Q2 BV 9.5. Q3 BV 9.6 Q5.

Recall the backtracking rule:

Given a descent direction Δx for f at $x \in \mathrm{dom}\, f$,
$\alpha \in (0, \frac{1}{2})$, $\beta \in (0,1)$,

$t \leftarrow 1$

decrease of f is not enough.

while $f(x + t\Delta x) > f(x) + \alpha t \nabla f(x)^T \Delta x$
set $t \leftarrow \beta t$.

f: strongly convex with $\underbrace{mI \preccurlyeq \nabla^2 f(x)}_{\text{strongly cvx}} \preccurlyeq MI \quad \forall x.$ M-smooth.

Find backtracking stopping condition:

$$\nabla^2 f(x) \le MI \Rightarrow f(x + t\Delta x) = f(x) + t\nabla f(x)^T \Delta x + \frac{t^2}{2} \Delta x^T \underbrace{\nabla^2 f(z)}_{\le MI} \Delta x$$

z between x & $x + t\Delta x$.

$$f(x+t\Delta x) \le f(x) + t\nabla f(x)^T \Delta x + \frac{Mt^2}{2} \|\Delta x\|^2 .$$

We terminate whenever
$$f(x+t\Delta x) \le f(x) + \alpha t \nabla f(x)^T \Delta x$$

A sufficient condition is:

$$\cancel{f(x)} + t\nabla f(x)^T \Delta x + \frac{Mt^2}{2}\|\Delta x\|^2 \leq \cancel{f(x)} + \alpha t \nabla f(x)^T \Delta x$$

$$\cancel{t}(1-\alpha)\nabla f(x)^T \Delta x + \frac{Mt^{\cancel{2}}}{2}\|\Delta x\|^2 \leq 0 .$$

That is the exit condition holds with $t \in [0, t_0]$

$$1 \geq t_0 = -2\underbrace{(1-\alpha)}_{\uparrow\ 0 \leq \alpha \leq \frac{1}{2}}^{\leq 1} \frac{\nabla f(x)^T \Delta x}{M\|\Delta x\|^2} . \geq \underbrace{-\frac{\nabla f(x)^T \Delta x}{M\|\Delta x\|^2}}_{\geq 0} \neq t_0 .$$

How many iterations s.t. $t \leq t_0$

$$\beta^k \cdot 1 \leq t_0 \Rightarrow k \log \beta \leq \log t_0$$

$$k \geq \frac{\log t_0}{\log \beta} . = \frac{\log (1/t_0)}{\log (1/\beta)}$$

Q3. BV 9.6 $\qquad f(x) = \frac{1}{2}(x_1^2 + \gamma x_2^2)$

$x^* = (x_1^*, x_2^*) = (0,0)$

$$\nabla f(x) = \begin{bmatrix} x_1 \\ \gamma x_2 \end{bmatrix}$$

$$x^{(k+1)} = x^{(k)} - t\nabla f(x^{(k)})$$

$$= \begin{bmatrix} x_1^{(k)} \\ x_2^{(k)} \end{bmatrix} - t \begin{bmatrix} x_1^{(k)} \\ \gamma x_2^{(k)} \end{bmatrix} = \begin{bmatrix} (1-t)x_1^{(k)} \\ (1-t\gamma)x_2^{(k)} \end{bmatrix}.$$

$$x^{(0)} = \begin{bmatrix} \gamma \\ 1 \end{bmatrix}$$

$$x^{(1)} = \begin{bmatrix} (1-t)\gamma \\ (1-t\gamma)\cdot 1 \end{bmatrix}, \quad x^{(2)} = \begin{bmatrix} (1-t)^2\gamma \\ (1-t\gamma)^2 \end{bmatrix}, \ldots, x^{(k)} = \begin{bmatrix} (1-t)^k\gamma \\ (1-t\gamma)^k \end{bmatrix}.$$

Find the $t > 0$ that min

$$g(t) := f(\underbrace{x^{(k-1)} - t\nabla f(x^{(k-1)})}_{x^{(k)}})$$

$$f(x) = \frac{1}{2}(x_1^2 + \gamma x_2^2)$$

$$g(t) = \frac{1}{2}\Big(\overbrace{((1-t)^k\gamma)^2}^{\gamma^2(1-t)^{2k}} + \gamma(1-t\gamma)^{2k}\Big)$$

$$g'(t) = \cancel{\frac{1}{2}}\Big[\cancel{2k}(1-t)^{2k-1}(-1)\cancel{\gamma^2} + \cancel{\gamma}\cdot\cancel{2k}\cdot(1-t\gamma)^{2k-1}\cdot(-\cancel{\gamma})\Big]$$

$$= 0.$$

$$-(1-t)^{2k-1} = (1-t\gamma)^{2k-1}$$

$$\Downarrow$$

$$-(1-t) = 1-t\gamma$$

$$\Downarrow$$

$$t^* = \frac{2}{1+\gamma}$$

$$x^{(k)} = \begin{bmatrix} (1-t)^k \gamma \\ (1-t\gamma)^k \end{bmatrix} = \begin{bmatrix} \left(1 - \frac{2}{1+\gamma}\right)^k \gamma \\ \left(1 - t \cdot \frac{2}{1+\gamma}\right)^k \end{bmatrix} = \left(\frac{1-\gamma}{1+\gamma}\right)^k \begin{bmatrix} \gamma \\ 1 \end{bmatrix}.$$

Find out how $f(x^{(k)})$ evolves.

$$f(x) = \frac{1}{2}\left(x_1^2 + \gamma x_2^2\right) = \frac{\gamma(\gamma+1)}{2} \left(\frac{1-\gamma}{1+\gamma}\right)^{2k}$$

If $\gamma \approx 1$, then $\frac{1-\gamma}{1+\gamma} \approx 0$

If $\gamma \approx 0$, then $\frac{1-\gamma}{1+\gamma} \approx 1$

If $\gamma \approx 10^6$, then $\frac{1-\gamma}{1+\gamma} \approx -1$

5.

$$f(x) = \|x\|_2^{2+\beta} = \left(\left(x_1^2 + \cdots + x_n^2\right)^{\frac{1}{2}}\right)^{2+\beta}$$
$$= \left(x_1^2 + \cdots + x_n^2\right)^{1+\beta/2}.$$

$$\frac{\partial f}{\partial x_i} = (\nabla f(x))_i = \left(1 + \frac{\beta}{2}\right)\left(x_1^2 + \cdots + x_n^2\right)^{\beta/2} (2x_i)$$
$$= (2+\beta)\, \|x\|^\beta x_i \qquad \forall i = 1, \ldots, n.$$

$$\nabla f(x) = (2+\beta)\, \|x\|^\beta \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = (2+\beta)\, \|x\|^\beta x$$

GD with step size $s > 0$.

$$x^{(k+1)} = x^{(k)} - s \nabla f(x^{(k)})$$

$$= x^{(k)} - s\,(2+\beta)\,\|x^{(k)}\|^{\beta}\, x^{(k)}$$
$$= x^{(k)}\left[\,1 - s\,(2+\beta)\,\|x^{(k)}\|^{\beta}\,\right].$$

Claim: If $\|x^1\| < \|x^0\|$, then $\|x^{k+1}\| < \|x^k\|$ for all $k \geq 1$.

Pf: Suppose $\|x^1\| < \|x^0\|$. Assume $\|x^k\| < \|x^{k-1}\|$ for all $k \leq n$, $n \geq 2$.

$$\frac{\|x^{n+1}\|}{\|x^n\|} = \left|\,1 - s(2+\beta)\,\|x^n\|^{\beta}\,\right|$$

We only need to show that $\left|1 - s(2+\beta)\|x^n\|^{\beta}\right| < 1$.

$|a| < 1 \Leftrightarrow -1 < a < 1$

$$-1 < 1 - s(2+\beta)\,\|x^n\|^{\beta} < 1$$
$$0 < s(2+\beta)\,\|x^n\|^{\beta} < 2$$

By the induction hypothesis,

$$\frac{\|x^n\|}{\|x^{n-1}\|} < 1 \Rightarrow \quad 0 < s(2+\beta)\,\|x^{n-1}\|^{\beta} < 2 \qquad \text{have}$$

$$0 < s(2+\beta)\,\|x^n\|^{\beta} \overset{\text{I.H}}{<} s\,(2+\beta)\,\|x^{n-1}\|^{\beta} < 2.$$

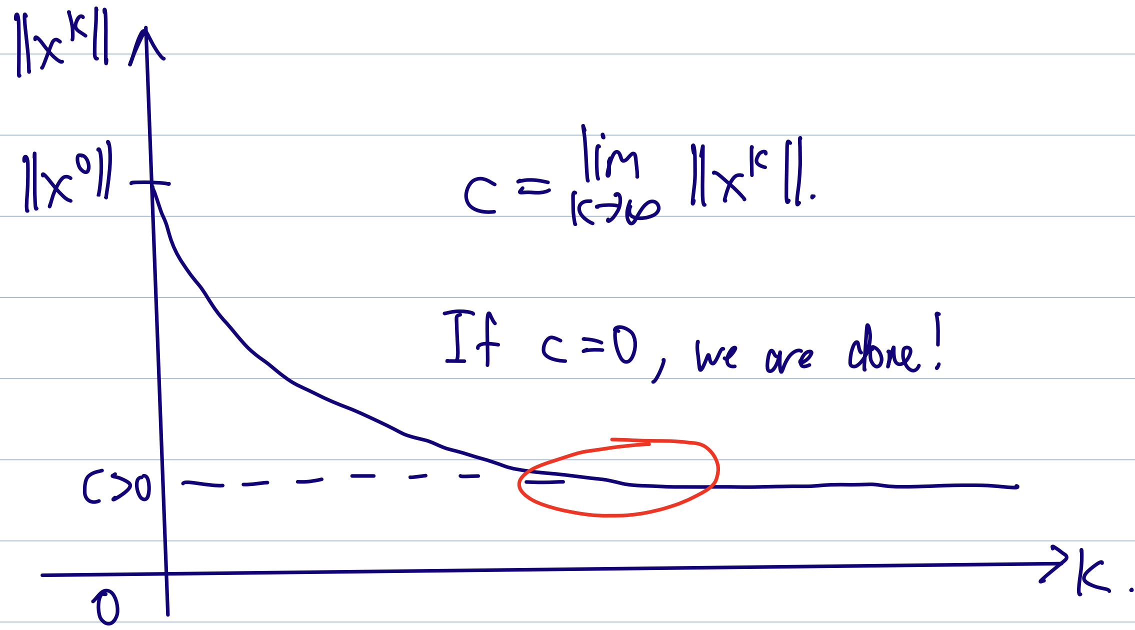
Claim: If $\|x^1\| < \|x^0\|$, then $\|x^{k+1}\| < \|x^k\|$.
for all $k \geq 1$.

$$x^1 = \left(1 - s(2+\beta)\|x^0\|^\beta\right) x^0$$

Thus for the GD method to converge, we need $\|x^1\| < \|x^0\|$.

$$\left|1 - s(2+\beta)\|x^0\|^\beta\right| < 1. \quad \text{---}(*)$$

If $(*)$ is satisfied $\{\|x^k\|\}$ is monotonically decreasing.



$$c = \lim_{k\to\infty} \|x^k\|.$$

If $c=0$, we are done!

$$\left|1 - s(2+\beta)\|x^0\|^\beta\right| < 1. \quad \text{---}(*)$$

$\|x^0\| \geq c$, $\quad 0 < s(2+\beta)\|x^0\|^\beta < 2$

$\Rightarrow \quad 0 < s(2+\beta)c^\beta < 2.$

$\Rightarrow \quad \left|1 - s(2+\beta)c^\beta\right| < 1 \quad \text{---}(**)$

$$\lim_{k\to\infty} \frac{\|x^{k+1}\|}{\|x^k\|} = 1$$

Combining (**) with the iterations, we see that

$$\lim_{k\to\infty} \frac{\|x^{k+1}\|}{\|x^k\|} = \left|1 - s(2+\beta)\underset{\lim_{k\to\infty}\|x^k\|}{\underset{\|}{c^\beta}}\right| \overset{(**)}{<} 1 \quad \Rightarrow\Leftarrow$$

c cannot be > 0. $\Rightarrow$ $c = 0$.

$$x^{k+1} = x^k\left(1 - s(2+\beta)\|x^k\|^\beta\right).$$

$$\lim_{k\to\infty} \frac{\|x^{k+1}\|}{\|x^k\|} = \lim_{k\to\infty} \left|1 - s(2+\beta)\|x^k\|^\beta\right|$$

$$= \left|1 - s(2+\beta)c^\beta\right|.$$