Initialization `init.m`	
`pars`	Parameter structure for simulation environment
`Tsim`	Total simulation time for one run
`Nruns`	Number of runs

Key fields in `pars`	
`estBufferSize`	Buffer size for model estimator
`optCtrlMode`	Optimal controller mode
`ctrlStackSize`	N
`criticStackSize`	M
`rcostS, rcostR`	S, R

`optCtrlMode`	
1, 2 - model-predictive control (MPC)	$J\left(y_1, \{u\}_1^N\right) = \sum_{k=1}^N r(y_k, u_k)$
3, 4 - RL/ADP via stacked Q-learning	$J\left(y_1, \{u\}_1^N\right) = \sum_{k=1}^N \hat{Q}(y_k, u_k)$
5, 6 - RL/ADP via N roll-outs of r	$J\left(y_1, \{u\}_1^N\right) = \sum_{k=1}^{N-1} r(y_k, u_k) + \hat{Q}(y_N, u_N)$

Modes 1, 3, 5 use true model (f, h) for prediction
Modes 2, 4, 6 use an a state-space model estimated online

Legend	
`[•,•]`	a matrix
`struct`	a structure
r	running cost
N	control horizon
M	critic stack size
J	controller cost function
J_c	critic cost function
$Q, \hat{Q}$	Q-function and its approximate
e	temporal difference
W, W^-	current and previous critic weights
R, S	matrices in $r(y,u) = y^\top S y + u^\top R u$
γ	discounting factor
L	number of critic weights
`dt`	controller sampling time

Sample with step `dt`

System		
Signal	**Info**	**Type**
In	Control u	`[1, 1]`
Out	System output y	`[p, 1]`
Par.	`pars`	`struct`

$$\begin{aligned} \dot{x} &= f(x,u) \\ y &= h(x) \end{aligned}$$

f, g are described by `sys_rhs, sys_out` resp.

Nominal controller		
Signal	**Info**	**Type**
In	System output y	`[p, 1]`
Out	Control u	`[1, 1]`
Par.	`pars`	`struct`

$$u = \kappa(y)$$

This controller is used as a baseline for comparison with the optimal one

Optimal controller `dataDrivOptCtrl`		
Signal	**Info**	**Type**
In	System output y	`[p, 1]`
Out	Control u	`[1, 1]`
Par.	`pars`	`struct`

Buffer of size `estBufferSize`

Buffer of size `estBufferSize`

Model estimator `mySSest.m`		
Signal	**Info**	**Type**
In	Buffers of previous controls and measured outputs `us, us`	`[1, estBufferSize]` `[p, estBufferSize]`
In	Sampling time `dt`	scalar > 0
In	`modelOrder`	natural
In	Initial state-space model `(Ainit, Binit, Cinit, Dinit)`	`[modelOrder,modelOrder]` `[modelOrder,l]` `[p,modelOrder,]` `[p,l]`
Out	Par. of estimated state-space model (A, B, C, D) of desired order `modelOrder` and initial state `x0`	`[modelOrder,modelOrder]` `[modelOrder,l]` `[p,modelOrder,]` `[p,l]` `[modelOrder, 1]`

Fits a state-space model

$$\begin{aligned} \hat{x}^+ &= A\hat{x} + Bu \\ y^+ &= C\hat{x} + Du, \end{aligned}$$

from data `us, ys` that are organized in buffers of previous controls and measured outputs

Buffer of size M

Buffer of size M

Critic `critic.m`		
Signal	**Info**	**Type**
In	Buffers of previous controls and measured outputs `U, Y`	`[l, M]` `[p, M]`
In	Running cost par. S, R	`[p, p]` `[l, l]`
In	Discounting factor γ	scalar in $(0, 1]$
Out	Critic weights and cost W, J_c	`[L, 1]` scalar

Critic solves the problem

$$\min_W J_c = \frac{1}{2} \sum_{k=1}^M e^2$$

where $e = W^\top \varphi(y^-, u^-) - \gamma W^- \varphi(y, u) - r(y^-, u^-)$ and $(y, u), (y^-, u^-)$ are the current and, resp., previous output and control taken from the buffers `U, Y`.
The Q-function approximate is described by

$$\hat{Q}(y, u) = W^\top \varphi(y, u)$$

Actor `optCtrl.m`		
Signal	**Info**	**Type**
In	System output y	`[p, 1]`
In	Running cost par. S, R	`[p, p]` `[l, l]`
In	Critic weights W	`[l, 1]`
In	Sampling time `dt`	scalar > 0
In	Par. of estimated state-space model (A, B, C, D) of desired order `modelOrder` and initial state `x0`	`[modelOrder,modelOrder]` `[modelOrder,l]` `[p,modelOrder,]` `[p,l]`
Out	Control u and actor cost function J	`[l, 1]` scalar

Actor solves the problem

$$\begin{aligned} \min_U \quad & J(y, U) \\ \text{s.t.} \quad & y^+ = \mathcal{M}(\bullet, u) \end{aligned}$$

w.r.t. to a **sequence** of controls $U = \{u\}_1^N$ over a horizon of N (see definition of J in `optCtrlMode`) starting with the current output y. Here, $\mathcal{M}$ is a model that predicts each next output from the current one along with the control.
According to the mode `optCtrlMode`, $\mathcal{M}$ can be either (a discretized) version of (f, h), i. e.,

$$\begin{aligned} x^+ &= x + \delta f(x, u) \\ y^+ &= h(x), \end{aligned}$$

where δ is `dt;` or an estimated state-space model

$$\begin{aligned} \hat{x}^+ &= A\hat{x} + Bu \\ y^+ &= C\hat{x} + Du, \end{aligned}$$