# Generalizaciones del MDP

1) Recompensas Estado-Acción $r(s,a)$
2) Horizonte Finito $(T)$ $T$
3) Sistemas Dinámicos Lineales
   - LQR
   - Basados Modelo

## 1. Recompensas Estado-Acción

IVA

$$S \rightarrow R$$

$$S \times A \rightarrow R$$

$$s_0 \xrightarrow{a_0} s_1 \xrightarrow{a_1} s_2 \; \ldots\ldots$$

$$R(s_0, a_0) + \gamma^1 R(s_1, a_1) + \gamma^2 R(s_2, a_2) \ldots$$

Ec. Bellman

$$\boxed{V^*(s) = \max_a \left[ R(s,\underline{a}) + \gamma \sum_{s'} P_{sa}(s') \underline{V^*(s')} \right]}$$

las recompensas dependen tbn de la acción

$$\boxed{\pi^*(s) = \arg\max_a R(s,\underline{a}) + \gamma \sum P_{sa}(s') V^*(s')}$$

2da Generalización:

Horizonte Finito

MDP $\{S, A, \{P_{sa}\}, \underline{T}, R\}$

(sin $\gamma$!)

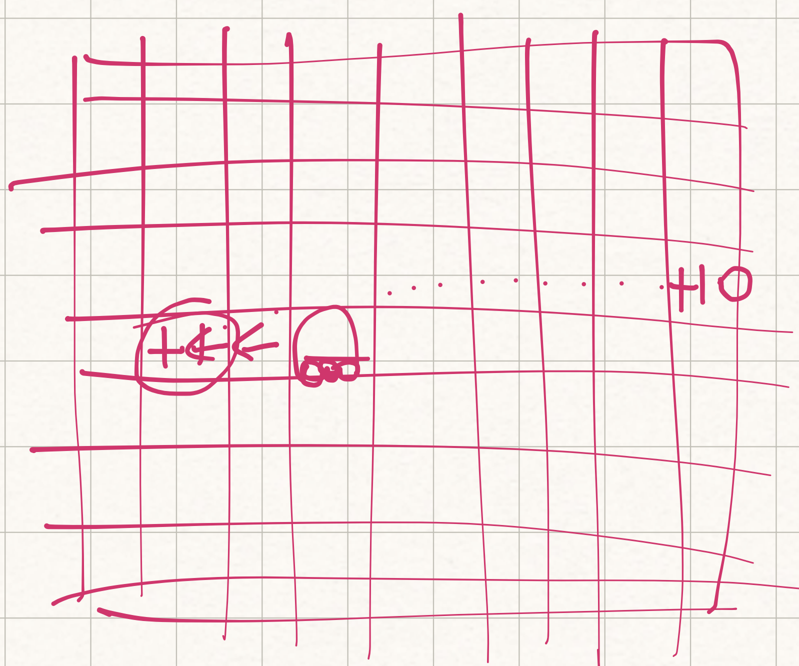
Suponde la Gen 1ra

$$\underbrace{R(s_0, a_0) + R(s_1, a_1) + \ldots\ldots + R(s_T, a_T)}_{\text{ganancia total}} \quad t$$

$$E\left[ \quad \right]$$

maximiza

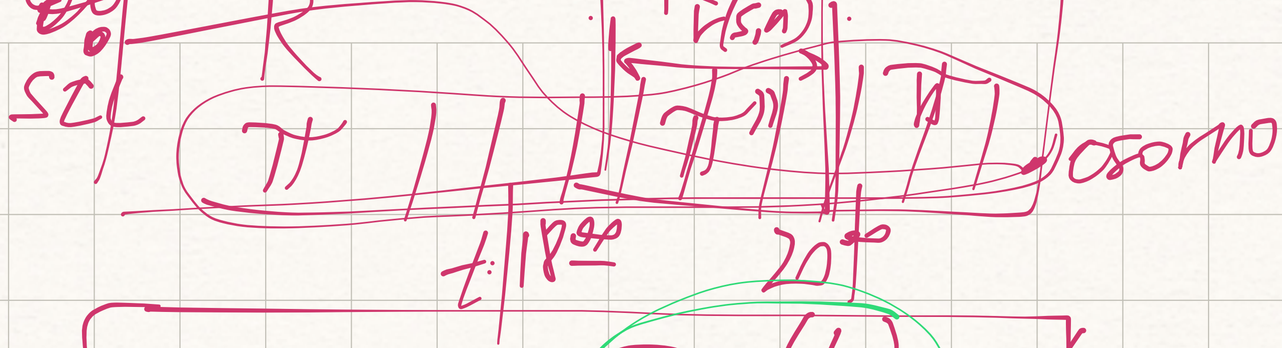
## E vs E




La acción óptima depende del reloj del robot.

$\pi_T^*(s) \rightarrow$ No Estacionaria (cambia con el tiempo)

$\pi^*(s) \rightarrow$ Estacionario




$$\boxed{S_{t+1} \sim P^{(t)}_{s_t a_t} \qquad R^{(t)}_{(s,a)}}$$

**Ejemplo: Avión con combustible**

$$V_T^*(s) = E\left[R(s_t, a_t) + R(s_{t+1}, a_{t+1}) + \cdots + R(s_T, a_T) \mid \pi, s_0 = s\right]$$

$$V_t^*(s) = \max_a \underbrace{R^{(t)}(s,a) + \sum_{s'} P^{(t)}_{sa}(s') V^*_{t+1}(s')}$$

$$\pi^*(s) = \arg\max_a \left[ \qquad \right]$$

$V_T \rightarrow$ Obte... $V$

$V_{10}$, Obtener $V_9$

$$\boxed{V_t^*(s) = \max_a R(s,a)}$$

## Programación Dinámica

## 3) Sist. Dinámicos Lineales

$\Rightarrow$ MDP $\{S, A, P_{sa}, T, R\}$

$\Rightarrow S = \mathbb{R}^n$

$\Rightarrow A = \mathbb{R}^d$

$\Rightarrow P_{sa} \rightarrow \boxed{S_{t+1} = A^{(t)} S_t + B^{(t)} a_t + \boxed{W_t}}$

ruido

$\Rightarrow W_t \sim \mathcal{N}(\underline{0}, \underline{\underline{\Sigma_w}})$ - - - - Tarea

$\Rightarrow$ Recompensas Cuadráticas

$R(s,a) = -\left(s^T U_t s_t + a_t^T W_t a_t\right)$

$$R(s,a) = -\left(\|s\|^2 + \|a\|^2\right)$$

¿Cómo se encuentran estas matriz $A$ y $B$ ?

→ $\boxed{\text{Regresión lineal}}$

$$\boxed{\begin{matrix} s_0 & \cdots \\ s_0^{(n)} & \cdots \end{matrix}} \rightarrow As_t + Ba_t$$

→ $\boxed{\text{Linearización}}$

$$\boxed{s_{t+1} = f(s_t, a_t)}$$

CartPole

$$\begin{pmatrix} x_{t+1} \\ \dot{x}_{t+1} \\ \theta_{t+1} \\ \dot{\theta}_{t+1} \end{pmatrix} = f\left(\begin{pmatrix} x_t \\ \dot{x}_t \\ \theta_t \\ \dot{\theta}_t \end{pmatrix}, a_t\right)$$

## Series de Taylor

$s_{t+1}$ ↑ , $f$


$$s_{t+1} = f(s_t)$$

$$\boxed{f(s_t) \approx f(\bar{s}_t) + f'(\bar{s}_t)(s_t - \bar{s}_t)}$$

Caso Gral

$$s_{t+1} = f(s_t, a_t)$$

$$s_{t+1} \approx \Big[ f(\bar{s}_t, \bar{a}_t) + (\nabla_s f(\bar{s}_t, \bar{a}_t)(s_t - \bar{s}_t) + (\nabla_a f(\quad)(a_t - \bar{a}_t) \Big]$$

[Subir Webcursos, este material en PDF]