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):

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

(Equation 35):

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

(Equation 36):

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

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:

$$\text{Prob}(X \geq t) \leq \frac{\mathbf{E}[X]}{t}. \tag{8}$$

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

$$\text{Prob}(||X - \mathbf{E}[X]||^2 \geq t^2) \leq \frac{\mathbf{E}\left[||X - \mathbf{E}[X]||^2\right]}{t^2}.$$

Since

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

then

$$\text{Prob}(||X - \mathbf{E}[X]|| \geq t) \leq \frac{\mathbf{Var}[X]}{t^2}. \tag{10}$$

[P7]

E24

If

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

then

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

E26

[*** partial ***]

[P8]

E33

Let

$$\chi_i = \begin{cases} 1 & i \in S \\ 0 & i \notin S \end{cases}.$$

Since

$$p_i = \frac{1}{n},$$

and

$$|S| = \tau,$$

then

$$\mathbf{E}[\chi_i] = \mathrm{Prob}(i \in S) = \sum_{i=1}^{n} p_i \chi_i = \frac{1}{n}\sum_{i=1}^{n}\chi_i = \frac{\tau}{n}.$$

E35

For any vectors, $b_1, ..., b_n \in \mathbb{R}^d$:

$$\begin{aligned}
\left\|\sum_{i=1}^{n} b_i\right\|^2 - \sum_{i=1}^{n}\|b_i\|^2 &= \underbrace{\sum_{i=1}^{n}\langle b_i, b_i\rangle + \sum_{i\neq j}\langle b_i, b_j\rangle}_{\left\|\sum_{i=1}^{n} b_i\right\|^2} - \sum_{i=1}^{n}\langle b_i, b_i\rangle,\\
\left\|\sum_{i=1}^{n} b_i\right\|^2 - \sum_{i=1}^{n}\|b_i\|^2 &= \sum_{i\neq j}\langle b_i, b_j\rangle.
\end{aligned}$$

[P9]

E37

Assumptions of $\mathcal{C} : \mathbb{R}^d \to \mathbb{R}^d$:

1. $\mathbf{E}[\mathcal{C}(x)] = x, \quad \forall x \in \mathbb{R}^d$
2. $\mathbf{E}\left[||\mathcal{C}(x) - x||^2\right] \leq \omega ||x||^2 + \delta, \quad \forall x \in \mathbb{R}^d, \quad \exists \omega, \delta \geq 0$

Proof of convergence for CGD with $n = 1$:
Since $\mathcal{C} \in \mathbb{B}^d(\omega)$,

$$\mathbf{E}\left[||g(x)||^2\right] = \mathbf{E}\left[||\mathcal{C}(\nabla f(x))||^2\right] \leq (\omega + 1)||\nabla f(x)||^2. \tag{11}$$

In case of $\nabla f(y) = 0$,

$$\begin{aligned} G(x,y) &:= \mathbf{E}\left[||g(x) - \nabla f(y)||^2\right] \\ &= \mathbf{E}\left[||g(x)||^2\right] \\ &\overset{(11)}{\leq} (\omega + 1)||\nabla f(x) - \nabla f(y)||^2, \\ &\leq 2(\omega + 1) L D_f(x, y). \end{aligned}$$

In case of $\nabla f(y) \neq 0$,

$$\begin{aligned} G(x,y) &:= \mathbf{E}\left[||g(x) - \nabla f(y)||^2\right] \\ &= \mathbf{E}\left[||g(x) - \nabla f(x)||^2\right] + ||\nabla f(x) - \nabla f(y)||^2 \\ &= \mathbf{E}\left[||\mathcal{C}(\nabla f(x)) - \nabla f(x)||^2\right] + ||\nabla f(x) - \nabla f(y)||^2 \\ &\leq \omega ||\nabla f(x)||^2 + \delta + ||\nabla f(x) - \nabla f(y)||^2 \\ &= \omega ||\nabla f(x) - \nabla f(y) + \nabla f(y)||^2 + ||\nabla f(x) - \nabla f(y)||^2 + \delta \\ &\leq 2\omega ||\nabla f(x) - \nabla f(y)||^2 + 2\omega ||\nabla f(y)||^2 + ||\nabla f(x) - \nabla f(y)||^2 + \delta \\ &= (2\omega + 1)||\nabla f(x) - \nabla f(y)||^2 + 2\omega ||\nabla f(y)||^2 + \delta \\ &\leq 2 \underbrace{(2\omega + 1) L}_{A} D_f(x, y) + \underbrace{2\omega ||\nabla f(y)||^2 + \delta}_{C}. \end{aligned}$$

If $0 < \gamma < \frac{1}{A}$, then

$$\mathbf{E}\left[||x^k - x^*||^2\right] \leq (1 - \gamma\mu)^k ||x^0 - x^*|| + \frac{2\gamma\omega ||\nabla f(x^*)||^2 + \gamma\delta}{\mu}.$$

E39

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

P9

P10

E39) no comp: $\mathcal{C}(x) = x$

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

From pg 267 $\xrightarrow{\text{implies}}$ $g(x) = \mathcal{C}(\bar{g}(x))$

$= \bar{g}(x)$

$= \frac{1}{n}\sum_{i=1}^{n} g_i(x)$

$g_i(x) = \mathcal{C}_i(\nabla f_i(x))$

• Bound for $E[\|g(x) - \nabla f(x)\|^2]$:

$\ldots = E[E[\|\underbrace{\mathcal{C}(\bar{g}(x)) - \bar{g}(x)}_{0} + \bar{g}(x) - \nabla F(x)\|^2 \mid \mathcal{C}_1 .. \mathcal{C}_n]]$

= like in lect. notes (113)

$= \text{or} \le E[\|\bar{g}(x) - \nabla F(x)\|^2] + 2LD_F$ (39.1)

?• Bound (37.1) $- 2LD_F$

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

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

- (like in lect. notes) (114)

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

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

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

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

• Bound $\|\nabla f_i(x)\|^2$:

(116)

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

• $G(x,y) \le E[\|g(x) - \nabla F(x)\|] + 2LD_F$

$\overset{(39.2)}{\le} \frac{\omega}{n^2}\sum_i \|\nabla f_i(x)\|^2 + 2LD_F$

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

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

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

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

For $\begin{cases} \sigma^2(y) = \frac{1}{n}\sum_i \|\nabla f_i(y)\|^2 = 0 \quad (a) \\ \qquad \text{when } \nabla_{f(y)} = 0. \quad (c) \\ \omega_i = \omega \quad (b) \end{cases}$: $E[\|g(x) - \nabla_{f(y)}\|] = E[\|g(x)\|]$

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

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

$+ 2LD_F + 2\|\nabla f(y)\|^2$

$\overset{(a)(c)}{=} \frac{2\omega}{n}\sum_i \frac{2L_i D_{f_i}}{n} + \underbrace{\sum_i \frac{\|\nabla f_i(y)\|^2}{n}}_{0} + 2LD_F + \underbrace{2\|\nabla_{f(y)}\|^2}_{0}$

$= \frac{2\omega}{n}\sum_i \frac{2L_i D_{f_i}}{n} + 2LD_F$

take max to u.bound sum:

$\le 4\frac{\omega L_{max}}{n} D_F + 2LD_F$

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

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

E37) trivial;

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

until) Convergence of CGD

$E[\|x^k - x^*\|^2] \le (1-\gamma\mu)^k \|x^0 - x^*\|^2$

$+ \frac{\gamma C}{\mu}$

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