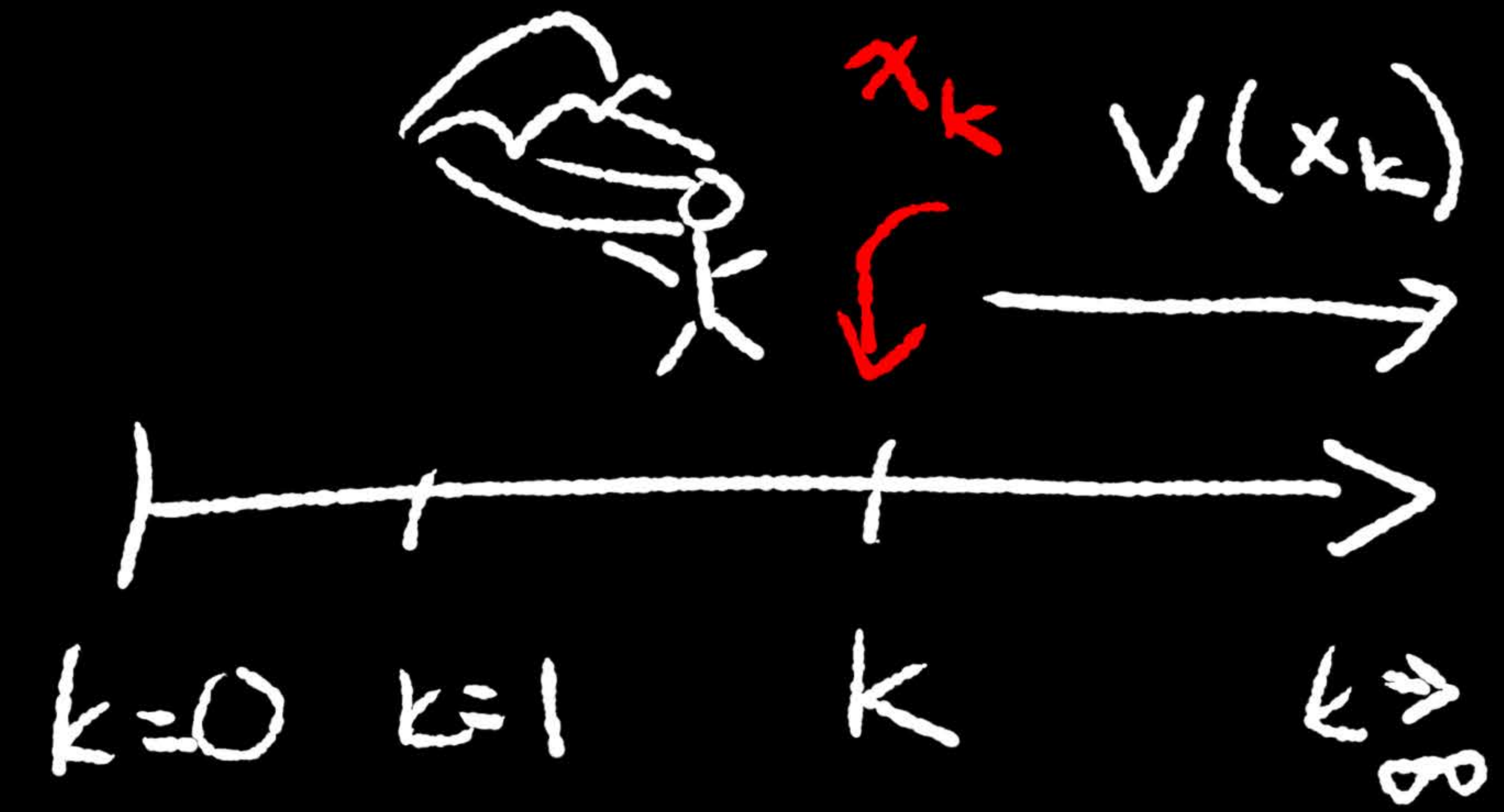
II. Dynamic Programming (DP)



Algorithmic method to solve optimal control problems

Consider the discounted optimal control formulation

$\gamma \in [0,1]$

$$\text{minimize } J = \sum_{k=0}^{\infty} \gamma^k \cdot c(x_k, u_k)$$

$$\text{subject to: } x_{k+1} = f(x_k, u_k) \quad k = 0, 1, \ldots$$

Defn (Value Function): Define $V(x_k)$ as the cumulative cost from k onward toward ∞, given the current state is x_k.

"cost-to-go"

Let $V_{\pi}(x_k)$ represents the value function corresponding to control policy $\pi(\cdot)$, that $u_k = \pi(x_k)$ which may or may not be optimal.

Note $V_{\pi}(x_k) = \underbrace{c(x_k, u_k)}_{\text{instantaneous cost}} + \gamma \cdot \underbrace{\sum_{\tau=k+1}^{\infty} \gamma^{\tau-(k+1)} \cdot c(x_\tau, u_\tau)}_{= V_{\pi}(x_{k+1})}$

$$V_{\pi}(x_k) = c(x_k, u_k) + \gamma \cdot V_{\pi}(x_{k+1})$$

which can be used to recursively calculate $V_{\pi}(\cdot)$

We are now positioned to write Bellman's Principle of Optimality Equations.

$$\boxed{\begin{aligned} V(x_k) &= \min_{\pi(\cdot)} \left\{ c(x_k, \pi(x_k)) + \gamma \cdot V(x_{k+1}) \right\} \\ \text{where} \quad x_{k+1} &= f(x_k, \pi(x_k)) \end{aligned}}$$

The optimal policy is

$$\pi^*(x_k) = \arg\min_{\pi(\cdot)} \left\{ c(x_k, \pi(x_k)) + \gamma \cdot V(x_{k+1}) \right\}$$

Remark: Bellman's Principle of Optimality Eqn is also known as the Hamilton-Jacobi-Bellman (HJB) equation.

Notes ① The eqn is recursive in $V(x_k)$

② Offline "planning" method.