AVEC'96
International Symposium on Advanced Vehicle Control
at the Aachen University of Technolgy, 24th - 28th of June, 1996

The Application of Interconnected Stochastic Learning Automata to Controller Design for Semi-Active Automobile Suspensions

M. N. HOWELL
G. P. FROST
T. J. GORDON

Department of Aeronautical and Automotive Engineering and Transport Studies
Loughborough University of Technology

ABSTRACT

Interconnected stochastic learning automata are applied to the on-line optimal control of a semi-active suspension system. The control objective is to minimise vertical body acceleration and thereby improve passenger ride comfort. This is a difficult problem due to the stochastic nature of the unknown road input compounded by sensor noise, damper non-linearities and interaction with other vehicle components such as the engine on its mounts. Rig tests conducted on a four wheeled passenger vehicle fitted with high bandwidth continuously variable suspension dampers demonstrate the suitability of this approach. A significant improvement in ride quality was achieved after just a few hours of learning.

1. INTRODUCTION

Promoting passenger ride comfort by isolating the passenger compartment (sprung mass) from the vibrational effects of road unevenness is one of the principal requirements of a vehicle suspension system. Considerable improvements in vehicle ride performance over a traditional passively operated system can potentially be achieved through the use of active and semi-active vehicle suspension systems /5/. A semi-active suspension system can achieve some of the benefits of a full-active system without the need for expensive hydraulics that a full-active system requires and thus offers a good compromise between the limitations of passive damping and the added expense and reliability problems of a full-active system.

The traditional approach to the synthesis of a suitable controller for such a system is a difficult and time-consuming process. Complex systems modelling techniques are required to take account of the non-linearities caused by the nature of the semi-active dampers, suspension geometry and friction. Vibrational modes involving other parts of the vehicle, such as the engine on its mounts, should be accounted for and bandwidth limitations and electrical noise should be considered. Ideally the effects of various disturbance inputs also need to be investigated to ensure the system can cope with a wide range of different operating conditions.

Machine learning and intelligent control techniques offer an alternative approach to the synthesis of controllers for complex systems. These techniques are directed at the on-line learning of the controller parameters, with the capability of adapting to changing conditions and with robustness to high levels of noise. On-line learning has the potential to completely eliminate modelling errors, the resulting controller being ideally suited to the particular hardware and operating conditions.

2. REINFORCEMENT LEARNING BY AUTOMATA

Reinforcement learning can be applied to the problem of the optimal control of an incompletely known system, where only indirect information about the system behaviour is available. The approach is based on the common-sense principle that if a certain action causes an improved performance is some sense, then the tendency to produce that action in the future should be increased. Actions are therefore tested in the relevant environment, using a selection process which 'rewards' the actions that turn out to be the most effective.

Other intelligent control techniques, such as, supervised learning rely on the prior knowledge of output and this information is used during a training phase. This is primarily an off-line process and for many control tasks it is difficult to define the desired outputs in advance. In contrast, reinforcement learning techniques only require a signal which relates to the quality of control achieved and can therefore be applied directly to real time on-line adaptive optimal control.

Learning automata are a form of reinforcement learning technique which has its roots in mathematical psychology and animal learning /4/. Learning automata interact with an environment through a set of possible actions which are selected stochastically. The environment returns a 'teaching' signal to the learning automaton which indicates the effect of the selected action on the environment. Each action has a selection probability associated with it, and this is altered by means of a learning algorithm. This improves the chances of selection of actions that lead to a response which can be considered favourable.

Learning automata have been shown to be suitable for learning optimal control /7/. They are particularly suited to the control of stochastic processes and have previously been applied to the controller design of active suspension systems /2,3/. In this earlier work a single learning automata was used to obtain all the parameters of the controller for a single wheel-station, using a simple quarter car vehicle model. In keeping with the concept of learning automata as a practical on-line technique, in this paper a real test vehicle has been used and interconnected learning automata applied to the control of the semi-active suspension system.

3. INTERCONNECTED LEARNING AUTOMATA

A single learning automaton can be considered as a building block for more complex structures, where several automata operate in a distributed, decentralised fashion within a co-ordinated intelligent control system. The benefits of this modular approach are that a greater flexibility in the controller structure can be achieved, and shorter learning times are possible. Interconnected learning automata /6/ provide one such structure. Here in fact, the automata are only connected through the dynamics of the environment and through a shared control objective; each learning automaton has no knowledge that it is working in a team - see *Figure 1*.

Suppose a single automaton acts on its environment by setting each of m action variables to one of r possible discrete values. There are then r^m different choices of overall action available, so if r and m are reasonably large the number of probabilities to be stored becomes overwhelming. Individual probabilities are initially very small and it actually takes a very long time for progress to be made by the reinforcement learning algorithm. A team of interconnected automata can be used to overcome this dimensionality problem. If each of the automata is assigned one of the action variables, then the number of probabilities to be updated is only $r \times m$, so the number of actions for an individual automaton is reduced while maintaining the overall high dimensionality.

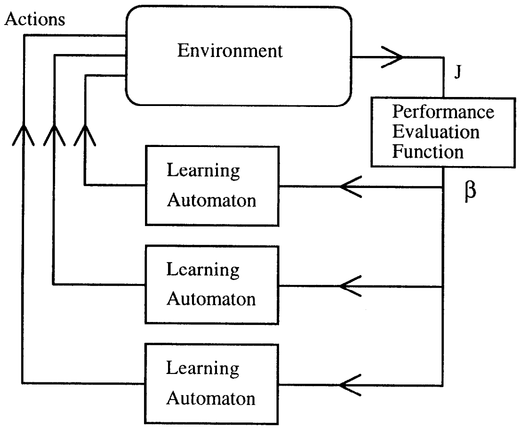

Figure 1.

4. LEARNING A FULL-VEHICLE SUSPENSION CONTROLLER

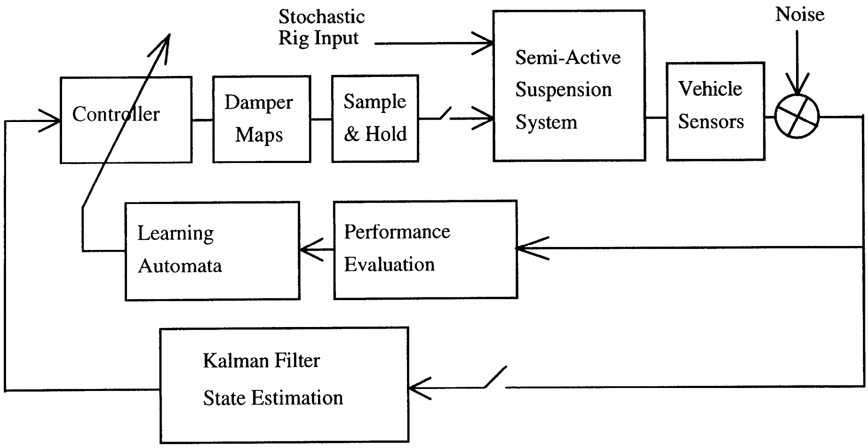

Figure 2.

Learning automata are to be used to minimise the body bounce accelerations on an experimental vehicle by adjusting the controller parameters of the suspension system, as represented in *Figure 2*. In this case, controller feedback gains K are identified as the action variables, and the environment is taken as the vehicle system, including sensors and actuators, together with the random excitation process, in this case delivered by a four-poster road simulator rig.

A team of four interconnected learning automata, one for each wheel-station, are operated in a synchronous manner. At every iteration, each automaton selects a set of controller gains from only a finite set of different gain vectors. Each automaton has a limited action set of 27 different gain vectors available, which is obtained by quantising three feedback gains into three levels. Each of the actions has a corresponding probability distribution

$$P = \{p_i \ : \ i = 1,2,\ldots,27\} \tag{1}$$

thus an overall space of $27^4 \cong 5 \times 10^5$ actions is described by a reduced set of $27 \times 4 = 108$ probabilities.

As a slight variation on the interconnected automata structure of *Figure 1*, four cost functions are evaluated on-line with mean square bounce accelerations at each suspension mounting point stored as a reference set R which is used to define the success criteria for subsequent actions. Let $a = 1,2,3,4$ denote the four corners of the vehicle; the corresponding cost at each iteration for corner a is given by

$$J^a(k) = \frac{1}{\Delta}\sum_{t-\Delta}^{t}\left(\dot{v}_b^a(t)\right)^2 \qquad (2)$$

where $\dot{v}_b^a$ is the vehicle body acceleration and Δ is the time over which each set of fixed controller parameters are evaluated on the vehicle. Each automaton uses the corresponding cost as its basis for action reinforcement.

The probability distribution, for each automaton, is updated according to a reward-inaction scheme

$$\left.\begin{aligned} p_i(k+1) &= p_i(k) + \theta\beta(k)(1-p_i(k)) \\ p_j(k+1) &= p_j(k) - \theta\beta(k)p_j(k) \end{aligned}\right\}(j \neq i) \qquad (3)$$

where a reinforcement index $\beta \in [0,1]$, corresponding to an S-model type learning automata /4/, is defined by

$$\beta(k) = \max\left\{0, \frac{J^{a}_{\text{med}} - J^a(k)}{J^a_{\text{med}} - J^a_{\text{min}}}\right\} \qquad (4)$$

and is used as a measure of success for any particular action chosen.

Decentralised control is achieved by having each of the four automata determining parameters for a single damper controller and using a localised reinforcement signal. The design of the car gives a natural division of responsibility for the automata, with the cost associated with each automaton being primarily influenced by the corresponding corner dynamics. A level of interaction will occur between wheel-stations via the body dynamics, but the subdivision into a four corner team of automata reduces the number of probabilities with the likely beneficial effect of increasing convergence speed. The arrangement can be viewed as a decentralised co-operative game between the automata where they are all aiming to reduce the overall body acceleration affects. Though the number of automata could be increased further by allowing a single automaton to determine each individual controller parameter, this might lead to problems with local minima.

In order to further reduce the number of actions available to each automaton, the learning was conducted in stages. At each learning stage, a finite set of actions is made available for each automaton. When the probability of a single action exceeds a pre-assigned threshold η (in this study $\eta = 0.5$) the action space for that automaton is re-defined. The controller parameters are re-quantised about the successful gain vector, with scales reduced by a factor λ (in this study $\lambda = 0.4$). In this way, the search region is successively reduced until the parameter ranges are sufficiently small for learning to be effectively complete /2/. The detailed algorithm for each learning automaton is given in the Appendix.

5. EXPERIMENTAL IMPLEMENTATION

The vehicle was mounted on a hydraulic four-poster road simulator rig with each corner being driven independently. Each rig actuator had a displacement spectral density function $S(f) = c \cdot f^{-2.5}$ with $c = 2.18 \times 10^{-4}$ and was limited to the frequency range 0.2 to 20Hz. The vehicle, a medium sized saloon car, was fitted with semi-active dampers which can continuously vary the amount of energy dissipation in response to a control signal. Sensors to measure wheel acceleration, body acceleration and suspension deflection were fitted to each corner of the car and a previously designed Kalman filter /1/ was used to provide state estimates for feedback control. The Kalman filter designs are based on quarter car models of the car but provide adequate state feedback signals. The following variables were used for feedback: suspension deflection (x_s) wheel velocity (v_w) and body velocity (v_b) at each corner of the vehicle.

As mentioned above, the objective of the learning system was to minimise the mean square vertical body acceleration at each of the four suspension mounting points; each corner of the car had a single learning automaton which determined a three-parameter control law for the corresponding damper. The required damper force F_s^a at corner a is given by

$$F_s^a = k_1^a x_s^a + k_2^a v_w^a + k_3^a v_b^a \qquad (5)$$

with the learning system having to determine the controller parameters. The desired force was used together with the velocity difference across the damper to determine the required control setting; this was calculated from a non-linear map of the damper characteristics. The initial quantisation for each controller parameter was

$$k_1^a = [0, 7500, 15000]$$

$$k_2^a = [0,1000,2000]$$
$$k_3^a = [-4000,-2000,0] \quad (6)$$

with re-quantisation as defined in the previous section.

Though each automaton was configured in a decentralised structure, all the automata were operating in a synchronised manner, so that the gain values at each corner were constant over the same 16 second learning iteration. The learning algorithms were implemented in 'C' on a 486DX2-50 PC with the controller gains being sent to a TMS320 digital signal processor operating at 500 Hz. This provided the state estimation, performed the closed loop control, and returned the dynamic cost.

6. CONTROLLER PERFORMANCE

Table 1 shows the results obtained from three independent trials of the learning algorithms methodology, with each test using a different randomised rig input. The trials were run for approximately 1000 iterations, representing less than 5 hours of rig time. All the automata achieved two stages of convergence and would probably have converged further had more rig time been allocated. Results from firm and soft damper settings are also shown in the table for comparison.

Figure 3 shows a typical reduction in the cost that is achieved during learning, the occasional rise coinciding with the re-quantising of the learning automata. The learning automata are shown to be capable of learning despite a high level of noise. *Figure 4* shows a power spectral density plot of body acceleration for a learnt controller compared with a nominal passive damper setting. The passive damper setting is equivalent to the vehicle being fitted with standard production dampers. The learnt controller has improved the system response at the body bounce frequency, although it has performed slightly worse at other frequencies. A skyhook law of the form

$$F_s^a = -k_{\text{sky}} v_b^a \quad (7)$$

where $k_{\text{sky}} = 2000$ was also used as a comparison of the controllers performance, the value of the skyhook damping rate being obtained by informally optimising over a test section of road. The rms body acceleration is lower overall for the learnt controller parameters (1.277) than for the nominal passive damper setting (1.430) and the skyhook damping control law (1.380) demonstrating the capability of the learning methodology.

	Corner A	Corner B	Corner C	Corner D
Firm	-49.1734	-39.8324	-46.6319	-32.8596
Soft	2.0767	-5.1710	-10.4914	-0.4073
LA Controller 1	7.7817	6.0137	2.5103	8.5991
LA Controller 2	7.6008	6.2334	4.7675	8.2051
LA Controller 3	6.6475	4.7039	4.8640	6.5046

Table 1. Percentage Improvement in Body Acceleration [Comparison with a nominal passive damper setting. Negative values denote increased body acceleration].

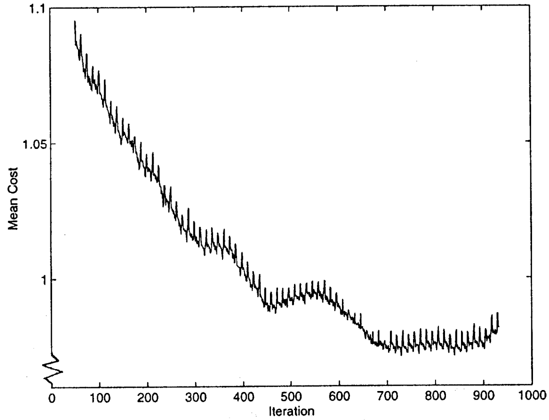

Figure 3. Typical Cost Reduction Over Time

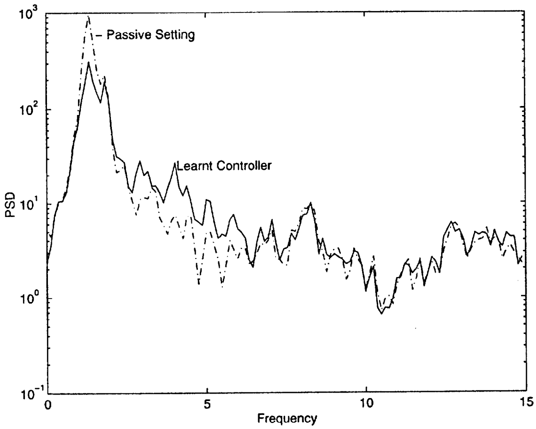

Figure 4. A Comparison of the Power Spectral Density Functions

7. DISCUSSION AND CONCLUSIONS

Learning automata have been shown to operate successfully in a complex and highly stochastic environment. The algorithm has relatively few free parameters, can be used in situations where the problem is ill defined, and can eliminate the need for systems modelling. The learnt controller produced a significant improvement in rms body acceleration in under 5 hours of rig time this represents a considerable improvement over conventional design techniques.

The learning automata methodology uses a 'divide and conquer strategy' to reduce the parameter space dimensions. The effect of each action (gain vector) in the environment has to be assessed and there is a trade off between exploring new areas of the action space and exploiting the information already gained. Learning automata reduce the amount of exploration over time by increasing the probability of selecting those actions that perform better.

The interconnected learning automata discussed in this paper used a four corner team of automata, using primarily localised information. This structure has the benefit of reducing the number of probabilities for updating at each iteration and this reduction is further enhanced by the scheme for re-quantising gain vectors. This study has demonstrated the capability of using learning automata in a co-ordinated decentralised design.

Work is currently underway to extend the operation of the continuous learning automata to perform the state estimation functions and eliminate the need for a Kalman filter. It is also anticipated that road based learning should be feasible.

ACKNOWLEDGEMENTS

This research is supported by a grant (No. GR/J82652) from the Engineering and Physical Science Research Council. It is also in collaboration with the Chassis Systems Integration Department, part of the European Chassis Advanced Vehicle Technology Group within the Ford Motor Company.

REFERENCES

/1/ Best, M.C., Gordon, T.J. Crawford I.L, Hand P., 1994, Real Time Estimation of Vehicle Suspension Characteristics Using Kalman Filtering, XXV FISITA Congress, 17-21 Oct. 1994, Beijing.

/2/ Frost, G.P., Gordon, T.J., and Wu, Q.H., 1994, The Application of Learning Automata to Advanced Vehicle Suspension Control, IUTAM Symposium, the active control of vibration, 5-8 Sept. 1994, Bath.

/3/ Gordon, T.J., March, C., and Wu, Q.H., 1993, Stochastic Optimal Control of Active Vehicle Suspensions Using Learning Automata, Proc. I.Mech.E. Part I, 1993, 207, pp143 - 152 .

/4/ Najim, k., and Posnyak, A.S., 1994, Learning Automata; Theory and Applications, 1994, Pergamon

/5/ Sharp, R.S., and Hassan, S.A., 1986, The Relative Performance Capabilities of Passive, Active and Semi-Active Car Suspension Systems, Proc. I.Mech.E., Vol. 200 (D3), pp219-228.

/6/ Tang C.K.K., and Mars, P., 1993, Games of Stochastic Learning Automata and Adaptive Signal Processing, IEEE Trans. Systems, Man and Cybernetics, Vol. 23, No. 3, May/June 1993.

/7/ Wu Q,H., 1993, Learning Co-ordinated Control of Synchronous Machines in Multi-machine Power Systems, Proc. IEEE Systems, Man and Cybernetics 1993 Conference , Vol. 3, pp728 - 733.

APPENDIX

NOTATION

Δ	action test period
k	discrete time variable
J	action cost
R	reference set of stored costs
K	gain vector
k_i	controller gains
r	number of quantisation levels
α_i	i^{th} automaton action
p_i	probability of i^{th} action
θ	automaton learning rate

λ	re-quantisation reduction scale factor
η	re-quantisation threshold
$\dot{v}_b$	vehicle body acceleration at corner a
a	corner index, $a = 1,2,3,4$

Learning Algorithm

$n = 0$

Repeat While $\lambda^n > 0.01$

    define $A(n)$ via equations (11), (12)

    $k = 1$

    $P(k) =$ uniform distribution

    $R = \varnothing$

    Repeat While $\max\{p \in P(k)\} < \eta$

        randomly select index i according to the probability distribution $P(k)$

        select gain vector $\mathbf{K} \in A(n)$ corresponding to index i

        evaluate suspension system performance over a 16 second time interval

        evaluate the performance index J via equation (8) and append to R

        evaluate minimum and median values $J_{\min}$ and J_{med} of R

        evaluate $\beta(k)$ via equation (10)

        evaluate $P(k+1)$ via equation (9)

        increment k

    End Repeat

    increment n

    $\mathbf{K}(n) = \arg\max p_i(k)$

End Repeat

Formulae

$$J(k) = \frac{1}{\Delta} \sum_{t-\Delta}^{t} \left(\dot{v}_b^2\right) \tag{8}$$

$$\left.\begin{aligned} p_i(k+1) &= p_i(k) + \theta\beta(k)(1 - p_i(k)) \\ p_j(k+1) &= p_j(k) - \theta\beta(k)p_j(k) \end{aligned}\right\}(j \neq i) \tag{9}$$

$$\beta(k) = \max\left\{0, \frac{J^a_{\text{med}} - J^a(k)}{J^a_{\text{med}} - J^a_{\min}}\right\} \tag{10}$$

$$s(n) = \lambda^n s(0) \tag{11}$$

$$\begin{aligned} A(n) = \{K_1(n), K_1(n) \pm s_1(n)\} \times \{K_2(n), K_2(n) \pm s_2(n)\} \times \ldots \\ \{K_3(n), K_3(n) \pm s_3(n)\} \end{aligned} \tag{12}$$

Parameter values

$K(0) = [7500, 1000, -2000]$

$s(0) = [7500, 1000, 2000]$

$w_3 = 1,\ \Delta = 16$ seconds

$\theta = 0.1,\ \lambda = 0.4,\ \eta = 0.5$