21st International Symposium on
Mathematical Theory of Networks and Systems
July 7-11, 2014. Groningen, The Netherlands

Learnability for Dynamical Systems

Rowland O'Flaherty and Magnus Egerstedt

Abstract— **This paper takes a step towards defining what it means for a dynamical system, such as a robot, to be able to learn through the notion of *learnability*. It takes a system-theoretic view to define what learning is and establishes when a system can and can not learn. Equipped with this definition of learnability, we provide a learnability result for linear systems with quadratic costs that is then applied to two different types of mobile robots, namely a simulated two-wheel inverted pendulum robot and a real, locomoting robotic platform.**

I. INTRODUCTION

As robotic systems become more interconnected, complex, and large-scale, machine learning tools are used increasingly when designing controllers – typically in conjunction with more standard control design methodologies. However, it is not clear when such methods are appropriate or effective, and this has to do, in part, with a lack of a formal characterization of when the solution to a problem is "learnable" by a particular system. In this paper, we attempt to remedy this by proposing a formal definition of what learnability means when viewed from a system-theoretic vantage-point, and then apply this definition to different types of mobile robots.

In order to produce a definition of learnability that is both relevant and applicable, one must first understand what "learning" itself might mean. In the Merriam-Webster dictionary [1], *to learn* is "to gain knowledge or understanding of a skill by study, instruction, or experience." But, some skills are clearly not learnable by a particular system. For example, a wheeled ground robot cannot learn to fly no matter the sophistication of its algorithms. Or, a robot with no sensors whatsoever can not learn how to approach a landmark. These two rather extreme examples both hint at the roles that classic system theoretic notions such as controllability and observability might take on when defining learnability.

In the traditional machine learning community, learning has been analyzed and characterized in different settings, and a working definition [2], [3], states that "A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E". From this rather informal definition one can deduce that if the performance measure never improves no matter what the experience might be, then the computer program can not learn. We will follow this intuitive notion when defining the learnability concept for dynamical systems.

The main contribution of this paper is a definition of what it means for a dynamical system to have the learning capability to improve its performance relative to a given performance cost. We call this notion *learnability*. Armed with such a definition, one can test whether a given dynamical system can learn a given task; similar to how one can test whether a given system is controllable or observable. The condition for learnability differs from these other standard conditions in that not only does it depend on the dynamical model of the system, but also on the cost function that is associated with the task. This means that a given system might not have learnability for one type of task but does for another type.

The term "learnability" has been used in other contexts before. The notion of learnability is most widely used in the field of statistical learning theory, which is the mathematical analysis of machine learning. Campi and Vidyasagar define *probably approximately correct* (PAC) learnable for a function class if there exists an algorithm that PAC-learns that function. An algorithm is PAC if the probability of the generalized error of the hypothesis for a target function being greater than ϵ goes to zero as time goes to infinity [4] [5]. This notion was proposed by Valiant in 1984 [6]. Distribution-free PAC learnablility is equivalent to a finite VC-dimension, which is named after the foundation work done in learning theory by Vapnik and Chervonenkis [7]. Learnability is also used in other areas. For example, learnability is used in software testing and is defined in ISO 9126 [8] as attributes of a software product that have a tolerance on the effort required by its users to learn its application.

Within the domains of controls, robotics, and dynamical systems, related ideas have been pursued in the contexts of system identification and optimal control through the investigation of conditions under which system parameters can be identified or optimal solutions do exist [9]. For example, Vidyasagar and Karandikar use a learning theoretic approach for system identification and use PAC learning as a condition on which the system can be identified [10].

Optimal control is also related to reinforcement learning, which was developed in the artificial intelligence community. In fact, strong connections between optimal control and reinforcement learing are made by Sutton and Barto [11]. This is because both optimal control and reinforcement learning seek to find a policy that minimizes some cost function (or maximizes some reward function) for a given task. The difference is in the type of problems that each technique is better suited for solving. Given this connection

R. O'Flaherty is with the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA. rowland.oflaherty@gatech.edu

M. Egerstedt is with the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA. The work by M. Egerstedt was supported by the DARPA M3 program. magnus@gatech.edu

it is reasonable to assume that a condition on when a system can learn depends on the cost function and the optimal control policy that minimizes that cost function. This is in fact what is used in this paper to define learnability.

This paper is organized in the following manner. In Section II we define learnability. In Section III we further motivate learnability with a simple example. In Section IV we describe learnability for a linear system and introduce a theorem defining when a linear system is learnable. In Section V we illustrate learnability on a two-wheel inverted pendulum robot. In Section VI we describe an experiment with learnability on real robot and show results for these experiments. We conclude the paper in Section VII.

II. Learnability Definition

Even though the word "learnability" has been used in a number of different fields and has different (mostly informal) definitions, the semantics, which defines when knowledge can be gained, is consistent. The motivation behind this work is to produce a precise system theoretic definition based on the idea that a system is learnable if and only if the system is capable of both performing actions and observing outputs that relate to the cost function that the system is attempting to minimize. Particularly, the system will be able to acquire the necessary knowledge to complete its objective if and only if (i) there are different initial conditions that result in different optimal control policies and (ii) there are different initial conditions that produce different perceived costs for at least some input signal. We use perceived cost because it is assumed that the system does not have direct access to its state information; therefore, it must estimate both its state and the cost that is associated with that estimated state.

The intuition for the first criterion is that if the best action to take from every initial condition gives the same terminal cost then there is nothing that the system is capable of doing to reduce the cost. Therefore, every initial condition will have the same optimal control policy. Thus, learning can not take place because the performance is not being improved with added experience. The intuition for the second criterion is that if the perceived cost is the same no matter where the system is located then no matter what the system does the cost will not be reduced. Again, learning will not be possible.

To make this more precise, consider a system whose dynamics are of the form

$$\dot{x} = f(x,u), \ y = h(x,u), \ z = g(x,u), \tag{1}$$

where $x(t) \in \mathbb{X}$ is the state, $u(t) \in \mathbb{U}$ is the input, $y(t) \in \mathbb{Y}$ is the measurement, and $z \in \mathbb{Z}$ is the signal-of-interest to the problem that is to be solved. The system dynamics are encoded by the function f, the sensors are encoded through h, and the signal-of-interest is encoded by the function g. The signal-of-interest is the signal that actually affects the cost. The space of input signals is denoted as $\mathcal{U}$. Moreover, given that the state can not be measured directly it must be estimated. Let us assume that the estimated state has been obtained through an observer, e.g.,

$$\dot{\hat{x}} = l(\hat{x}, y, u), \tag{2}$$

for some observer dynamics l. We will denote the estimated signal-of-interest as $\hat{z} = g(\hat{x}, u)$.

Learning only makes sense if something is actually supposed to be learnt, i.e., when there is a cost to be minimized. This cost could for instance be total distance traveled by a robot towards a landmark, and so forth. In this paper, we let the cost function that the system is minimizing be of the form

$$J(u) = \int_0^T \Lambda(u,z)dt + \Psi(z(T)), \tag{3}$$

where $\Lambda : \mathbb{U} \times \mathbb{Z} \to \mathbb{R}_+$ is the instantaneous cost and $\Psi : \mathbb{Z} \to \mathbb{R}_+$ is the terminal cost for the final time T. For our purposes it is convenient to include the initial state of the system $x_0 = x(0)$ into the argument of the cost. Thus, $J(u, x_0)$ denotes the cost defined over two arguments; the control input and the initial condition. Let us denote the minimizer to the cost for a given initial condition as

$$u^*_{x_0} = \arg\min_{u \in \mathcal{U}} J(u, x_0), \tag{4}$$

and we assume that this minimizer exists. Given that the state can not be measured directly and must be estimated, we denote the estimated cost (or perceived cost) as

$$\hat{J}(u, x_0) = \int_0^T \Lambda(u, \hat{z})dt + \Psi(\hat{z}(T)). \tag{5}$$

With the above system definition we can formally define *learnability*.

Definition 1 (Learnability): The tuple (f, h, g, l, J) is said to be learnable if

1) the state informs the control, i.e.

$$\exists x_0, x_0' \text{ s.t. } u^*_{x_0} \neq u^*_{x_0'}, \tag{6}$$

2) the cost is influenced by the output, i.e.

$$\exists x_0, x_0', u \text{ s.t. } \hat{J}(u, x_0) \neq \hat{J}(u, x_0'). \tag{7}$$

This definition may seem obvious, but it does capture the two essential ingredients, namely that what the system does has an impact on the cost, and that the state of the system has an impact on what the control input should be.

III. A Simple Yet Illustrative Example

Let us demonstrate the learnability condition on a simple example. Imagine a caterpillar that is inching along on a thin straight branch that is lying on the ground. The caterpillar can only move along the length of the branch – let us call that direction the x-direction – and its head is turned to the side, so it can only see to the side of the branch – the y-direction. The caterpillar uses its eyes as perfect observers to measure distances in the y-direction. Moreover, the caterpillar's goal is to reach a delicious green leaf high above in the tree that hangs directly over the caterpillar – the z-direction – at distance η. Will the caterpillar be able to learn the necessary actions to reach its desired meal?

Let us set up the problem as follows to answer our question; the dynamics are

$$\dot{x} = [1, 0, 0]^{\mathsf{T}} u, \ y = [0, 1, 0]x, \ z = [0, 0, 1]x, \tag{8}$$

and the cost is

$$J(u, x_0) = \int_0^T \|u\|^2 dt + (z(T) - \eta)^2. \qquad (9)$$

This cost says that the caterpillar would like to exert the least amount of energy while getting as close as possible (in terms of height off the ground) to the leaf. However, no matter where the caterpillar starts on that branch the optimal input is to do nothing. This is because $z = 0$ for all possible x values that the caterpillar can obtain. This implies that $\Psi = \eta^2$ and therefore $u^* = 0 \ \forall x_0$. The first condition of learnability is violated. Therefore, learning is not possible. Let us look at the second condition (7) just for demonstration. For any given input that the caterpillar uses the cost will always be $\hat{J}(u, x_0) = \hat{J}(u, x_0')$, this is because the measurements that the caterpillar is taking in the y-direction do not affect the cost. Therefore, the second condition is also violated. The caterpillar has no hope of learning how to get to that delicious green leaf.

IV. Learnability for Linear Systems

The definition of learnability establishes a particularly direct result for when a linear system is learnable with respect to a quadratic cost. In this section we formulate this linear system problem for learnability and state the linear learnability theorem.

A. Problem Statement

Consider a continuous-time linear time-invariant system of the form

$$\dot{x} = Ax + Bu, \ x \in \mathbb{R}^n, \ u \in \mathbb{R}^k, \qquad (10)$$
$$y = Cx, \ y \in \mathbb{R}^m, \ m \leq n,$$

where x is the state, u is the control signal, and y is the measurement. In addition, let us define the signal-of-interest as

$$z = Gx, \ z \in \mathbb{R}^l, \ l \leq n. \qquad (11)$$

Furthermore, let us assume that an estimate of the state is accomplished using a Luenberger observer of the form

$$\dot{\hat{x}} = A\hat{x} + Bu + L(y - C\hat{x}). \qquad (12)$$

Therefore, our estimated signal-of-interest is

$$\hat{z} = G\hat{x}. \qquad (13)$$

We do not want to impose unnecessary constraints on the choice of observer gain in (12); however, we do need to assume that the observer is not harmful in the sense that the observer should reflect different state estimates (after a while) if the difference between two different initial states are observable. In other words, we assume that[1]

$$\mathcal{R}(C^\mathsf{T}) \subseteq \mathcal{R}(L), \qquad (14)$$

which is sufficient to ensure that at least some initial conditions on the state result in different state estimates.

[1] $\mathcal{R}()$ and $\mathcal{N}()$ denote the range space and null space, respectively.

The objective is for the system to learn how to transition from $z(t)$ (a point in $\mathbb{R}^l$ corresponding to the current output) to another point η (a desired point in $\mathbb{R}^l$) with minimal input. The cost and estimated cost for the problem can be written as

$$J(u, x_0) = \int_0^T \|u\|^2 dt + (z(T) - \eta)^\mathsf{T} S (z(T) - \eta), \qquad (15)$$
$$\hat{J}(u, x_0) = \int_0^T \|u\|^2 dt + (\hat{z}(T) - \eta)^\mathsf{T} S (\hat{z}(T) - \eta), \qquad (16)$$

where $x_0 = x(0)$, T is the in final time, and $S = S^\mathsf{T} \succ 0$ is a weighting on the distance to the goal for each output component as compared with the input signal.

B. Linear System Learnability Theorem

Here we state the theorem for learnability on a linear system given a quadratic cost. Let us first introduce some notation. The controllability Gramian is denoted as Γ and the observability Gramian is denoted as Ω and are defined as

$$\Gamma = \int_0^T e^{As} B B^\mathsf{T} e^{A^\mathsf{T} s} ds, \qquad (17)$$
$$\Omega = \int_0^T e^{A^\mathsf{T} s} C^\mathsf{T} C e^{As} ds. \qquad (18)$$

Theorem 1 (Learnability): A system is learnable if

$$\mathcal{R}(G^T) \subseteq \mathcal{R}(\Gamma) \cap \mathcal{N}(\Omega)^\perp \text{ and } G \neq 0. \qquad (19)$$

Please refer to Appendix A for the proof of Theorem 1.

Theorem 1 declares that the states that affect the output must be states that are both controllable and observable. However, it is worth noting that learnability does not imply complete controllability or complete observability. A partially controllable and partially observable linear system may still be learnable. And, one might argue that this theorem is almost a truism, i.e., that it states almost exactly what one would expect. This is indeed the case and we do not take the lack of any major surprises here as a negative. Rather, it shows that the definition of learnability coincides well with what learnability should indeed mean.

V. Inverted Pendulum Robot

Let us give an example of learnability by determining if a two-wheel inverted pendulum robot can learn to move towards a goal. The dynamics of a two-wheel inverted pendulum robot can be found in [12] and [13] and an illustration of the system is shown in Fig. 1.

The state of this systems is

$$x = [x_1, \ x_2, \ v, \ \psi, \ \dot{\psi}, \ \phi, \ \dot{\phi}]^\mathsf{T}, \qquad (20)$$

where x_1 and x_2 is the position of the robot in the 2D plane, v is the forward velocity, ψ is the angle of the robot's orientation relative to the x_1-axis, and ϕ is the angle of the robot's posture relative to the straight up unstable configuration.

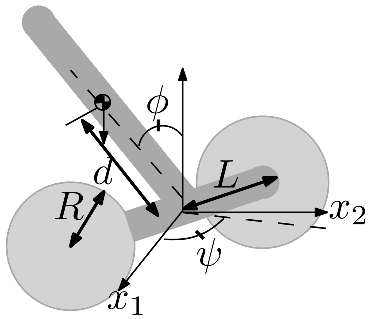

Fig. 1: Illustration of two-wheel inverted pendulum robot [13].

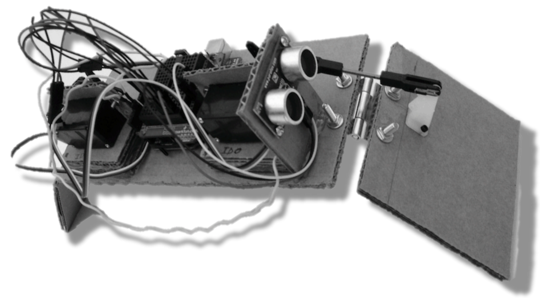

Fig. 2: Physical robotic system used to demonstrate learnability.

The linearization of these dynamics around an operating point of

$$\bar{x} = [\bar{x}_1,\ \bar{x}_2,\ 0,\ \bar{\psi},\ 0,\ 0,\ 0]^{\mathsf{T}} \text{ and } \bar{u} = [0,\ 0]^{\mathsf{T}},$$

for a particular set of parameters that were used in [12] are shown here,

$$A = \begin{bmatrix} 0 & 0 & cos(\bar{\psi}) & 0 & 0 & 0 & 0 \\ 0 & 0 & sin(\bar{\psi}) & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2.16 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 72.49 & 0 \end{bmatrix},$$

$$B = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ -1.67 & -1.67 \\ 0 & 0 \\ 0.03 & -0.03 \\ 0 & 0 \\ -24.15 & -24.15 \end{bmatrix}.$$

We will say that the system has direct measurements of its x_1 and x_2 position, thus

$$C = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

The cost function $J(u, x_0)$ is the same as it is written in (15). It is straight forward to show that this is not completely controllable $(rank(\Gamma) = 6)$ and not completely observable $(rank(\Omega) = 5)$; however, that does not mean it is not learnable.

For this system the intersection of the controllable space and observable space is

$$\mathcal{R}(\Gamma) \cap \mathcal{N}(\Omega)^{\mathsf{T}} = span \left\{ \begin{bmatrix} 0\\0\\0\\0\\0\\1\\0 \end{bmatrix}, \begin{bmatrix} 0\\0\\0\\0\\0\\0\\1 \end{bmatrix}, \begin{bmatrix} 1\\0\\0\\0\\0\\0\\0 \end{bmatrix}, \begin{bmatrix} 0\\0\\1\\0\\0\\0\\0 \end{bmatrix} \right\}.$$

First, we will assume that the goal location is along the x_1-axis and that the robot is linearized around the orientation of $\bar{\psi} = 0$. In this case G is written as

$$G = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

and $\mathcal{R}(G^{\mathsf{T}})$ is obviously subset of $\mathcal{R}(\Gamma) \cap \mathcal{N}(\Omega)^{\mathsf{T}}$; therefore, the system is learnable.

Now let us assume that the goal location is not on the x_1-axis and that the robot is still linearized around the orientation of $\bar{\psi} = 0$. In this case G is written as

$$G = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Here $\mathcal{R}(G^{\mathsf{T}})$ is not a subset of $\mathcal{R}(\Gamma) \cap \mathcal{N}(\Omega)^{\mathsf{T}}$; therefore, the *linearized* system is not learnable.

This illustrative example demonstrates that if a system is not inherently linear then one must be careful in using Theorem 1 to test for learnability with the linearization of the system. This is similar to how one must be careful in testing for controllability by using the linearization of the system.

VI. Simple Example Revisited On A Real Robot

In Section III we show why a caterpillar is unable to learn how to get to a delicious green leaf when it is constrained to a branch on the ground. To further demonstrate this caterpillar example we built a simple robot that captures the constrained controllable actions and observable measurements of the caterpillar. A picture of this robot is shown in Fig. 2.

This robot consists of a body and two moveable appendages. Each appendage has only one rotational degree of freedom that are parallel. This limits the robot to moving only forward and backward along a straight line. In addition, the robot has a sensor to measure distance. The objective of the robot is to learn how to move forward towards a goal that is some distance away.

The model we used to approximate the robot is a linear system of form used in Section IV-A. Let us use Theorem 1 to show when the robot can and can not learn. For this system

$$A = \mathbf{0}^{2\times 2},\ B = [1, 0,]^{\mathsf{T}},\ \text{and } G = [1, 0].$$

Here the states are the (x, y) location of the robot in the ground plane. We have assumed the input drives the robot forward or backward along the x-axis and is not an input to each individual appendage. It is easy to see from these dynamics that

$$\mathcal{R}(G^{\mathsf{T}}) = span\{[1,0]^{\mathsf{T}}\} \text{ and } \mathcal{R}(\Gamma) = span\{[1,0]^{\mathsf{T}}\}.$$

We ran two experiments where we changed the position of the sensor for the robot. In the first experiment we position the robot's sensor so that it is aligned with the direction of movement (x-axis). Thus,

$$C_1 = [1,0] \text{ and } \mathcal{N}(\Omega)^{\perp} = span\{[1,0]^{\mathsf{T}}\}.$$

In the second experiment we position the robot's sensor so that it is aligned orthogonal to the direction of movement (y-axis). Thus,

$$C_2 = [0,1] \text{ and } \mathcal{N}(\Omega)^{\perp} = span\{[0,1]^{\mathsf{T}}\}.$$

According to Theorem 1 the objective is learnable in the first experiment and not learnable in the second experiment. For both experiments we ran a standard reinforcement learning algorithm, specifically Q-learning [11], and observed how well the robot learned to move towards the goal. The results are shown in Fig. 3.

In both experiments the robot is trying to learn how to reach a goal that is at 35 cm in the x-direction (robot's position is dark dash line). In the first experiment the robot explores for roughly the first 60 seconds of the experiment attempting to figure out which actions brings it closer to the goal. After exploring (dark areas of horizontal line) it starts exploiting (light areas of horizontal line) what it has learned and eventually reaches the goal. The total accumulated cost does not continue to increase (light gray dotted line). In the second experiment the robot continuously explores the different actions attempting to learn which action will consistently move it towards the goal. It is unable to learn the actions because it does not have the measurement information to inform it on which actions have moved it forward. The accumulated cost continues to increase [2].

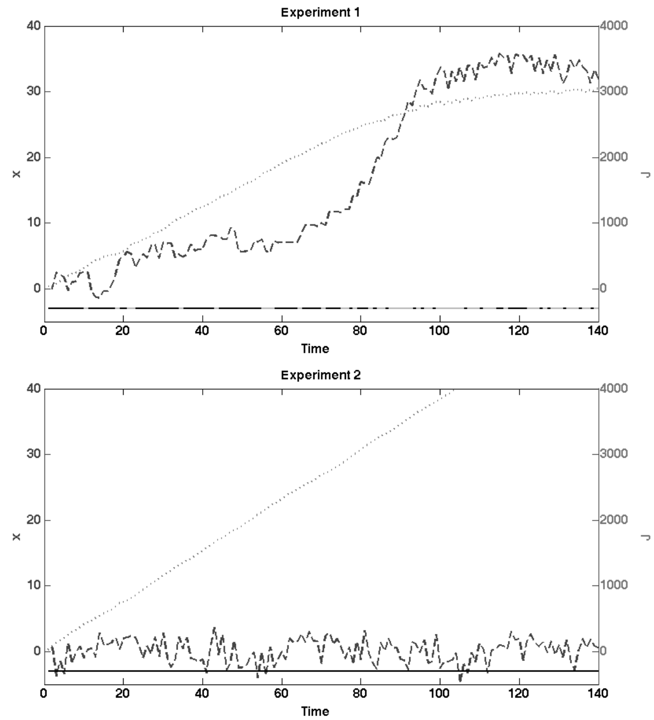

Fig. 3: The top plot shows the results from experiment 1, which is when the robot is measuring distances in the direction it is moving. The bottom plot shows the results from experiment 2, which is when the robot is not measuring distances in the direction it is moving. The dashed black line is the distance the robot has moved. The dotted light gray line is the accumulated cost. The horizontal line along the bottom of the plots show when the robot is exploring new learned actions (dark areas of the line) and exploiting previous learned actions (light areas of the line)

VII. Conclusion

In this paper we defined a notion of learnability with respect to dynamical systems in general, and robots in particular. Using this definition we introduce a theorem defining when a linear system with a Luenberger observer and a quadratic cost is learnable. We apply the result to a linearized system of a two-wheel inverted pendulum robot and to a real, locomoting robot that learns with reinforcement learning. The experiments with the real robot preformed as expected based on the learnability conditions from our theorem. When a system has both the control actions to affect its cost function and the sensor measurements to detect changes in its cost it will be able to learn.

Acknowledgment

The authors would like to thank Greg Droge for helpful discussions regarding the two-wheel inverted pendulum robot.

[2]A video of the physical robot learning can be viewed at http://www.youtube.com/watch?v=cniTpI2oOjI

References

[1] MerriamWebster.com, "learn." `http://www.merriam-webster.com/dictionary/learn`, June 2012. Web. 17 September 2012.

[2] T. M. Mitchell, *Machine Learning*. McGraw-Hill Science/Engineering/Math, 1 ed., Mar. 1997.

[3] Wikipedia.com, "Machine learning." `http://en.wikipedia.org/wiki/Machine_learning`, June 2012. Web. 17 September 2012.

[4] M. C. Campi and M. Vidyasagar, "Learning with prior information," *Automatic Control, IEEE Transactions on*, vol. 46, pp. 1682–1695, Sept. 2001.

[5] M. Vidyasagar, *Learning and Generalization: With Applications to Neural Networks (Communications and Control Engineering)*. Springer, softcover reprint of hardcover 2nd ed. 2002 ed., Dec. 2010.

[6] L. G. Valiant, "A theory of the learnable," in *STOC '84: Proceedings of the sixteenth annual ACM symposium on Theory of computing*, ACM Request Permissions, Dec. 1984.

[7] V. N. Vapnik and A. Y. Chervonenkis, "On the uniform convergence of relative frequencies of events to their probabilities," *Theory of Probability and its Applications*, vol. 16, no. 2, pp. 264–280, 1971.

[8] ISO/IEC, *Part 9126-1:2001: Software engineering – Product quality – Part 1: Quality model.* International Organization for Standardization, Geneva, Switzerland., 18. June 2011.

[9] J. A. E. Bryson and Y.-C. Ho, *Applied Optimal Control: Optimization, Estimation and Control*. Taylor & Francis, revised ed., Jan. 1975.

[10] M. Vidyasagar and R. L. Karandikar, "A learning theory approach

to system identification and stochastic adaptive control," *Journal of Process Control*, vol. 18, pp. 421–430, Mar. 2008.
[11] R. Sutton and A. Barto, *Reinforcement Learning: An Introduction*. Adaptive Computation and Machine Learning, The MIT Press, the mit press ed., Mar. 1998.
[12] Y. Kim, S. H. Kim, and Y. K. Kwak, "Dynamic Analysis of a Nonholonomic Two-Wheeled Inverted Pendulum Robot," *Journal of Intelligent and Robotic Systems*, vol. 44, pp. 25–46, Jan. 2006.
[13] G. Droge and M. Egerstedt, "Multi-Modal Parametrized MPC for Mobile Robot Navigation," *IEEE Transactions on Control Systems Technology*, 2012.
[14] M. Egerstedt and C. Martin, *Control Theoretic Splines: Optimal Control, Statistics, and Path Planning (Princeton Series in Applied Mathematics)*. Princeton University Press, Dec. 2009.

APPENDIX

A. Linear System Learnability Proof

To prove Theorem 1 it must be shown that both conditions in Definition 1 will always be satisfied if (19) is true. The proof is divided into two parts: (i) the first condition of Definition 1 is satisfied by $\mathcal{R}(G^T) \subseteq \mathcal{R}(\Gamma)$ and (ii) the second condition is satisfied by $\mathcal{R}(G^T) \subseteq \mathcal{N}(\Omega)^{\perp}$. For the first part of the proof we need an expression for the minimizer $u^*_{x_0}$ and for the second part of the proof we need an expression for $\hat{J}(u, x_0) - \hat{J}(u, x_0')$.

Before we begin let us write the controllability and observability Gramians, respectively, as

$$\Gamma = K(T)[K^*(t)], \tag{21}$$
$$\Omega = M(T)[M^*(t)]. \tag{22}$$

In (21) $K(t)$ is a linear operator such that

$$K(t)[u(t)] = \int_0^t e^{As} Bu(s)ds, \tag{23}$$

and $K^*(t)$ is the adjoint of K,

$$K^*(t) = B^{\mathsf{T}} e^{A^{\mathsf{T}} t}. \tag{24}$$

Additionally, in (22) $M(t)$ is a linear operator such that

$$M(t)[u(t)] = \int_0^t e^{A^{\mathsf{T}} s} C^{\mathsf{T}} u(s)ds \tag{25}$$

and $M^*(t)$ is the adjoint of M,

$$M^*(t) = Ce^{At}. \tag{26}$$

1) First condition: Let us solve for the minimizer $u^*_{x_0}$ by using a method of projections in a Hilbert space. Egerstedt and Martin outline this method in more detail in [14]. We will begin by defining a Hilbert space as the combination of two inner product spaces $\mathcal{H} : L_2 \times \mathbb{R}^l_S$, where

$$L_2 :< v, w >_{L_2} = \int_t^T v^{\mathsf{T}} w d\tau, \tag{27}$$
$$\mathbb{R}^l_S :< p, q >_{\mathbb{R}^l_S} = p^{\mathsf{T}} S q, \tag{28}$$

i.e.

$$< (v; p), (w; q) >_{\mathcal{H}} = < v, w >_{L_2} + < p, q >_{\mathbb{R}^l_S} . \tag{29}$$

Here we are using v and w as dummy variables for control signals in $\mathcal{U}$ and p and q as dummy variables for points in $\mathbb{R}^l$. We know that

$$x(t) = e^{At}x_0 + \int_0^t e^{A(t-s)} Bu(s)ds = \beta(x_0) + Ku, \tag{30}$$

where K is a linear operator on u as it is defined in (23) and $\beta(x_0) = e^{At}x_0$. We can then express the output as

$$z = G\beta + GKu, \tag{31}$$

where we will drop the dependence on t and x_0 for now. Let us define an affine variety $V_\beta \subseteq \mathcal{H}$, where

$$V_\beta = \{(u; z) \mid GKu - z = -G\beta\}. \tag{32}$$

Now, we can reformulate the minimization of (4) for the linear system in (10) and the cost in (15) as

$$\begin{aligned} &\min_{(u;z)\in V_\beta} \|u\|^2_{L_2} + \|z - \eta\|^2_{\mathbb{R}^l_S} \Rightarrow \\ &\min_{(u;z)\in V_\beta} \|(u; z) - (0; \eta)\|^2_{\mathcal{H}}. \end{aligned} \tag{33}$$

Let us define $p \equiv (0; \eta) \in \mathcal{H}$, which is the point in the Hilbert space $\mathcal{H}$ with the absolute minimum value to (33). However, p does not satisfy the dynamics of the system (i.e. $p \notin V_\beta$). The projection of p onto V_β (a point that satisfies the dynamics of the system) will be the point in V_β that minimizes (33), which is also defines the optimal u^* for (4). This projection is done in four steps:

1) Find the subspace V_0 that is parallel to V_β,

$$V_0 = \{(u; z) \mid GKu - z = 0\}.$$

2) Find the subspace $V_0^{\perp}$ that is perpendicular to V_0,

$$V_0^{\perp} = \{(w; \rho) \mid \; < (w; \rho), (u; z) >_{\mathcal{H}} = 0, \; (u; z) \in V_0\}.$$

Which is rewritten by using the fact that we can write the inner product equation as

$$\begin{aligned} &< (w; \rho), (u; z) >_{\mathcal{H}} = 0 && \Leftrightarrow \\ &< w, u >_{L^2} + < \rho, z >_{\mathbb{R}^l_S} = 0 && \Leftrightarrow \\ &< w, u >_{L^2} + < \rho, GKu >_{\mathbb{R}^l_S} = 0 && \Leftrightarrow \\ &< w, u >_{L^2} + < S\rho, GKu >_{\mathbb{R}^l} = 0 && \Leftrightarrow \\ &< w, u >_{L^2} + < K^* G^{\mathsf{T}} S\rho, u >_{L^2} = 0 && \Leftrightarrow \\ &< w + K^* G^{\mathsf{T}} S\rho, u >_{L^2} = 0 && \Leftrightarrow \\ &w + K^* G^{\mathsf{T}} S\rho = 0. \end{aligned}$$

Thus,

$$V_0^{\perp} = \{(w; \rho) \mid w = -K^* G^{\mathsf{T}} S\rho\}.$$

3) Find the affine variety V_p that is parallel to $V_0^{\perp}$ and contains p. The affine variety V_p is $V_0^{\perp} + p$, or

$$\begin{aligned} V_p &= \{(w'; \rho') \mid w' = w, \; \rho' = \rho + \eta, \; (w; \rho) \in V_0^{\perp}\} \\ &= \{(w'; \rho') \mid w' = -K^* G^{\mathsf{T}} S(\rho' - \eta)\}. \end{aligned}$$

4) Find the intersection of V_β and V_p, which is the projection of p onto V_β,

$$\begin{aligned}
&V_\beta \cap V_p && \Leftrightarrow\\
&GKu - z = -G\beta \ \big|_{u=-K^*G^\mathsf{T}S(z-\eta)} && \Leftrightarrow\\
&GK(-K^*G^\mathsf{T}S(z-\eta)) - z = -G\beta && \Leftrightarrow\\
&-G\Gamma G^\mathsf{T}S(z-\eta) - z = -G\beta && \Leftrightarrow\\
&-G\Gamma G^\mathsf{T}Sz + G\Gamma G^\mathsf{T}S\eta - z = -G\beta && \Leftrightarrow\\
&(G\Gamma G^\mathsf{T}S + I)z = G\Gamma G^\mathsf{T}S\eta + G\beta && \Leftrightarrow\\
&z = (G\Gamma G^\mathsf{T}S + I)^{-1}(G\Gamma G^\mathsf{T}S\eta + G\beta).
\end{aligned}$$

Note that $(G\Gamma G^\mathsf{T}S + I)$ is always positive definite.

Therefore, the minimizer $u^*_{x_0}$ to (4) is

$$u^*_{x_0} = -K^*G^\mathsf{T}S((G\Gamma G^\mathsf{T}S + I)^{-1}(G\Gamma G^\mathsf{T}S\eta + G\beta(x_0)) - \eta), \tag{34}$$

remember we described β as a function of x_0 in (30).

2) Second condition: Now, let us derive the an expression for $\hat{J}(u, x_0) - \hat{J}(u, x_0')$. Before we do, let us recall that $x_0 \neq x_0'$ and $\hat{z} = G\hat{x}$. In addition, we will say that $\hat{z}' = G\hat{x}'$. So,

$$\begin{aligned}
&\hat{J}(u, x_0) - \hat{J}(u, x_0') \Leftrightarrow\\
&\int_0^T \|u\|^2 dt + (\hat{z}(T) - \eta)^\mathsf{T}S(\hat{z}(T) - \eta) -\\
&\qquad \int_0^T \|u\|^2 dt + (\hat{z}'(T) - \eta)^\mathsf{T}S(\hat{z}'(T) - \eta) && \Leftrightarrow\\
&(\hat{z}(T) - \eta)^\mathsf{T}S(\hat{z}(T) - \eta) - (\hat{z}'(T) - \eta)^\mathsf{T}S(\hat{z}'(T) - \eta) && \Leftrightarrow\\
&\hat{z}(T)^\mathsf{T}S\hat{z}(T) - \hat{z}'(T)^\mathsf{T}S\hat{z}'(T)^\mathsf{T} - 2\eta^\mathsf{T}S(\hat{z}(T) - \hat{z}'(T)) && \Leftrightarrow\\
&(\hat{z}(T) + \hat{z}'(T))^\mathsf{T}S(\hat{z}(T) - \hat{z}'(T)) - 2\eta^\mathsf{T}S(\hat{z}(T) - \hat{z}'(T)) && \Leftrightarrow\\
&((\hat{z}(T) + \hat{z}'(T))^\mathsf{T}S - 2\eta^\mathsf{T}S)(\hat{z}(T) - \hat{z}'(T)) && \Leftrightarrow\\
&(\hat{z}(T) + \hat{z}'(T) - 2\eta)^\mathsf{T}S(\hat{z}(T) - \hat{z}'(T)).
\end{aligned} \tag{35}$$

Notice that $\hat{J}(u, x_0) - \hat{J}(u, x_0')$ is only guaranteed to be zero for all η values if $\hat{z}(T) - \hat{z}'(T) = 0$.

Let us expand $\hat{z}(T) - \hat{z}'(T)$ to show when this term equals zero. But first we will need an expression for $\hat{x}(t)$, which is estimated by the observer. We defined the express for the observer in (12), let us simplify that expression to

$$\dot{\hat{x}} = \hat{A}\hat{x} + Bu + LCx, \tag{36}$$

where $\hat{A} = (A - LC)$. Therefore,

$$\hat{x}(t) = e^{\hat{A}t}\hat{x}_0 + \int_0^T e^{\hat{A}(t-s)}(Bu(s) + LCx(s))ds.$$

From this we can get

$$\hat{z}(t) = G\left(e^{\hat{A}t}\hat{x}_0 + \int_0^T e^{\hat{A}(t-s)}(Bu(s) + LCx(s))ds\right). \tag{37}$$

Now we plug (37) into $\hat{z}(T) - \hat{z}'(T)$ to get

$$\begin{aligned}
&\hat{z}(T) - \hat{z}'(T) =\\
&G\Bigg(e^{\hat{A}t}\hat{x}_0 + \int_0^T e^{\hat{A}(t-s)}(Bu(s) + LCx(s))ds -\\
&\qquad e^{\hat{A}t}\hat{x}_0 - \int_0^T e^{\hat{A}(t-s)}(Bu(s) + LCx'(s))ds\Bigg) =\\
&G\int_0^T e^{\hat{A}(t-s)}LC(x(s) - x'(s))ds.
\end{aligned}$$

We can expand this further by using the expression for $x(t)$ from (30),

$$\begin{aligned}
&\hat{z}(T) - \hat{z}'(T) =\\
&G\int_0^T e^{\hat{A}(t-s)}LC((\beta(x_0) + Ku) - (\beta(x_0') + Ku))ds =\\
&G\int_0^T e^{\hat{A}(t-s)}LC(\beta(x_0) - \beta(x_0'))ds =\\
&G\int_0^T e^{\hat{A}(t-s)}LCe^{As}(x_0 - x_0')ds =\\
&G\int_0^T e^{\hat{A}(t-s)}LM^*ds(x_0 - x_0').
\end{aligned} \tag{38}$$

We will condense (38) further to

$$\hat{z}(T) - \hat{z}'(T) = G\hat{M}M^*(x_0 - x_0'), \tag{39}$$

where $\hat{M}$ is a linear operator such that

$$\hat{M}(t)[u(t)] = \int_0^t e^{\hat{A}(t-s)}Lu(s)ds. \tag{40}$$

Note that with our assumption on L from (14) we have that

$$\begin{aligned}
&\mathcal{R}(C^\mathsf{T}) \subseteq \mathcal{R}(L) && \Leftrightarrow\\
&\mathcal{R}(M) \subseteq \mathcal{R}(\hat{M}) && \Leftrightarrow\\
&\mathcal{R}(MM^*) \subseteq \mathcal{R}(\hat{M}M^*).
\end{aligned} \tag{41}$$

With the expression in (34), (35), (39), and (41) we have all the parts in place to prove Theorem 1, which we restate below.

Theorem 1 (Learnability): A system is learnable if

$$\mathcal{R}(G^T) \subseteq \mathcal{R}(\Gamma) \cap \mathcal{N}(\Omega)^\perp \text{ and } G \neq 0. \tag{42}$$

Proof: Let us begin with $\mathcal{R}(G^\mathsf{T}) \subseteq \mathcal{R}(\Gamma)$, which implies the following

$$\begin{aligned}
&\mathcal{R}(G^\mathsf{T}) \subseteq \mathcal{R}(\Gamma) && \Leftrightarrow\\
&\mathcal{R}(G^\mathsf{T}) \subseteq \mathcal{R}(K) && \Leftrightarrow\\
&\mathcal{R}(G^\mathsf{T}) \subseteq \mathcal{N}(K^*)^\perp && \Leftrightarrow\\
&\exists\zeta \ s.t.\ K^*G^\mathsf{T}S\zeta \neq 0 && \Leftrightarrow\\
&\exists x_0, x_0' \text{ s.t. } u^*_{x_0} \neq u^*_{x_0'}.
\end{aligned}$$

Next, $\mathcal{R}(G^{\mathsf{T}}) \subseteq \mathcal{N}(\Omega)^{\perp}$ implies the following

$$\begin{aligned}
&\mathcal{R}(G^{\mathsf{T}}) \subseteq \mathcal{N}(\Omega)^{\perp} && \Leftrightarrow \\
&\mathcal{R}(G^{\mathsf{T}}) \subseteq \mathcal{R}(\Omega^{\mathsf{T}}) && \Leftrightarrow \\
&\mathcal{R}(G^{\mathsf{T}}) \subseteq \mathcal{R}(MM^{*}) && \Rightarrow \\
&\mathcal{R}(G^{\mathsf{T}}) \subseteq \mathcal{R}(\hat{M}M^{*}) && \Rightarrow \\
&\mathcal{R}(\hat{M}M^{*}) \nsubseteq \mathcal{R}(G^{\mathsf{T}})^{\perp} && \Leftrightarrow \\
&\mathcal{R}(\hat{M}M^{*}) \nsubseteq \mathcal{N}(G) && \Leftrightarrow \\
&\exists \zeta \ s.t. \ G\hat{M}M^{*}\zeta \neq 0 && \Leftrightarrow \\
&\exists x_0, x_0' \ s.t. \ G\hat{M}M^{*}(x_0 - x_0') \neq 0 && \Leftrightarrow \\
&\exists x_0, x_0' \text{ s.t. } \hat{z}(T) - \hat{z}'(T) \neq 0 && \Leftrightarrow \\
&\exists x_0, x_0', u \text{ s.t. } \hat{J}(u, x_0) - \hat{J}(u, x_0') \neq 0 && \Leftrightarrow \\
&\exists x_0, x_0', u \text{ s.t. } \hat{J}(u, x_0) \neq \hat{J}(u, x_0').
\end{aligned}$$

Thus,

$$\begin{aligned}
&\mathcal{R}(G^{\mathsf{T}}) \subseteq \mathcal{R}(\Gamma) \ \cap \ \mathcal{N}(\Omega)^{\perp} \text{ and } G \neq 0 \Rightarrow \\
&\exists x_0, x_0', u \text{ s.t. } u^{*}_{x_0} \neq u^{*}_{x_0'} \text{ and} \\
&\exists x_0, x_0', u \text{ s.t. } \hat{J}(u, x_0) \neq \hat{J}(u, x_0'),
\end{aligned}$$

which satisfies the two conditions for learnability. ■