

3.2 Test2

The current problem is formulated as

$$\min_{u,x} \quad \frac{\gamma^N}{2} x_N^T P_N x_N + \sum_{k=0}^{N-1} (f^t[x_k, u_k]^T) + \sum_{k=0}^{N-1} (\frac{1}{2} \gamma^k (x_k^T Q_{rl} x_k + u_k^2 R_{rl})) \quad (13)$$

$$s.t \quad (14)$$

$$x_{k+1} = x_k + \frac{T}{N} (Ax_k + Bu_k + E), \quad k \in [0, N-1] \quad (15)$$

$$x_0 = s \quad (16)$$

The real system dynamics is stated as

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -1 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \epsilon \quad \epsilon \approx N(1, 1) \quad (17)$$

The RL-algorithms acts upon 10 parameters, $\theta = [q_1, q_2, q_3, q_4, a_1, a_2, a_3, a_4, e1, e2]$, that is the elements of the MPCmodel A and E, and the quadratic stage cost matrix Q. The RL-algorithm is as follows

$$L = x_s^T x_s + u_s^T u_s \quad (18)$$

$$TD = L + \gamma V_{s+1} - V_s \quad (19)$$

$$\theta_{s+1} = \theta_s + \alpha TD \nabla_{\theta} \mathcal{L} \quad (20)$$

The sequence of operations is

$$x_s, u_s, \theta_s, V_s, \nabla_{\theta, s} \mathcal{L} \xrightarrow{\text{simulation}} x_{s+1} \xrightarrow{\text{optimization}} u_{s+1}, V_{s+1}, \nabla_{\theta, s+1} \mathcal{L} \quad (21)$$

$$\downarrow \quad (22)$$

$$x_{s+1}, u_{s+1}, \theta_{s+1}, V_{s+1}, \nabla_{\theta, s+1} \mathcal{L} \xleftarrow{\text{progression}} (L(x_s, u_s) + \gamma V_{s+1} - V_s) \alpha + \theta_s = \theta_{s+1} \xleftarrow{\text{RL-update}} \quad (23)$$

$V(s)$ is defined as the solution to the problem of 13 - 16, and $Q(s, a)$ is defined as the solution to the same problem 13 - 16, only with the added constraint $u_0 = a$. Since the input a is deterministic, then the solution to both of these problems becomes the same, and for simplicity, $V(s)$ is used instead of $Q(s, a)$

MPCParameters:

The MPC model is simulated by discrete eulersmethod, where the dt is given by the prediction horizon N and time horizon T . The vector f is set to zero, and the discount is set to $\gamma = 0.9$. A and B is initialized as the real system dynamics, and E is initialized as zero. In addition, $T = 1$ & $N = 5$.

Other Parameters:

The system as a whole is simulated by 4th order runge-kutte with timestep of 0.01. The RL-algorithm has parameters $\gamma = 0.9, \alpha = 0.0001$

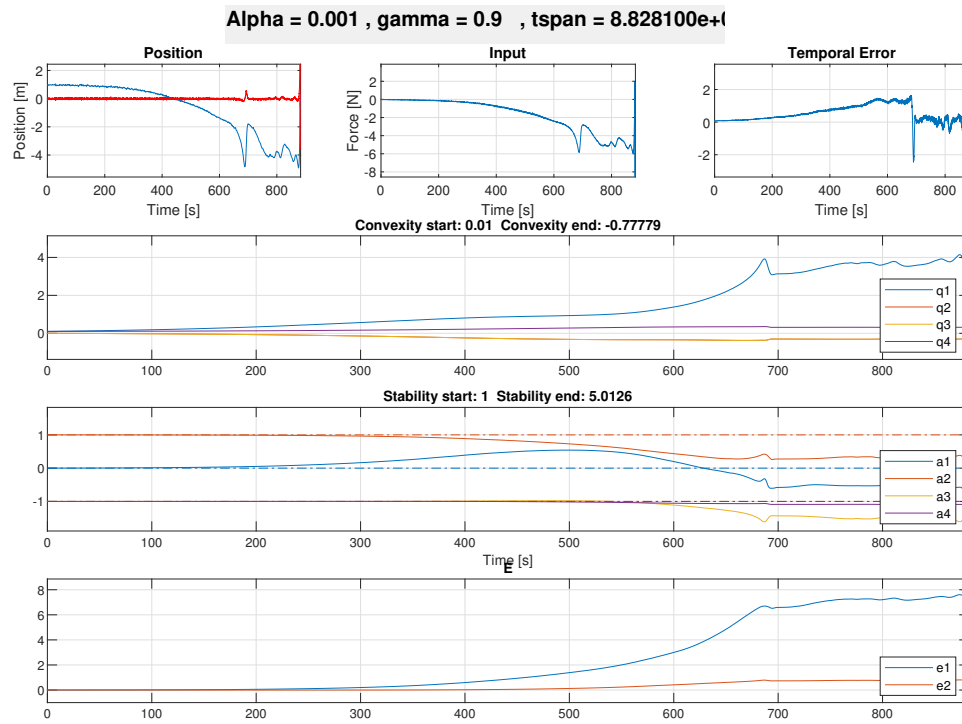


Figure 1: Caption

4 Discussion

5 Results

6 Known Problems