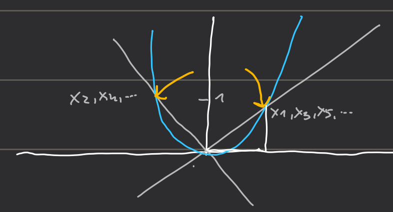
HW II

① L-Lipschitz:

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



$f(x) = ax^2 \qquad f'(x) = 2ax$

$x_{k+1} = (1-2a\gamma)^k x_1 = (1-2a\gamma)\, x_k$

↳ If we want periodic, we need $(1-2a)^k = 1$

→ 2-periodic: $(1-2a\gamma)^2 = 1$

↳ CASE 1: $1-2a\gamma = 1 \Rightarrow 2a\gamma = 0 \Rightarrow \gamma = 0$

↳ CASE 2: $1-2a\gamma = -1 \Rightarrow 2a\gamma = 2 \Rightarrow a\gamma = 1 \Rightarrow \underline{\gamma = \frac{1}{a}}$ (Gives $(-1)^k \cdot x_1$, projection over y)

For function $f(x) = 2x^2$ the GD will result in periodic sequence of period 2 for learning rate $\gamma = \frac{1}{2}$

Starting point $x_1 = 1$, $f(1) = 2 \qquad f'(x) = 4x \qquad \gamma = \frac{1}{2}$

$x_2 = 1 - \frac{1}{2} \cdot 4 = 1 - 2 = -1$

$x_3 = -1 - \frac{1}{2} \cdot (-4) = 1 \qquad x_{k+1} = (1-2)^k x_1 = (-1)^k \cdot 1$

$\underline{x_{k+1} = (-1)^k \cdot 1}$ Therefore GD can result in non-constant 2 periodic sequence

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

3-sequence $\quad 0, 0, 1 \rightarrow x_1 = 0 \quad x_2 = 0 \quad x_3 = 1$

$x_3 = 1 = 0 - \gamma \nabla f(0) + \mu(0+0) = -\gamma \nabla f(0) \Rightarrow \gamma = -\frac{1}{2 \cdot \nabla f(0)}$

$x_4 = 0 = 1 - \left(-\frac{1}{2\nabla f(0)}\right) \cdot \nabla f(1) + \mu(1-0)$

$1 + \frac{\nabla f(1)}{2\nabla f(0)} + \mu = 0 \Rightarrow \mu = -\frac{\nabla f(1)}{2\nabla f(0)} - 1$

$\nabla f(0) \neq 0$

$x_5 = 0 = 0 + \frac{\nabla f(0)\ (=\nabla f(x_3))}{2\nabla f(0)\ (=\gamma)} - \mu$

$\nabla f(x) = ax^2 + b$ (convex)

$\frac{1}{2} \cdot 1 - \mu = 0 \qquad -\frac{\nabla f(1)}{2\nabla f(0)} - 1 = \frac{1}{2}$

$\underline{\mu = \frac{1}{2}}$

$-\frac{\nabla f(1)}{2\nabla f(0)} = \frac{3}{2} \Rightarrow -\nabla f(1) = 3\nabla f(0)$

$-(a+b) = 3b$

$-a = 4b \qquad b = -\frac{3}{2} \quad a = 6$

$a = -4b \qquad \nabla f(x) = 6x^2 - \frac{3}{2}$

Compute γ:

$\underline{\gamma} = \frac{1}{2\nabla f(0)} = \frac{1}{2 \cdot \frac{1}{2}} = \underline{1}$

$f(x) = \frac{6x^3}{3} - \frac{3}{2}x = 2x^3 - \frac{3}{2}x$

$x_3 = 0 - 1 \cdot (-\frac{3}{2}) + 0 = 0$

$\nabla f(x) = 4x^2 - 1$

Test: $f(x) = \frac{4x^3}{3} - x \qquad \gamma = 1 \quad \mu = \frac{1}{2} \qquad x_1 = 0,\ x_2 = 0$

$x_3 = x_2 - \gamma \cdot \nabla f(x_2) + \mu (x_2 - x_1)$

$= 0 - 1 \cdot (-\frac{1}{2}) + 0 = \frac{1}{2}$

Let's try an arbitrary 3 sequence: $[0, 0, a, 0, 0, a, \dots]$

$x_1 = 0 \quad x_2 = 0 \quad x_3 = a$

$x_3 = 0 - \gamma \nabla f(0) + \mu(0) \Rightarrow -\gamma \nabla f(0) = a$

$\gamma = -\frac{a}{\nabla f(0)} = -a \cdot (-2) = 2a$

$x_4 = 0 = a + \frac{a}{\nabla f(0)} \cdot \nabla f(a) + a\mu \Rightarrow \mu = \left(-a - \frac{a}{\nabla f(0)}\right) : a = \frac{-\cancel{a}\left(1 + \frac{1}{\nabla f(0)}\right)}{\cancel{a}} = -1 - \frac{1}{\nabla f(0)} = -1+2 = 1$

$x_5 = 0 = 0 + \frac{a}{\nabla f(0)} \cdot \nabla f(0) + \left(-1 - \frac{1}{\nabla f(0)}\right)(-a) = 0 + a + a + \frac{a}{\nabla f(0)} = 2a + \frac{a}{\nabla f(0)} = 0$

$a\left(2 + \frac{1}{\nabla f(0)}\right) = 0$

$a = 0$

No, or sequence is constant

$\frac{1}{\nabla f(0)} = -2$

$-2\nabla f(0) = 1$

$\nabla f(0) = -\frac{1}{2}$

The intercept of ∇f is $-\frac{1}{2}$

$\hookrightarrow \underline{\mu = 1} \quad \underline{\gamma = 2a}$

$a + \frac{a}{-\frac{1}{2}} \cdot \nabla f(a) + a = 0$

$2a - 2a \nabla f(a) = 0$

$2a \nabla f(a) = 2a$

$\underline{\nabla f(a) = 1}$ Another condition

Let's pick

$f(x) = x^2 - \frac{1}{2}x \quad \mu = 1 \quad \gamma = 1 \qquad \nabla f(\tfrac{1}{2}) = \frac{1}{2}$

$\nabla f(x) = 2x - \frac{1}{2}$

$x_3 = 0 - 1 \cdot \nabla f(0) + 1 \cdot 0 = 0 + \frac{1}{2} + 0 = \frac{1}{2}$

$x_4 = \frac{1}{2} - 1 \cdot \nabla f(\tfrac{1}{2}) + 1 \cdot \frac{1}{2} = \frac{1}{2} - \frac{1}{2} + \frac{1}{2} = \frac{1}{2}$

$a \neq 0$ we pick $a = \frac{1}{2}$

$\Rightarrow$ Find $f(x)$, s. that $\nabla f(a) = 1 \quad \nabla f(0) = -\frac{1}{2} \quad \gamma = 2a$ and $\mu = 1$

derivative of first term must be $\frac{3}{2}$ & first term must be of power > 1

$\nabla f(\tfrac{1}{2}) = 1 = cx - \frac{1}{2}$

$3x - \frac{1}{2} \rightarrow f(x) = 1.5x^2 - \frac{1}{2}x$

Test: $x_3 = 0 - 1 \cdot \nabla f(0) + 0 = \frac{1}{2}$

$x_4 = \frac{1}{2} - 1 \cdot \nabla f(\tfrac{1}{2}) + \frac{1}{2} = \frac{1}{2} - 1 + \frac{1}{2} = 0$ ✓

$x_5 = 0 - 1 \nabla f(0) - \frac{1}{2} = 0 + \frac{1}{2} - \frac{1}{2} = 0$ ✓

$x_6 = 0 - 1 \nabla f(0) + 0 = 0 + \frac{1}{2} + 0 = \frac{1}{2}$ ✓

If for $f(x) = \frac{3}{2}x - \frac{1}{2}$ we pick $\gamma = 1$ $\mu = 1$ we get a sequence $[0, 0, \frac{1}{2}, 0, 0, \frac{1}{2}, \dots]$

Nesterov

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

sequence $0, 0, a$

$$x_3 = 0 - \gamma \cdot \nabla f(0) + 0 = -\gamma \nabla f(0) = -a \Rightarrow \gamma = -\frac{a}{\nabla f(0)} \quad (1)$$

$$x_4 = a + \frac{a}{\nabla f(0)} \cdot \nabla f(a + a\mu) + a\mu = 0 \qquad \mu = -a - \frac{a}{\nabla f(0)} \cdot \nabla f(a + a\mu) \,/\, a$$

Fix the ∇f: $\nabla f(x) = ax + \cancel{b}$ $b=1$

$[0, 0, z]$: $x_3 = 0 - \gamma \cdot b + 0 = -\gamma b = z \Rightarrow \boxed{\gamma = -\frac{z}{b}}$ $b=1$

$$x_4 = 0 = z + \frac{z}{b} \cdot \nabla f(z + \mu \cdot z) + \mu z = z + \frac{z}{b} \cdot (a(z + \mu z) + b) + \mu z = z + \frac{z}{b}(az + \mu a z + b) + \mu z$$

$$= z + \frac{az^2}{b} + \frac{\mu a z^2}{b} + z + \mu z$$

$$= \frac{\mu a z^2}{b} + \mu z = -2z - \frac{az^2}{b}$$

$$= \mu\left(\frac{az^2}{b} + z\right) = -2z - \frac{az^2}{b}$$

$$\mu = \frac{-2z - \frac{az^2}{b}}{\frac{az^2}{b} + z}$$

$a = 1$ $b = 1$ $z = 1{,}436$

$$\mu = \frac{-2{,}872 - 2{,}062}{2{,}062 + 1} = -1{,}410$$

$$\frac{-2z - z^2}{z^2 + z}$$

$$x_5 = 0 + \frac{z}{b} \cdot \nabla f\left(0 - z\left(\frac{-2z - \frac{az^2}{b}}{\frac{az^2}{b} + z}\right)\right) - z\left(\frac{-2z - \frac{az^2}{b}}{\frac{az^2}{b} + z}\right)$$

$$= 0 + \frac{z}{b} \cdot a\left(\frac{2z^2 + \frac{az^3}{b}}{\frac{az^2}{b} + z}\right) + \frac{2z^2 + \frac{az^3}{b}}{\frac{az^2}{b} + z} = 0$$

$$0 + z \cdot \frac{2z^2 + z^3}{\cancel{z^2 + z}} + \frac{2z^2 + z^3}{\cancel{z^2 + z}} = 0$$

$$\frac{\frac{2az^2}{b} + \frac{a^2z^4}{b^2}}{\cancel{\frac{az^2}{b} + z}} + \frac{\frac{2z^2}{1} + \frac{az^3}{b}}{\cancel{\frac{az^2}{b} + z}} = 0$$

$$2z^3 + z^4 + 2z^2 + z^3 = 0$$

$$z^2(z^2 + 3z + 2) = 0$$

$z_1 = -2$ $z_2 = 1$

$$\frac{2az^2}{b} + \frac{a^2z^4}{b^2} + \frac{2z^2}{1} + \frac{az^3}{b} = 0$$

$z = -5$ $\gamma = 5$ μ

$$\frac{2az^2 + az^3}{b} + \frac{a^2z^4}{b^2} + 2z^2 = 0$$

$[0, 0, -5, 0, 10.]$

$b=1$

$$2az^2 + az^3 + a^2z^4 + 2z^2 = 0$$

$x_6 =$

$$z^2(\underbrace{z^2(a^2)}_{a} + \underbrace{z(a)}_{b} + \underbrace{(2+2a)}_{c}) = 0$$

$z_{1,2} = 0$ NOT OK, CONSTANT SEQENCE

$$z_{3,4} = \frac{-a^2 \pm \sqrt{a^2 - 4a^2(2+2a)}}{2a^2} = \frac{-a^2 \pm \sqrt{a^2 - 8a^2 - 8a^3}}{2a^2}$$

$$= \frac{-a^2 \pm \sqrt{a^2(-7-8a)}}{2a^2}$$

$\boxed{a=1} \Rightarrow z = \frac{-1 \pm \sqrt{15}}{2}$

$z_1 = \frac{-1 + \sqrt{15}}{2}$

$\left[0, 0, \frac{-1+\sqrt{15}}{2}\right]$ ($\frac{-1+\sqrt{15}}{2} = 1{,}436$)

$\nabla f(x) = x + 1$ $f(x) = \frac{x^2}{2} + x$

$\gamma = -z = -1{,}436$ $\mu =$

$z = -2$ seq $= [-2, 0, -2, 0 \ldots)$

but $\mu = 0$ so NOT ok :-/

(2) Determine γ_{opt} : $f(x,y,z) = x^2+2y^2-2yz+4z^2+3x-4y+5z$

$$\frac{\partial f}{\partial x} = 2x+3 \qquad \frac{\partial f}{\partial y} = 4y-2z-4 \qquad \frac{\partial f}{\partial z} = -2y+8z+5$$

$$H = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 4 & -2 \\ 0 & -2 & 8 \end{bmatrix}$$

$$\det(H-\lambda I) = \begin{vmatrix} 2-\lambda & 0 & 0 \\ 0 & 4-\lambda & -2 \\ 0 & -2 & 8-\lambda \end{vmatrix} = (2-\lambda)\cdot\left((4-\lambda)(8-\lambda)-(4)\right) = (2-\lambda)(32-12\lambda+\lambda^2-4) = (2-\lambda)(\lambda^2-12\lambda+28) = 0$$

$\lambda_1 = 2 = \alpha$

$$\lambda_{2,3} = \frac{12 \pm \sqrt{12^2-4\cdot 28}}{2} = 6 \pm 2\sqrt{2}$$

$\lambda_2 = 6-2\sqrt{2}$ $\quad \lambda_3 = 6+2\sqrt{2} = \beta$

$$\sqrt{\mu} = \frac{\sqrt{\beta}-\sqrt{\alpha}}{\sqrt{\beta}+\sqrt{\alpha}} = \frac{\sqrt{6+2\sqrt{2}}-\sqrt{2}}{\sqrt{6+2\sqrt{2}}+\sqrt{2}} \approx 0,35504 \Rightarrow \underline{\mu \approx 0,126058}$$

$$\gamma = \frac{4}{(\sqrt{\alpha}+\sqrt{\beta})^2} = \frac{4}{\left(\sqrt{6+2\sqrt{2}}+\sqrt{2}\right)^2} \approx \underline{0,20798}$$