1. Which can result in non-constant 2 or 3-periodic sequences?

Gradient descent: $x_{k+1} = x_k - \gamma \nabla f(x_k)$

Polyak GD: $x_{k+1} = x_k - \gamma \nabla f(x_k) + \mu(x_k - x_{k-1})$

Nesterov GD: $x_{k+1} = x_k - \gamma \nabla f(x_k + \mu(x_k - x_{k-1})) + \mu(x_k - x_{k-1})$

2-periodic: $\exists N: \; x_{N+j} = x_{N+j+2} \quad \forall j \in \mathbb{N}$

3-periodic: $\exists N: \; x_{N+j} = x_{N+j+3} \quad \forall j \in \mathbb{N}$

GD:
$$x_{k+1} = x_k - \gamma \nabla f(x_k)$$
$$x_{k+2} = x_{k+1} - \gamma \nabla f(x_{k+1}) = x_k - \gamma \nabla f(x_k) - \gamma \nabla f(x_{k+1})$$

2-periodic: $x_k = x_{k+2} = x_k - \gamma \nabla f(x_k) - \gamma \nabla f(x_{k+1})$ for all $k > N$

$$\iff \gamma(\nabla f(x_k) + \cancel{\gamma} \nabla f(x_{k+1})) = 0 \qquad \gamma \neq 0 \text{, because non-const.}$$
$$\Rightarrow \nabla f(x_k) + \nabla f(x_{k+1}) = 0$$

If we choose $f(x) = x^2$: $\nabla f(x) = 2x$

if $x_1 = 1$: $\nabla f(x_1) = 2 = -\nabla f(x_2) \Rightarrow x_2 = -1$.

$$x_2 = 1 - \gamma \cdot 2 \Rightarrow \gamma = 1$$
$$x_3 = -1 - 1 \cdot (-2) = 1$$
☒ We have a 2-per. sequence

Polyak GD: We will try to construct a 3-periodic seq.: 0, 0, 1, 0, 0, ...

$$\underbrace{x_0 = x_1 = 0}, \; \underbrace{x_2 = 1} = 0 - \gamma \nabla f(0) + \mu \cdot 0 \Rightarrow \gamma = -\frac{1}{\nabla f(0)}$$
$$\underbrace{0 = x_3} = 1 + \frac{1}{\nabla f(0)} \cdot \nabla f(1) + \mu(1 - 0) \Rightarrow \mu = -1 - \frac{\nabla f(1)}{\nabla f(0)}$$
$$\underbrace{0 = x_4} = 0 + \frac{1}{\nabla f(0)} \cdot \nabla f(0) + \mu(0 - 1) \Rightarrow \boxed{\mu = 1}$$

Constraints: $1 = -1 - \frac{\nabla f(1)}{\nabla f(0)}$, $\nabla f(0) \neq 0$, $-\frac{1}{\nabla f(0)} > 0$, $f''(x) > 0$ (strictly convex)

We can choose linear $\nabla f(x) = kx + n$, with $n < 0$, $k > 0$

$$2 \nabla f(0) = -\nabla f(1)$$
$$2n = -k - n$$
$$3n = -k \xrightarrow{\text{choose}} k = 3 \Rightarrow n = -1$$

$$\Rightarrow \boxed{\gamma = 1}, \quad \nabla f(x) = 3x - 1, \quad f(x) = \frac{3}{2}x^2 - x$$

Check:
$$x_2 = 0 - 1 \cdot (-1) + \mu \cdot 0 = 1$$
$$x_3 = 1 - 1 \cdot 2 + 1 = 0$$
$$x_4 = 0 - 1 \cdot (-1) + 1 \cdot (0 - 1) = 0$$
→ we know (from x_2) that we have 1 after two zeros. ☒

Nesterov GD:

We will try to construct a 2-per. sequence : 0, 1, 0, 1, ...

$$x_{k+1} = x_k - \gamma \nabla f(x_k + \mu(x_k - x_{k-1})) + \mu(x_k - x_{k-1})$$

$$x_0 = 0$$

$$x_1 = 1$$

$$x_2 = 0 = 1 - \gamma \nabla f(1 + \mu \cdot 1) + \mu \cdot 1$$

$$\Rightarrow \quad \gamma = \frac{1+\mu}{\nabla f(1+\mu)}$$

$$x_3 = 1 = 0 - \frac{1+\mu}{\nabla f(1+\mu)} \nabla f(0-\mu) - \mu$$

$$\Rightarrow \quad (\cancel{1+\mu}) \cdot \nabla f(1+\mu) = -(\cancel{1+\mu}) \nabla f(-\mu)$$

$$x_4 = 0 = 1 - \frac{1+\mu}{\cancel{\nabla f(1+\mu)}} \cancel{\nabla f(1+\mu)} + \mu$$

$$1+\mu = 1+\mu \quad \checkmark$$

Our constraints : $\gamma = \frac{1+\mu}{\nabla f(1+\mu)} > 0 \Rightarrow \nabla f(1+\mu) > 0$

$\nabla f(1+\mu) = -\nabla f(-\mu) \Rightarrow \nabla f(-\mu) < 0$

We could reach that by linear function: $\nabla f(x) = kx + n$. $k > 0$, so that f is convex.

We can set $\mu = 1$.

$$2k + n = -(-k+n) \qquad 2k + n > 0 \Rightarrow n < 0$$

$$k = -2n$$

Let's set $n = -1$, $k = 2$ $\Rightarrow \nabla f(x) = 2x - 1$

$\Rightarrow f(x) = x^2 - x$

$$\gamma = \frac{2}{2 \cdot 2 - 1} = \frac{2}{3}$$

$$\Rightarrow \quad x_{k+1} = x_k - \frac{2}{3}\left(2(x_k + x_k - x_{k-1}) - 1\right) + x_k - x_{k-1} =$$

$$= 2x_k - x_{k-1} - \frac{8}{3} x_k + \frac{4}{3} x_{k-1} + \frac{2}{3} = \frac{2}{3} - \frac{2}{3} x_k + \frac{1}{3} x_{k-1}$$

$x_0 = 0$, $x_1 = 1$, $x_2 = \frac{2}{3} - \frac{2}{3} = 0$, $x_3 = \frac{2}{3} + \frac{1}{3} = 1, \ldots$ $\boxtimes$

2. $f(x,y,z) = x^2 + 2y^2 - 2yz + 4z^2 + 3x - 4y + 5z =$

$$= \begin{bmatrix} x & y & z \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & -1 \\ 0 & -1 & 4 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} + \begin{bmatrix} 3, & -4, & 5 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

→ Quadratic function. (convex ✓)

⇒ For Polyak GD the optimal choice for γ and μ is:

$$\gamma = \frac{4}{(\sqrt{\alpha} + \sqrt{\beta})^2}, \quad \sqrt{\mu} = \frac{\sqrt{\beta} - \sqrt{\alpha}}{\sqrt{\beta} + \sqrt{\alpha}}$$

$$H = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 4 & -2 \\ 0 & -2 & 8 \end{bmatrix} \qquad \lambda_1 = 2 \longrightarrow \text{H is PSD}$$

$$(2-\lambda)((4-\lambda)(8-\lambda) - 4) = 0$$

$$32 - 12\lambda + \lambda^2 - 4 = 0$$

$$\lambda^2 - 12\lambda + 28 = 0$$

$$D = 12^2 - 4 \cdot 28 = 32$$

$$\lambda_{2,3} = \frac{12 \pm \sqrt{32}}{2} = 6 \pm 2\sqrt{2}$$

→ $\alpha = 2, \quad \beta = 6 + 2\sqrt{2}$

$$\gamma = \frac{4}{(\sqrt{2} + \sqrt{6+2\sqrt{2}})^2} \qquad \mu = \left(\frac{\sqrt{6+2\sqrt{2}} - \sqrt{2}}{\sqrt{6+2\sqrt{2}} + \sqrt{2}}\right)^2 = \frac{6 + 2\sqrt{2} - \sqrt{12+4\sqrt{2}} + 2}{6 + 2\sqrt{2} + \sqrt{12+4\sqrt{2}} + 2} =$$

$$= \frac{8 + 2\sqrt{2} - \sqrt{12+4\sqrt{2}}}{8 + 2\sqrt{2} + \sqrt{12+4\sqrt{2}}}$$