## Linear System

(type of state transitions for continuous MDPs)

$$\underset{\in \mathbb{R}^{n_s}}{s_{t+1}} = As_t + B\underset{\in \mathbb{R}^{n_a}}{a_t} + \underset{\sim \mathcal{N}(0, \sigma^2 I)}{w_t}$$

Last time:

$\rho(A)$ spectral radius

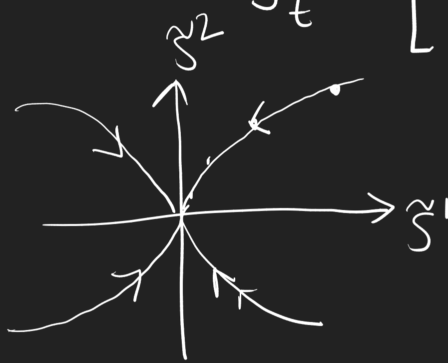
determining stability when $\underline{a_t = 0}$ and $\underline{w_t = 0}$

Key idea: $s_{t+1} = \underline{A} s_t$ $\qquad A = \underbrace{V \overset{\text{diagonal}}{D} V^{-1}}$ (if diagonalizable)

$$\begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_{n_s} \end{bmatrix}$$

Then $\tilde{s}_t = V^{-1} s_t$

$$\tilde{s}_t = \begin{bmatrix} \lambda_1^t & & \\ & \ddots & \\ & & \lambda_{n_s}^t \end{bmatrix} \tilde{s}_0$$




$1 > \lambda_2 > \lambda_1 > 0$

Alternative:

$s_{t+1} = As_t$

$s_0 = \underline{v_i}$ (eigenvector, so $Av_i = \lambda_i v_i$)

$s_1 = As_0 = Av_i = \lambda v_i$

$s_t = \lambda^t v_i$

## 1) LQR (Linear Quadratic Regulator)

$$s_{t+1} = \underline{A} s_t + \underline{B} a_t + w_t \qquad w_t \sim \mathcal{N}(0, \sigma^2 I)$$

$$c(s, a) = s^T Q s + a^T R a$$

$Q, R$ symmetric $Q^T = Q$
and positive definite (positive eigenvalues)

OCP

$$\min_{\pi} \mathbb{E}\left[ s_H^T Q s_H + \sum_{t=0}^{H-1} s_t^T Q s_t + a_t R a_t \;\middle|\; \begin{matrix} s_{t+1} = A s_t + B a_t + w_t, & w_t \sim \mathcal{N}(0, \sigma^2 I) \\ a_t = \pi_t(s_t), & s_0 \sim \mu_0 \end{matrix} \right]$$

ex 1D robot

$$s_{t+1} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} s_t + \begin{bmatrix} 0 \\ 1/m \end{bmatrix} a_t \qquad s_t = \begin{bmatrix} p_t \\ v_t \end{bmatrix}$$

$$c(s, a) = \gamma_p p_t^2 + \gamma_v v_t^2 + \gamma_a a_t$$

Value & Q functions

"cost-to-go"

$$\rightarrow V_t^\pi(s) = \mathbb{E}\left[ s_H^T Q s_H + \sum_{k=t}^{H-1} s_k^T Q s_k + a_k^T R a_k \;\middle|\; \begin{matrix} \text{dynamics} \\ a_k = \pi_k(s_k) \\ s_t = s \end{matrix} \right]$$

$Q_t^\pi(s, a)$ " " $a_t = a$

Because of terminal cost,

$$- V_H^\pi(s) = s^T Q s$$

2) Optimal LQR Policy

Dynamic Programing for OCP:

Start: $V_H^\pi(s) = c_H(s)$

for $t = H-1, H-2, \ldots, 0$:

$$Q_t^*(s, a) = c_t(s, a) + \mathbb{E}_{s' \sim P(s,a)}\left[V_{t+1}^*(s')\right] \qquad s' = f(s, a, w),\ w \sim \mathcal{D}$$

$$\pi_t^*(s) = \operatorname*{argmin}_a Q_t^*(s, a)$$

$$V_t^*(s) = Q_t^*(s, \pi_t^*(s))$$

Theorem (LQR optimal value function & policy)

Given $(A, B, Q, R, \sigma^2)$ .

$$V_t^*(s) = s^\top P_t s + p_t$$

$$\pi_t^*(s) = -K_t^* s \qquad K_t^* \in \mathbb{R}^{n_s \times n_a}$$

where $P_t, K_t, p_t$ depend on $(A, B, Q, R, \sigma^2)$

Proof: by induction

claim 1: (Base case) $V_H^*(s) = s^\top P_H s + p_H$ is quadratic.

claim 2: (induction) if $V_{t+1}^*(s) = s^\top P_{t+1} s + p_{t+1}$

Then

1) $Q_t^*(s,a)$ is quadratic in $s,a$

2) $\pi_t^*(a) = \underset{a}{\operatorname{argmin}}\, Q_t^*(s,a)$ is linear in $s$

Thus $V_t^*(s) = s^\top P_t s + p_t$ is quadratic.

$P_H = Q$ and $p_H = 0$

$$Q_t^*(s,a) = c(s,a) + \mathbb{E}_{s'}[V_{t+1}^*(s')]$$

$$V_{t+1}^*(s) = s^\top P_{t+1} s + p_{t+1}$$

$$\mathbb{E}_{w_t \sim \mathcal{N}(0,\sigma^2 I)}\left[V_{t+1}^*(As_t + Ba_t + w_t)\right]$$

$$= (As)^\top P_{t+1}(As) + (As)^\top P_{t+1} Ba + 0$$

$$(Ba)^\top P_{t+1} As + (Ba)^\top P_{t+1} Ba + 0$$

$$0 + 0 + \underbrace{\mathbb{E}_w[w^\top P_{t+1} w]} + p_{t+1}$$

$$\rightarrow \mathbb{E}_w[w^\top P w] = \sigma^2 \operatorname{Tr}(P)$$

$$\mathbb{E}(\operatorname{Tr}(w^\top P w)) = \operatorname{Tr}(P\,\mathbb{E}[ww^\top]) = \quad \sigma^2 I$$

$$\mathbb{E}_{w \sim \mathcal{N}(0,\sigma^2 I)}[Mw] = M\,\mathbb{E}[w] = M 0 = 0$$

$$\mathbb{E}[w^\top w] \neq \mathbb{E}w^\top \mathbb{E}w$$

$$Q_t^*(s,a) = s^T\underbrace{(Q + A^T P_{t+1} A)}_{M_1} s + a^T \underbrace{(R + B^T P_{t+1} B)}_{M_2} a + 2\underbrace{s^T A^T P_{t+1} B}_{M_3} a + \underbrace{\sigma^2 tr(P_{t+1}) + p_{t+1}}_{c}$$

$$\pi_t^*(s) = \arg\min_a Q_t^*(s,a)$$

$M_1, M_2$ symetric $\quad D_a(w^T a) = w$

$$\rightarrow Q(s,a) = \underbrace{s^T M_1 s} + a^T M_2 a + 2\underbrace{s^T M_3} a + c$$

minimum occurs $\nabla_a Q(s,a)$

$$\nabla_a Q(s,a) = 0 + 2M_2 a + 2M_3^T s + 0 = 0$$

$$\pi_t^*(s) = -\underbrace{(R + B^T P_{t+1} B)^{-1} B^T P_{t+1} A}_{K_t^*} s \qquad a^* = -M_2^{-1} M_3^T s$$

$$(AB)^T = B^T A^T$$

$$(x^T M x) = x^T \underbrace{\left(\frac{M}{2} + \frac{M^T}{2}\right)} x$$

$$= (x^T M x)^T = x M^T x$$

$$V_t^*(s) = Q_t^* = (s, \pi^*(s)) \qquad Q(s,a^*) = s^T(M_1 - M_3 M_2^{-1} M_3^T) s + c$$