When are tasks "difficult" for learning controllers?

Shlomo Geva and Joaquin Sitte
Faculty of Information Technology
Queensland University of Technology
GPO Box 2434 Brisbane Q 4001 Australia
s.geva@qut.edu.au, j.sitte@qut.edu.au

Hebertt Sira-Ramírez
Departamento de Sistemas de Control
Universidad de Los Andes
Mérida, Venezuela
isira@ing.ula.ve

Abstract

Some control task, formerly believed to be difficult and used to demonstrate neural network unsupervised learning, can be accomplished with very simple controllers. There seems to be no learning method that discovers these controllers. This failure could be attributed to the learning algorithm or to starting with overly complex controller structures. We suggest using the probability that randomly selected controller is successful as measure of learning difficulty. To limit the size of the space of possible controllers we use Fliess' classification of control problems into *flat* and *non-flat*. We illustrate the procedure on *easy* and *difficult* variants of the pendulum control task.

1 Introduction

Several nonlinear control problems have become popular among machine learning researchers, particularly those using neural networks, for demonstrating learning methods. The cart-pole and the truck backer-upper first studied in this context by Widrow and coworkers have become classic examples [11] [8]. The learning tasks associated with such nonlinear control problems are often intuitively classified as *difficult* because human beings need a substantial amount of practice, based on trial and error, before they become capable of executing the proposed control task to a reasonable degree of satisfaction. One finds, however, that some of the simplest imaginable controllers: linear or piecewise linear state feedback controllers deliver very good regulatory behaviour for the closed loop dynamics of these systems. Moreover, once the controller structure has been decided upon, the probability that a random choice of the parameter values for the structure will result in a successful controller is surprisingly large. For example we have shown that a simple random search in coefficient space for a linear feedback controller is more effective, at least for the cart-pole and the articulated truck backer-upper experiments, than any of the known learning methods [5] [6].

The simplicity of state feedback solutions, and the apparent ease with which they can be obtained, seems paradoxical when compared with the considerable computational effort demanded by the unsupervised learning methods. It seems worthwhile to investigate this situation. What is the reason for the poor performance of the learning methods? Are the methods at fault?, or is it that the initial range of potential controllers is far too large, i.e. the allowed controller structures are far too complex. To decide this question it would be helpful to have a way of quantifying the inherent *difficulty* of a learning task.

In this paper we discuss a measure of difficulty based on linear or piecewise linear solutions to control tasks. Because of space limitations we confine the discussion to pendulum systems. We already discussed linear and piecewise linear solutions to reverse parking of vehicles elsewhere [6] [4].

2 Volume of controllers in parameter space

A structure defines a family of controllers. Particular sets of values of the variable parameters instantiate the individual controllers in the family. Following the ideas of Schwartz et al. [10] we use the relative volume in parameter space occupied by the set of successful feedback controllers as a measure of learning difficulty. This measure defines a probability on the space of controllers of given structure. It is the probability that a randomly chosen controller will meet a specified performance level. Each task needs a clear specification of the performance criteria for discriminating between successful and unsuccessful controllers. The most important question, however, for obtaining actual values of the probability is how to choose a structure of sufficient generality without being unnecessarily complicated. We find an answer in Fliess' classification of nonlinear control systems into *differentially flat* and *differentially non-flat* [2] [3] [9].

0-7803-1901-X/94 $4.00 ©1994 IEEE

2.1 Differentially flat and non-flat systems

Differentially flat systems are best represented by the class of systems that are linearisable to decoupled controllable systems by means of *endogenous* feedback, that is: feedback form internal variables. The inherent linearity of flat systems may then be exploited by means of either a single linear control law or, alternatively, a piecewise linear control law that efficiently compensates for the system nonlinearities and manages to stabilise the system to its desired set point. It has been argued that most of the commonly studied nonlinear control system examples are indeed differentially flat. For instance, the class of nonlinear Hamiltonian (conservative) systems are, generally speaking, "flat". A large class of mechanical systems with nonholonomic velocity constraints are also differentially flat. The class of differentially non-flat systems is made up by systems satisfying the so called *strong accessibility property* and they are essentially uncontrollable systems. Some non–flat systems can be reduced to differentially flat systems, in an average sense, by adding an independent high frequency control input.

Flat systems are characterised by a set of linearising outputs whose number equals that of the control inputs. Every variable in the system, including the control inputs, are expressible as a *differential function* of the linearising outputs. A differential function is a function of its variables and a finite number of their time derivatives. The variables to be regulated are therefore functions of the linearising outputs and a finite number of their time derivatives. A linear feedback policy is easily seen to always stabilise the linearising outputs independently of the distance of the actual, or initial states to the required equilibrium conditions.

3 Examples

3.1 The cart-pole experiment

The cartpole is a modified form of balancing a broomstick. The lower end of a pole is hinged to a cart in such a way that the pole can only swing in a vertical plane parallel to the direction of motion of the cart. The objective is to balance the pole by pushing the cart back and forth. In the most common version of the experiment, described by Barto, Sutton & Anderson (1983) [1], the cart runs on a track of limited length. Balancing *fails* when the inclination of the pole exceeds preset limits, or when the cart hits the stops at the end of the track. A more demanding version of the cartpole experiment requires the controller to balance the pole and bring the cart back to the centre of the track. The state vector of the cartpole has four components: the pole angle θ with the vertical, the angular velocity ω, the cart position x, and its velocity v. The dynamic equations of the system can be found in [5]. A linear state feedback controller that generates a control force F of the form

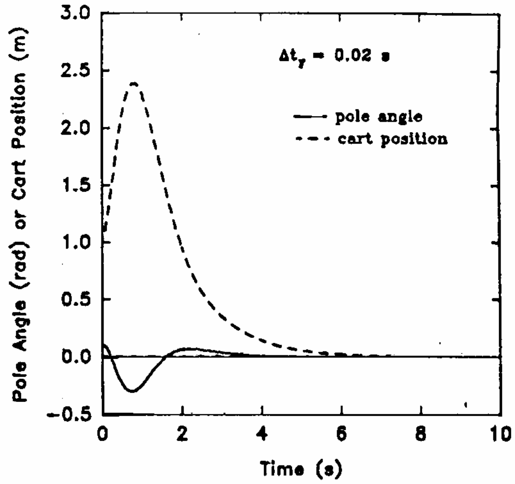

Figure 1: Cart position and pole angle vs. time for the *best* controller in the continuous force regime. Force update interval $\Delta t_F = 0.02s$. Initial condition: $x = 1m$, $v = 1m/s$, $\theta = 0.1rad$ and $\omega = 0.2rad/s$. RMS values over first 1000s: $\bar{x} = 0.0076m$, $\bar{\theta} = 0.0768rad$

$$F = k(w_\theta \theta + w_\omega \omega + w_x x + w_v v) = k(\vec{w} \cdot \vec{s}) \quad (1)$$

does a very good job in balancing the pole and centring the cart, as the trial run in Fig. 1 illustrates. The weight vector $\vec{w} = (w_\theta, w_\omega, w_x, w_v)^T$ is of unit length.

A simple qualitative analysis shows that in a successful controller all the components of the weight vector must be positive. We can measure the degree of difficulty of the problem by the fraction of the area of the positive quadrant of the four-dimensional unit sphere in weight space that yield satisfactory controllers. The degree of difficulty depends of course on the initial release condition. The closer this is to equilibrium the easier is the task. For the reasonably difficult initial condition of the trial run in Fig. 1 about 0.3 % of the randomly selected controllers had not failed 5 minutes after release. In other words the most unsophisticated learning method of testing random controllers will succeed in producing a successful controller after 330 trials on average. For a bang-bang control force the task of not failing is easier but the quality of the control is much worse [5].

3.2 The swingless crane

Another control problem involving the pendulum dynamics is that of the swingless crane. The simplest form of the task is the loading bridge configuration. A carriage runs on an overhead rail, and a load hangs on a rope attached to a carriage. The load moves in the vertical plane below the rail and the control task is to move the load from an initial position (x_i, z_i) to a final target position (x_f, z_f) in minimum time and without overshoot. This is to be achieved by moving the carriage along the rail and by pulling in or releasing rope. The dynamic equations for the loading bridge are:

$$L\ddot{\theta} = \frac{(1-\mu_l)a_{cx}\cos\theta}{1-\mu_l\cos^2\theta} - \frac{(\mu_l(2\dot{L}\dot{\theta}\sin\theta - (\ddot{L} - L\dot{\theta}^2)\cos\theta) + g)\sin\theta}{1-\mu_l\cos^2\theta} \quad (2)$$

$$\ddot{x}_c = \frac{(1-\mu_l)a_{cx} + \mu_l(\ddot{L} - L\dot{\theta}^2 - g\cos\theta)\sin\theta}{1-\mu_l\cos^2\theta} + \frac{(1-\mu_l)2\mu_l\dot{L}\dot{\theta}\cos\theta}{1-\mu_l\cos^2\theta} \quad (3)$$

$$\ddot{L} = a_L(t) \quad (4)$$

μ_l is the reduced mass $m_l/(m_c+m_l)$ of the load, θ is the angle of the rope with the vertical (measured clockwise), L is the free length of the rope. The controls are the reduced force $a_{cx} = F_x/(m_l + m_c)$ applied to the carriage and the lengthwise acceleration a_L of the rope.

The problem is flat [2]. The following linear state feedback controller achieves good control:

$$a_{cx} = 50\theta - 0\dot{\theta} - 1.4(x_t - x) - 4\dot{x} \quad (5)$$

$$a_L = (z_t - z) - 2\dot{L} \quad (6)$$

Figure (2) shows a sample trajectory of the load.

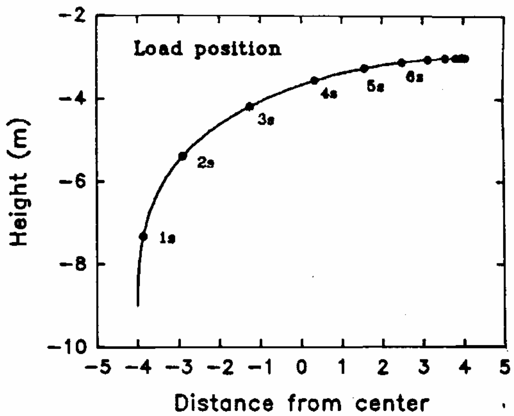

Figure 2: Load trajectory obtained by linear state feedback. The load is moved from a point 9 m below the rail and 4m to the left oft the centre to a point 3m (z_t) below the rail and 4 m (x_t) to the right of the centre. There is no residual oscillation. m_c = 200 Kg and m_l = 2000 Kg

3.3 Kapiza's pendulum

The Kapiza pendulum is a variation of the cartpole experiment that is ***non-flat*** [3]. Instead of balancing the pole by a linear horizontal motion of the pole's base, only vertical up and down motion is allowed. It is known that an inverted pendulum can be prevented from falling over by a high frequency (much higher than the natural frequency of oscillation of the pendulum) oscillatory vertical motion forced on the base of the pole [7]. This is an example of the class of non–flat problems that can be reduced, in an average sense, to a flat system by adding an independent high frequency control input. Given the assurance that stabilisation is possible, the control objective is then to apply a variable force to the base of the pole so as to keep it around a fixed vertical position while maintaining the pole in the upright position. The dynamic equations for a Kapiza pendulum of length 2L and mass m_p:

$$\ddot{\theta} = \frac{3\sin\theta(a_z + \dot{\theta}^2 L\mu_p\cos\theta)}{L(4 - 3\mu_p\sin\theta\cos\theta)} \quad (7)$$

$$\ddot{z} = \frac{4(a_z + \dot{\theta}^2 L\mu_p\cos\theta)}{4 - 3\mu_p\sin^2\theta} - g \quad (8)$$

where μ_p is the reduced mass $m_p/(m_p + m_c)$ of the pole. The mass of the pole mounting and driving gear is m_c. The controller applies the force $F = a_z(m_p + m_c)$ to the base of the pole. To simulate the system we have used Euler integration with integration time step $\Delta = 0.02s$, and that was also the control feedback interval. The system parameters were $m_p = 1Kg$, $m_c = 0$, $L = 1m$, $g = 9.81m/s$. There is indeed a structure for piecewise linear controllers that keeps the base of the pole near the origin and prevents the pole from falling over. We derived that structure from qualitative analysis of the pendulum dynamics. Acceptable parameters were found by experimentation. Remarkably, we did not have to rely on an additional high frequency input, however the trajectories of the controlled pendulum are oscillatory. The best controller we found is:

$$a_z = \begin{cases} 12g\sigma(8(z - .54) + \dot{z}) \\ \text{if } ((\theta\dot{\theta} < 0) \wedge (4|\theta| < |\dot{\theta}|)) \vee (|\theta| < 0.001) \\ -12g \qquad \text{otherwise} \end{cases} \quad (9)$$

The system is released from the initial configuration $z = -1.0m$, $\dot{z} = 0.0m/s$, $\theta = 0.5rad$, and $\dot{\theta} = 0.0rad/s$. Figure 3 shows the first 30 seconds of a sample trajectory. The residual root mean square error for the position and angle, measured

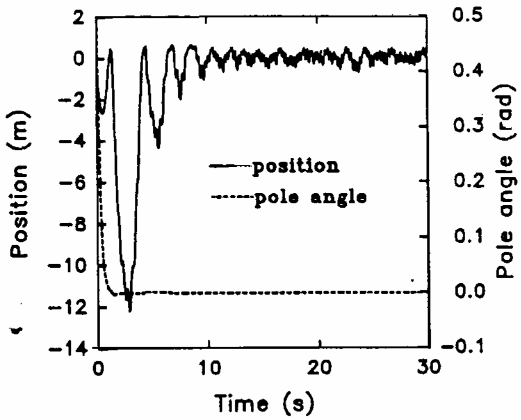

Figure 3: Sample trajectory for balancing the Kapiza pendulum

between 30 and 60 seconds after release, are $0.26m$ and $0.027rad$ respectively. While the controller is not sensitive to small changes in the numerical constants used in the control law, we were unable to improve on this result with a reasonable amount of experimentation..

4 Conclusion

The classification of dynamic systems into differentially flat and non-flat is useful for constraining the structure of learning controllers. Since many systems of practical importance turn out to be flat, characterising the difficulty of the learning problem by the volume of good controllers in parameters space leads to meaningful discrimination of learning control tasks. The set of task related to the control of a pendulum discussed in this paper, covering the range from easy (cart-pole) to difficult (Kapiza pendulum), clearly illustrate this.

References

[1] A.G. Barto, R.S. Sutton, and Anderson C.W. Neuronlike adaptive elements that can solve difficult learning control problems. *IEEE Trans. Syst., Man., Cybern.*, SMC 13:834–846, 1983.

[2] Michel Fliess, Jean Levine, Philippe Martin, and Pierre Rouchon. Sur les systems non lineaires differentiellment plats. *C. R. Acad. Sci. Paris, Serie 1*, 315:619–624, 1992.

[3] Michel Fliess, Jean Levine, Philippe Martin, and Pierre Rouchon. Defaut d'un systeme non lineaires et commande haute frequence. *C. R. Acad. Sci. Paris, Serie 1*, 316:513–518, 1993.

[4] Shlomo Geva and Joaquin Sitte. One neuron valet parking. In *Proceedings of the IEEE International Workshop on Emerging Technologies and Factory Automation (ETFA'92)*, pages 277 - 282, Melbourne, 1992.

[5] Shlomo Geva and Joaquin Sitte. A cartpole experiment benchmark for trainable controllers. *IEEE Control Systems Magzine*, 13(5):40–51, October 1993.

[6] Shlomo Geva and Joaquin Sitte. Simple controllers for reverse parking of articulated vehicles. *IEEE Transactions on Industrial Electronics*, 1994. accepted.

[7] L.D. Landau and Lifshitz E.M. *Mechanics.* Pergamon Press, 1969.

[8] D Nguyen and Bernard Widrow. The truck backer-upper: an example of self-learning in neural networks. In R.S. Sutton W.T. Miller and P.J. Werbos, editors, *Neural Networks for Control*, pages 287–299. 1990.

[9] Pierre Rouchon, Michel Fliess, Jean Levine, and Philippe Martin. Flatness and motion planning: the car with n trailers. In *Proceedings of the 2nd European Control Conference*, Groningen, 1993.

[10] D.B. Schwartz, V.K. Samalam, Sara A. Solla, and J.S. Denker. Exhaustive learning. *Neural Computation*, 2:374–385, 1990.

[11] Bernard Widrow. The original adaptive neural net broom-balancer. In *International Symposium on Circuits and Systems*, pages 351–357. IEEE, May 1987.