REINFORCEMENT LEARNING ALGORITHM FOR CONTROLLING THE TRANSIENT VIBRATIONS OF SEMI-ACTIVE STRUCTURES INDUCED BY UNKNOWN PERIODIC EXCITATION

D. Pisarski and Ł. Jankowski

Institute of Fundamental Technological Research of the Polish Academy of Sciences, Warsaw, Poland

e-mail: dpisar@ippt.pan.pl

1. Problem statement

The role of the adaptive control in the structural vibration problems is to guarantee a relevant adjustment of the control decisions as the environmental conditions or internal structural parameters are unknown or subject to changes. A typical approach to adapting the control functions is based on the model predictive control which employs repetitive solutions to an optimal control problem. This method exhibits high efficiency, nevertheless, due to the high computational burden, it is mostly dedicated to the applications in linear structural control problems (actively controlled structures) [2, 3]. To cope with the higher complexity of the dynamical systems concerned with semi-actively controlled structures, it is relevant to consider using iterative learning algorithms such as reinforcement learning (RL). Reinforcement learning control does not require complete information of the system [1] and the computational complexity concerned with successive iterations of the control policy is significantly lower than in the case of searching for the optimal solutions. In this work, we study the reinforcement learning actor-only algorithm for the problem of semi-active vibration control. The aim is to exploit the convergence property which is naturally inherited from the incorporated gradient descent method and to examine the stabilizing performance when compared to competitive control strategies.

We consider a class of semi-active vibrating structures described by the bilinear dynamical equation:

$$\dot{x}(t) = A\,x(t) + \sum_{j=1}^{m} u_j\,B_j\,x(t) + F(t)\,,\ \ x(0) = x_0\,. \tag{1}$$

In (1), x, $u_1, ..., u_m$, and A, $B_1, ..., B_m$ stands respectively for the state vector, control functions, and constant matrices. The vector F stands for a repetitive short-time excitation defined be a periodic function with unknown amplitude and frequency (concerned for example with drilling processes). The aim is to design the control functions $u_j \in [u_{min}, u_{max}]$, $j = 1, ..., m$ (later referred to as the policy functions), and reinforcement learning algorithm that allows these functions to adjust their parameters (policy parameters) guaranteeing the best stabilization of the transient vibrations of the system (1) induced by the excitation F.

2. Design of the RL control

For the considered process time $t \in [0, T]$, for (1) we assume the switching state-feedback policy functions $u_1, ..., u_m$ defined as:

$$u_j(t) = \begin{cases} u_{min}, & x^T(t)K_j(t)x(t) \geq 0, \\ u_{max}, & x^T(t)K_j(t)x(t) < 0. \end{cases} \tag{2}$$

In (2), K_j is the policy parameter and structured by two sub-parameters. The first one is K_j^*, iterated using the reinforcement learning actor-only algorithm for the time window $t \in [0, T_L]$ and then assumed for the time $t \in [0, T_F]$ ($T_F \geq T_L$) where the excitation force $F \neq 0$ is acting on a structure. The second sub-parameter is K_j^0, assumed for the free vibration for $t \in (T_F, T]$, i.e. when $F = 0$, and predefined based on the solution to the Lyapunov equation to guarantee the asymptotic stability of the autonomous form of the system (1) (see [5]). The aim of reinforcement learning is to minimize the cost functional J, defined as the integral of the structure's energy computed for the learning time window $t \in [0, T_L]$:

$$(3) \qquad J(T_L) = \frac{1}{2} \int_0^{T_L} x^T(t) Q x(t)\, \mathrm{d}t\,, \quad Q \succ 0\,.$$

For updating the policy parameter K_j^*, the reinforcement learning algorithm will use the following updating sequence and derivative of the cost functional:

$$(4) \qquad K_j^{*+} = K_j^* - \alpha_j \left(\frac{\mathrm{d}J}{\mathrm{d}K_j^*} \right)_{|K_j^*}\,, \quad \frac{\mathrm{d}J}{\mathrm{d}K_j^*} = (u_{max} - u_{min}) \sum_{i=1}^{s} p^T(\tau_i)\, B_j\, x(\tau_i)\, x(\tau_i)\, x^T(\tau_i)\,.$$

In (4), p and $\{\tau_i\}_{i=1,\ldots,s}$ denotes respectively the adjoint state associated to the cost functional (3) and the sequence of time instants $t \in [0, T_L]$ when $x^T(t) K_j^* x(t) = 0$. The sequence in (4) is initialized with K_j^0. The step size $\alpha_j > 0$ is selected by trial and error, to assure substantial decreases of the cost functional value.

3. Case study and results

We study an aluminum beam structure characterized by length $L = 1$ [m], moment of inertia $I = 0.7210^{-10}$[m^4], mass per unit length $\mu = 0.2$[kg/m]), and simply supported at both ends with the additional support of the semi-active device mounted at the position $0.4L$. The structure is subjected to the periodic force acting at the location $0.6L$. For the semi-active supports, we assume the elastic and damping forces that they generate depend linearly on the control variable u. Using this assumption and the finite element model for the beam structure, we eventually represent the considered system in the form of (1). To perform the reinforcement learning algorithm based on (4), we repeat 500 excitations of the duration $T_F = 0.2$[s], frequency 25[Hz], and amplitude 200[N]. The scenario is repeated three times assuming the learning time windows $T_L = 0.4T_F$, $T_L = 0.75T_F$, and $T_L = T_F$. After the learning process, we compare the established RL control to the optimal, passive, and heuristic [4] strategies. Summarizing the simulation results (see Table 1), we conclude that the RL control in the cases of $T_L = 0.4T_F$ and $T_L = 0.75T_F$ provides suboptimal performance. The observed minor drop of the performance in the case of $T_L = T_F$ results from a slower convergence of the update sequence (4) that is concerned with the increased difficulty in the selection of the step size α. The observations remain intact when for the learning process we include relatively small random perturbations of the excitation frequency (up to 20%).

Passive strategy	RL control			Heuristic control	Optimal control
	$T_L = 0.4T_F$	$T_L = 0.75T_F$	$T_L = T_F$		
1.0000	0.6416	0.6412	0.6956	0.8186	0.6340

Table 1: Cost functional J values for the considered control cases (normalized to the passive case).

Acknowledgments This research has been supported by the National Science Centre, Poland under grant agreements 2020/39/B/ST8/02615 and DEC-2017/26/D/ST8/00883.

References

[1] C. Dengler C and B. Lohmann. Actor-critic reinforcement learning for the feedback control of a swinging chain. *IFAC Papers*, 51:378–383, 2018.

[2] G. Takacs G and B Rohal-Ilkiv. Model predictive control algorithms for active vibration control: a study on timing, performance and implementation properties. *J Vib Control*, 20(13):2061–2080, 2013.

[3] A. Oveisi, M. Hosseini-Pishrobat M, T Nestorovic, and J. Keighobadi. Observer-based repetitive model predictive control in active vibration suppression. *Struct Control Health Monit*, 25(5):e2149–1–23, 2018.

[4] D. Pisarski. Decentralized stabilization of semi-active vibrating structures. *Mech Syst Signal Process*, 100:694–705, 2018.

[5] D. Pisarski, T. Szmidt, and R. Konowrocki. Decentralized semi-active structural vibration control based on optimal system modelling. *Struct Control Health Monit*, 27(11):e2624–1–20, 2020.