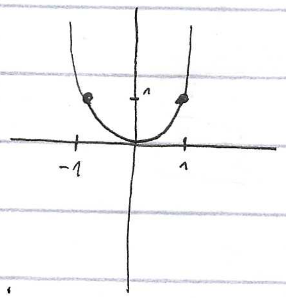
HOMEWORK 2

① A sequence $\{x_i\}_i$ is eventually p-periodic if for some $N, p \in \mathbb{N}$ we have $x_{N+j} = x_{N+j+p}$, $\forall j \in \mathbb{N}$. Which of the following procedures on a strictly convex function may result in non-constant 2- or 3-periodic sequences for an appropriate selection of fixed positive parameters γ, μ and starting point x_1.

a) Gradient descent $\boxed{x_{t+1} = x_t - \gamma \cdot \nabla f(x_t)}$

I) 2-periodic : GD may result in non-constant 2-periodic sequence. We provide an example:

$f(x) = x^2$ $\nabla f(x) = 2x$ $\gamma = 1$ $x_1 = 1$ or any $x \in \mathbb{R} \setminus \{0\}$



$x_1 = 1$

$x_2 = 1 - 1 \cdot 2 = -1$

$x_3 = -1 - 1 \cdot (-2) = 1$

$\vdots$

II) 3-periodic

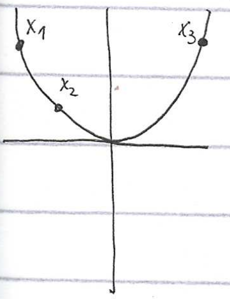
FIXED LEARNING RATE

GD may NOT result in non-constant 3-periodic sequence.

We know that for a strictly convex $\|\nabla f(x)\|$ increases as we go further away from minima. To achieve 3-periodicity ~~we have to cross one point~~ one value of gradient must be the opposite sign (let's say $\nabla f(x_3)$) of the other 2. At the same time $\|\nabla f(x_1)\| < \|\nabla f(x_2)\|$.

↓ So we are able to come back

↓ so sign $\nabla f(x_1) \overset{=}{\cancel{\neq}}$ sign $\nabla f(x_2)$ and we can jump to x_3.



However we choose these 3 points that follow gradient descent, we can NOT satisfy these conditions.

More intuitively: if ~~first~~ a step does not change sign $\nabla f(x_{t+1})$, neither will the next.