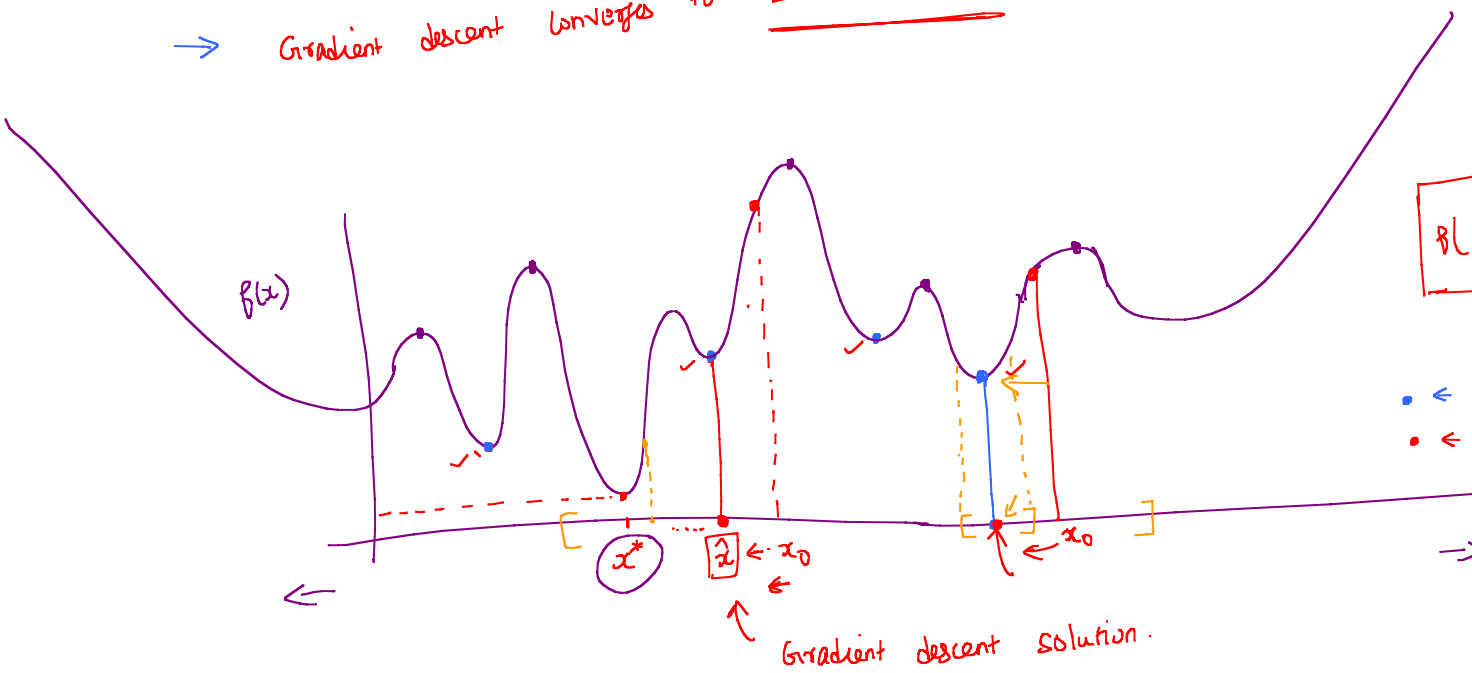
Gradient - descent

Properties

→ If $\eta_t = \frac{1}{t+1}$, the algorithm converges

→ Gradient descent converges to "LOCAL MINIMUM"



Local minimum

$\exists \epsilon > 0$ s.t

$f(\hat{x}) \leq f(x)$

$\forall x \in [\hat{x}-\epsilon, \hat{x}+\epsilon]$

Global minimum

$f(x^*) \leq f(x)$ $\forall x$

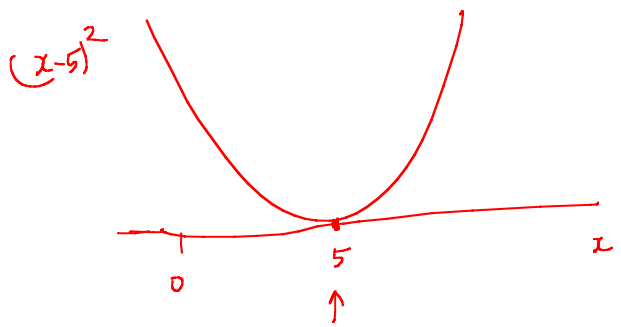
• ← Local minimum

• ← Local minimum + global minimum

Functions for which

local minimum ⇕ Global minimum

← CONVEX FUNCTIONS



What is special about

$d = \boxed{-f'(x)}$