Output Regulation of Unknown Linear Systems using RL

Farnaz Adib Yaghmaie (farnaz.adib.yaghmaie@liu.se)

Problem Formulation

Output regulation (OR) is defined as designing the controller

- to tacks a reference trajectory and
- to reject the effect of a disturbance with known model.

Usually, the dynamics of the reference and the disturbance are combined into a single dynamic system called *exo-system* in the literature. Consider the system as

$$\dot{x} = Ax + Bu + Dv,$$
$$y = Cx,$$

where $x \in \mathbb{R}^n$, $u \in \mathbb{R}^m$, $y \in \mathbb{R}^p$ denote the state, the control and the output of the system. Define the exo-system as

$$\dot{v} = Sv,$$
$$w = Fv,$$

where $v \in \mathbb{R}^q$, $w \in \mathbb{R}^p$. It is desired to design the controller u such that the output regulation error, denoted by $e \in \mathbb{R}^p$, converges to zero

$$e = y - w = Cx - Fv \to 0.$$

Theorem (Standard Solution to OR)

Assume that (A, B) is stabilizable and $Re(\lambda) \leq 0, \ \forall \lambda \in Spec(S)$. Assume that K_{fb} is selected such that $A + BK_{fb}$ is strictly stable. Then, the controller

$$u = K_{fb}x + (\Gamma - K_{fb}\Pi)v$$

solves the output regulation problem if and only if

$$\Pi S = A\Pi + B\Gamma + D,$$
$$C\Pi - F = 0.$$

have a solution.

Standard solution

- is model-based.
- needs offline computation.
- is not optimal.

Optimal OR

Introduce the quadratic performance index to be minimized

$$V_a(X(t), u(t)) = \lim_{T\to\infty} \frac{1}{T} \int_t^{t+T} e^\dagger Q e + u^\dagger R u \, d\tau. \quad (1)$$

The following result is useful in designing a reinforcement learning algorithm.

Lemma

Minimizing (1) is equivalent to minimizing

$$V_\infty(X(t), u(t)) = \int_t^{+\infty} (r(X(\tau), u(\tau)) - \lambda^*) \, d\tau \quad (2)$$

in the sense that an optimal control to (2) is also optimal for (1).

Model-Free Reinforcement Learning (RL) Algorithm

We apply a behavioral policy $u = u^{(k)} + e$ while learning a sequence of $u^{(k)}$ converging to the optimal controller u^*.

Model-free RL Algorithm

1: **Initialize:** $u^{(0)} = K^{(0)}X$, $k = 0$.
2: **repeat**
3: Apply $u = K^{(k)}X + e$ and collect the required information at N sample times.
4: Find $P^{(k)}$, $\lambda^{(k)}$, $K^{(k+1)}$ from

$$X^\dagger(t)P^{(k)}X(t) = \int_t^{t+\delta t} (e^\dagger Q e + u^{(k)\dagger} R u^{(k)}) \, d\tau + X^\dagger(t+\delta t)P^{(k)}X(t+\delta t)) + 2\int_t^{t+\delta t} (u - u^{(k)})^\dagger R K^{(k+1)} X \, d\tau - \delta\lambda^{(k)}.$$

5: **until** Convergence

Simulation Result

Output regulation for an F16 aircraft system

$$A = \begin{bmatrix} -1.01887 & 0.90506 & -0.00215 \\ 0.82225 & -1.07741 & -0.17555 \\ 0 & 0 & -1 \end{bmatrix}, B = \begin{bmatrix} 0 \\ 0 \\ 5 \end{bmatrix}, D = \begin{bmatrix} 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix},$$

$$S = \begin{bmatrix} 0 & 1 \\ -0.01 & 0 \end{bmatrix}, C = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}, F = \begin{bmatrix} 1 & 0 \end{bmatrix}.$$

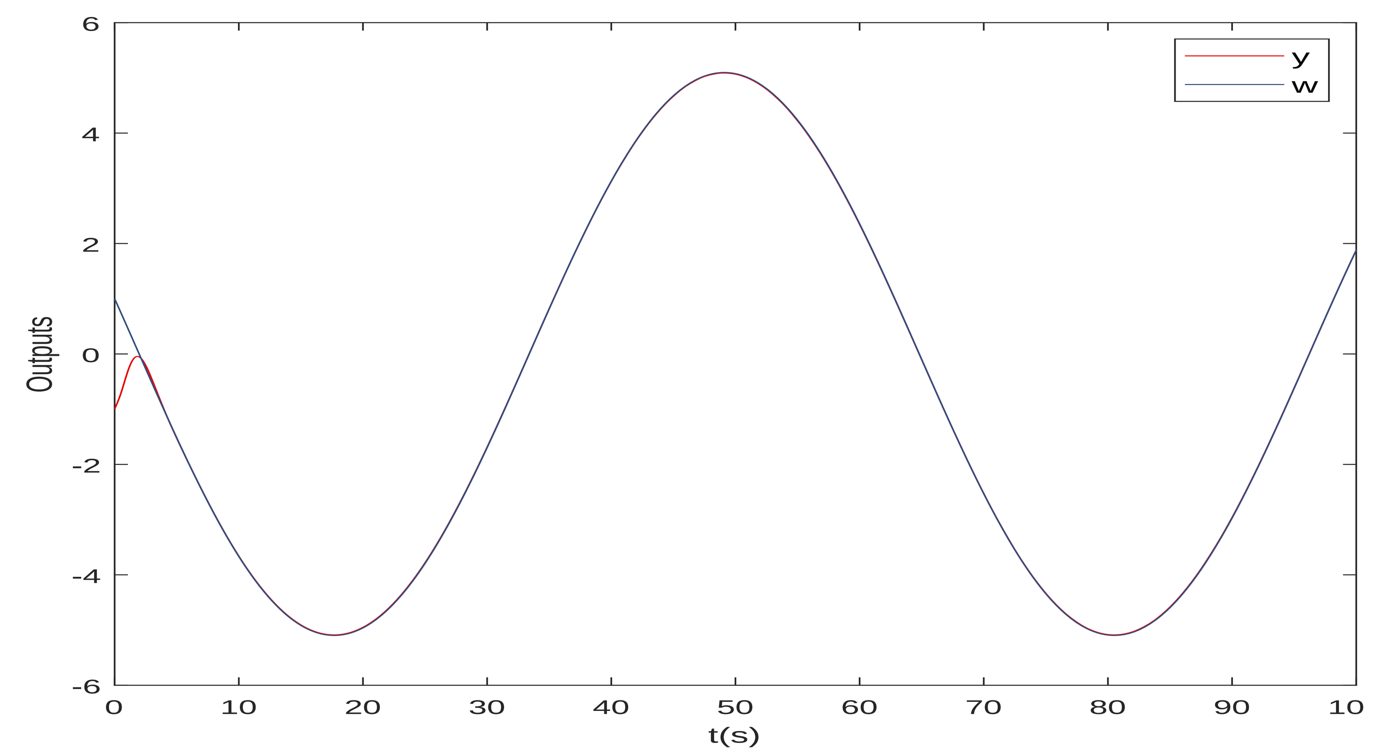
The weights are $Q = 100, \ R = 1$.

Analytical method

$vecs(P^*) =$	[50.614	17.264	-1.005	-72.609	-7.587
	9.118	-0.681	-29.408	-5.714	0.067
	1.995	0.755	◇	◇	◇]
$K^* =$	[5.025	3.406	-0.332	-9.998	-3.738]

Model-free Algorithm

$vecs(P^{(9)}) =$	[50.614	17.264	-1.005	-72.609	-7.587
	9.118	-0.681	-29.408	-5.714	0.067
	1.995	0.755	110.239	19.956	-1.102]
$K^{(9)} =$	[5.025	3.406	-0.332	-9.975	-3.774]



Conclusions

- Completely model-free.
- Solution is online and optimal.

LINKÖPING UNIVERSITY LINK-SIC

LINKÖPING UNIVERSITY
Division of Automatic Control