y||^2 + \frac{\mu}{2}||x-y||^2, \\ \langle \nabla f(x) - \nabla f(y), x - y \rangle &\geq \mu||x-y||^2. \end{aligned} \tag{3}$$

P6

E17

(Equation 34):

$$\begin{aligned}\langle a,b\rangle &\le \frac{||a||^2}{2t}+\frac{t||b||^2}{2},\\ \langle a,b\rangle &\le \frac{\langle a,a\rangle}{2t}+\frac{t\langle b,b\rangle}{2},\\ 2t\langle a,b\rangle &\le \langle a,a\rangle + t^2\langle b,b\rangle,\\ 0 &\le \langle a,a\rangle+\langle tb,tb\rangle-\langle a,tb\rangle-\langle tb,a\rangle,\\ 0 &\le ||a-tb||^2.\end{aligned} \tag{4}$$

(Equation 35):

$$\begin{aligned}||a+b||^2 \le 2||a||^2+2||b||^2,&\\ \langle a,a\rangle+\langle b,b\rangle+2\langle a,b\rangle \le 2\langle a,a\rangle+2\langle b,b\rangle,&\\ 0 \le \langle a,a\rangle+\langle b,b\rangle-2\langle a,b\rangle,&\\ 0\le ||a-b||^2.&\end{aligned} \tag{5}$$

(Equation 36):

$$\begin{aligned}\frac{1}{2}||a||^2-||b||^2 \le ||a+b||^2,&\\ \frac{1}{2}\langle a,a\rangle-\langle a,a\rangle \le \langle a,a\rangle+\langle b,b\rangle+2\langle a,b\rangle,&\\ \langle a,a\rangle-2\langle b,b\rangle \le 2\langle a,a\rangle+2\langle b,b\rangle+4\langle a,b\rangle,&\\ 0\le \langle a,a\rangle+\langle 2b,2b\rangle+\langle a,2b\rangle+\langle 2b,a\rangle,&\\ 0\le ||a+2b||^2.&\end{aligned} \tag{6}$$

E19

For random vector $X \in \mathbb{R}^d$:

$$\mathbf{Var}[X] := \mathbf{E}\left[||X-\mathbf{E}[X]||^2\right]. \tag{7}$$

Markov's inequality:

$$\mathrm{Prob}(X \ge t) \le \frac{\mathbf{E}[X]}{t}. \tag{8}$$

Proof of Chebyshev's inequality using Markov's inequality:

$$\mathrm{Prob}(||X-\mathbf{E}[X]||^2 \ge t^2) \le \frac{\mathbf{E}\left[||X-\mathbf{E}[X]||^2\right]}{t^2}.$$

Since

$$\mathrm{Prob}(||X-\mathbf{E}[X]||^2 \ge t^2) = \mathrm{Prob}(||X-\mathbf{E}[X]|| \ge t), \tag{9}$$

then

$$\mathrm{Prob}(||X-\mathbf{E}[X]|| \ge t) \le \frac{\mathbf{Var}[X]}{t^2}. \tag{10}$$

P7

E24

If

$$f = \frac{1}{n}\sum_{i=1}^{n} f_i,$$

then

$$\begin{aligned}
D_f(x,y) &= \frac{1}{n}\sum_{i=1}^{n} f_i(x) - \frac{1}{n}\sum_{i=1}^{n} f_i(y) - \frac{1}{n}\sum_{i=1}^{n}\langle \nabla f_i(y), x-y\rangle,\\
D_f(x,y) &= \frac{1}{n}\sum_{i}(f_i(x) - f_i(y) - \langle \nabla f_i(y), x-y\rangle),\\
D_f(x,y) &= \frac{1}{n}\sum_{i} D_{f_i}(x,y).
\end{aligned}$$

(maybe) this is the other direction, if matrix is psd then E is psd:

P9

P10

E39) no comp: $C(x) = x$

$\cdot\ \omega_i = \omega\ \forall i$.

Frommg Pg 267 $\xrightarrow{\text{implies}}$ $g(x) = C(\bar{g}(x)) = \bar{g}(x) = \frac{1}{n}\sum_{i=1}^{n} g_i(x)$

$g_i(x) = C_i(\nabla f_i(x))$

. Bound Fen $E[\|g(x) - \nabla F(x)\|^2]$:

$\ldots = E[E[\|\underbrace{C(\bar{g}(x)) - \bar{g}(x)}_{0} + \bar{g}(x) - \nabla F(x)\|^2 \mid C_i \ldots C_n]]$

= like in lect. note

$=$ or $\leq$? $E[\|\bar{g}(x) - \nabla F(x)\|^2] + 2L D_F$ (113)

. Bound (37.1) $- 2L D_F$ (39.1)

$= E[\|\frac{1}{n}\sum_{i=1}^{n}(g_i(x) - \nabla f_i(x))\|^2]$

$= E[\|\frac{1}{n}\sum_i a_i\|^2]$

(like in lect. note) (114)

$= \frac{1}{n^2}\sum_i E[\|a_i\|^2] + 0$

(107) $\leq \frac{1}{n^2}\sum_i \omega_i \|\nabla f_i(x)\|^2$

since $\omega_i = \omega\ \forall i$:

$\leq \frac{\omega}{n^2}\sum_i \|\nabla f_i(x)\|^2$ (39.2)

. Bound $\|\nabla f_i(x)\|^2$:

(116) $\leq 4L_i D_{f_i}(x,y) + 2\|\nabla f_i(y)\|^2$ (39.3)

. $G(x,y) \leq E[\|g(x) - \nabla F(x)\|] + 2L D_F$

$\overset{(39.2)}{\leq} \frac{\omega}{n^2}\sum_i \|\nabla f_i(x)\|^2 + 2L D_F$

$\overset{(39.3)}{\leq} \frac{\omega}{n^2}\sum_i 4L_i D_{f_i}(x,y) + 2\|\nabla f_i(y)\|^2 + 2L D_F$

$= 2\frac{\omega}{n}\left(\sum_i \frac{2L_i D_{f_i}}{n} + \sum_i \frac{\|\nabla f_i(y)\|^2}{n}\right) + 2L D_F$

$\overset{\max_i L_i}{\leq} \frac{\omega}{n} 4 L_{max} D_F + 2L D_F + 2\frac{\omega}{n}\sigma^2(y)$

rewrite $= 2(L + 2L_{max}\frac{\omega}{n}) D_F + 2\frac{\omega}{n}\sigma^2(y)$

For $\begin{cases} \sigma^2(y) = \frac{1}{n}\sum_i \|\nabla f_i(y)\|^2 = 0 \quad (a) \\ \omega_i = \omega \quad (b) \end{cases}$

when $\nabla F(y) = 0$. (c)

$E[\|g(x) - \nabla F(y)\|] = E[\|g(x)\|] = E[\|\bar{g}(x)\|]$

$\cdot\ E[\|\bar{g}(x)\|]$

(115, 116) $\leq \frac{1}{n^2}\left[\sum_i 4\omega L_i D_{f_i} + 2\omega\|\nabla f_i(y)\|^2\right] + 2L D_F + 2\|\nabla F(y)\|^2$

$= \frac{2\omega}{n}\sum_i \frac{2L_i D_{f_i}}{n} + \sum_i \underbrace{\frac{\|\nabla f_i(y)\|^2}{n}}_{0} + 2L D_F + 2\underbrace{\|\nabla F(y)\|^2}_{0}$

(d)(c) $= \frac{2\omega}{n}\sum_i \frac{2L_i D_{f_i}}{n} + 2L D_F$

take max to u.bound sum:

$\leq \frac{4\omega L_{max}}{n} D_F + 2L D_F$

$= 2\left(\frac{2\omega L_{max}}{n} + L\right) D_F$

rewrite: $= 2\left(L + 2L_{max}\frac{\omega}{n}\right) D_F(x,y)$

E37) trivial;

follow (108) - proof of lemma 2, w/ extra term δ

until) Convergence of CGD

$E[\|x^k - x^*\|^2] \leq (1 - \gamma\mu)^k \|x^0 - x^*\|^2 + \frac{\gamma C}{\mu}$

where $C = 2\omega\|\nabla F(y)\|^2 + \delta$