Assignment 1
Large-scale Optimization FTN0452

Li Ju

September 16, 2023

1 Problem 2.8

a).

Convexity

Proof. Given that f is m-strongly convex, by definition we have

$$\begin{aligned}
f(y) - f(x) &\geq \frac{m}{2}\|y-x\|^2 + (y-x)^\top \nabla f(x) \\
&= \frac{m}{2}(y-x)^\top(y-x) + (y-x)^\top(\nabla f(x) - mx) + m(y-x)^\top x \\
&= \frac{m}{2}(y-x)^\top(y-x+2x) + (y-x)^\top(\nabla f(x) - mx) \\
&= \frac{m}{2}(\|y\|^2 - \|x\|^2) + (y-x)^\top(\nabla f(x) - mx) \\
f(y) - \frac{m}{2}\|y\|^2 &\geq f(x) - \frac{m}{2}\|x\|^2 + (y-x)^\top(\nabla f(x) - mx)
\end{aligned}$$

With $q(x) \coloneqq f(x) - \frac{m}{2}\|x\|^2$ and $\nabla q(x) = \nabla f(x) - mx$, we have $q(y) \geq q(x) + (y-x)^\top \nabla q(x)$.
Let $z = (1-\alpha)x + \alpha y, \alpha \in [0,1]$, we have

$$\begin{aligned}
q(x) &\geq q(z) + (x-z)^\top \nabla q(z) \\
&= q(z) + \alpha(x-y)^\top \nabla q(z) \qquad (1) \\
q(y) &\geq q(z) + (y-z)^\top \nabla q(z) \\
&= q(z) + (1-\alpha)(y-x)^\top \nabla q(z) \qquad (2)
\end{aligned}$$

Summing up $(1-\alpha)\cdot(1) + \alpha\cdot(2)$, we have

$$\begin{aligned}
(1-\alpha)q(x) + \alpha q(y) &\geq q(z) + (1-\alpha)\alpha(y-x)\nabla q(z) + (1-\alpha)\alpha(x-y)\nabla q(z) \\
&= q((1-\alpha)x + y)
\end{aligned}$$

□

Smoothness

Proof. Given that ∇f is L-continuous and f is m-strongly convex, we have

$$
\begin{aligned}
f(y) &\le f(x) + (y-x)^\top \nabla f(x) + \frac{L}{2}\|y-x\|^2 \\
f(y) &\le f(x) + (y-x)^\top \nabla f(x) + \frac{L}{2}\|y-x\|^2 - \frac{m}{2}\|y-x\|^2 + \frac{m}{2}\|y-x\|^2 \\
f(y) - \frac{m}{2}\|y\|^2 &\le f(x) - \frac{m}{2}\|x\|^2 + (y-x)^\top(\nabla f(x) - mx) + \frac{L-m}{2}\|y-x\|^2
\end{aligned}
$$

With $q(x) \coloneqq f(x) - \frac{m}{2}\|x\|^2$ and $\nabla q(x) = \nabla f(x) + mx$, we have

$$
q(y) \le q(x) + (y-x)^\top \nabla q(x) + \frac{L-m}{2}\|y-x\|^2 \tag{3}
$$

Since we have proved the convexity of q, from Lemma **??**, we have co-coercivity

$$
\begin{aligned}
\frac{1}{L-m}\|\nabla q(y) - \nabla q(x)\|^2 &\le (y-x)^\top(\nabla q(y) - \nabla q(x)) \\
\|\nabla q(y) - \nabla q(x)\|^2 &\le (L-m)(y-x)^\top(\nabla q(y) - \nabla q(x)) \\
&\le (L-m)\|y-x\|^\top \|\nabla q(y) - \nabla q(x)\| \\
\|\nabla q(y) - \nabla q(x)\| &\le (L-m)\|y-x\|
\end{aligned}
$$

Lemma 1. *hello world*

□

b).

By applying *co-cocervicity* to $q(x)$, which has $(L-m)$-continuous gradients, we have

$$
\begin{aligned}
[\nabla q(x) - \nabla q(y)]^\top (x-y) &\ge \frac{1}{L-m}\|\nabla q(x) - \nabla q(y)\|^2 \\
[\nabla f(x) - \nabla f(y)]^\top (x-y) &\ge \frac{1}{L-m}\|\nabla f(x) - \nabla f(y) - m(x-y)\|^2 + m\|x-y\|^2 \\
&= \frac{\|\nabla f(x) - \nabla f(y)\|^2 - 2m(\nabla f(x) - \nabla f(y))^\top (x-y) + mL\|x-y\|^2}{L-m} \\
&= \frac{(L-m)\|\nabla f(x) - \nabla f(y)\|^2 + 2m\|\nabla f(x) - \nabla f(y)\|^2}{(L-m)(L+m)} + \\
&\quad \frac{-2m(L+m)[\nabla f(x) - \nabla f(y)]^\top (x-y) + mL(L+m)\|x-y\|^2}{(L-m)(L+m)} \\
&\ge \frac{-2m(L+m)\|\nabla f(x) - \nabla f(y)\|\|x-y\| + mL(L+m)\|x-y\|^2}{(L-m)(L+m)} + \qquad (4) \\
&\quad \frac{2m\|\nabla f(x) - \nabla f(y)\|^2}{(L-m)(L+m)} + \frac{\|\nabla f(x) - \nabla f(y)\|^2}{L+m} \\
&\ge \frac{-2m^2(L+m)\|x-y\|^2 + mL(L+m)\|x-y\|^2}{(L-m)(L+m)} + \frac{2m^3\|x-y\|^2}{(L-m)(L+m)} + \qquad (5) \\
&\quad \frac{\|\nabla f(x) - \nabla f(y)\|^2}{L+m} \\
&= \frac{(mL^2 - m^2 L)\|x-y\|^2}{(L-m)(L+m)} + \frac{\|\nabla f(x) - \nabla f(y)\|^2}{L+m} \\
&= \frac{mL}{m+L}\|x-y\|^2 + \frac{1}{m+L}\|\nabla f(x) - \nabla f(y)\|^2
\end{aligned}
$$

In Line (4) we utilize Cauchy-Schwarz inequality ($[\nabla f(x) - \nabla f(y)]^\top (x-y) \le \|\nabla f(x) - \nabla f(y)\| \|x-y\|^2$). In Line (5) we use the m-strongly convexity of function $f(x)$ ($\|\nabla f(x) - \nabla f(y)\| \ge m\|x-y\|$), from which $\|\nabla f(x) - \nabla f(y)\|^2 \ge m^2 \|x-y\|^2$ is also derived and used.

2 Problem 3.5

Proof. With step size $\eta = \frac{2}{m+L}$, for iteration t we have

$$
\begin{aligned}
f(x^{t+1}) &\le f(x^t) - \eta \|\nabla f(x^t)\|^2 + \eta^2 \frac{L}{2} \|\nabla f(x^t)\|^2 \\
&= f(x^t) + \frac{-2(L+m) + 2L}{(L+m)^2} \|\nabla f(x^t)\|^2 \\
&= f(x^t) - \frac{2m}{(L+m)^2} \|\nabla f(x^t)\|^2
\end{aligned}
$$

By subtracting $f^\star \coloneqq f(x^\star)$ from both side, we have

$$
\begin{aligned}
f(x^{t+1}) - f^\star &\le f(x^t) - f^\star - \frac{2m}{(L+m)^2} \|\nabla f(x^t)\|^2 \\
\frac{m}{2} \|x^{t+1} - x^\star\|^2 &\le \frac{L}{2} \|x^t - x^\star\|^2 - \frac{2m}{(L+m)^2} \|\nabla f(x^t)\|^2 \\
\|x^{t+1} - x^\star\|^2 &\le \frac{L}{m} \|x^t - x^\star\|^2 - \frac{4}{(L+m)^2} \|\nabla f(x^t)\|^2 \\
&\le \frac{L(L+m)^2 - 4mL^2}{m(L+m)^2} \|x^t - x^\star\|^2
\end{aligned}
$$

TOBEFILLED—

We have

$$
\|x^{t+1} - x^\star\| \le \frac{L-m}{L+m} \|x^t - x^\star\|
$$

After k iteration with $\kappa \coloneqq \frac{L}{m}$, we have

$$
\|x^k - x^\star\| \le (\frac{\kappa - 1}{\kappa + 1})^k \|x^0 - x^\star\|
$$

$\square$