Learning the Inverse Dynamics of a Robot Arm by Auto-Imitation

Karl Theodor KALVERAM
Cybernetical Psychology,
Institute of Experimental Psychology
University of Duesseldorf
40225 Duesseldorf, Germany
e-mail: kalveram@uni-duesseldorf.de

and

André SEYFARTH
Locomotion Laboratory,
Institute of Sport Science
University of Jena
07743 Jena, Germany
e-mail: oas@uni-jena.de

ABSTRACT

Auto-imitation is a type of unsupervised learning which enables a controller to quickly acquire the inverse dynamics of a plant. The basics of the method are demonstrated using a one-segment robot arm. If embedded into a learner-operator schema, the inverse model can be updated while the robot performs purposive or even arbitrary movements. So, the proposed algorithm would be an alternative to supervised learning, for instance via feedback-error. It is presumed that auto-imitation-like learning governs especially social learning in humans.

Keywords: inverse adaptive control, sensorimotor learning, inverse dynamics model, auto-imitation

1. INTRODUCTION

Control of human goal directed rotational forearm movements is commonly assumed to include feedforward through an internal inverse model of the arm’s dynamics, and negative feedback realized either through appropriately processed peripheral measurements [13,15,18], or through a tunable torsion spring [10,11]. The present paper focuses on how an artificial neural network can acquire the inverse dynamics with adequate precision. Technically, two learning algorithms come into consideration: (a) Supervised learning, for instance back-propagation of feedback-error [14], and (b) non-supervised learning through auto-imitation [9]. Both algorithms have been tested by simulation as principally feasible. Feedback-error learning [12,14] is a sequential learning rule which, however, revealed as time consuming and “loose”. In contrast, auto-imitation is a non-error based learning algorithm which is applicable in parallel, and reveals as precise and fast. The root idea of this learning algorithm has been conceived to characterize sensorimotor learning as acquiring and extending the “sensorimotor self and its reverse” [3] in the framework of the re-afference principle [1]. Formally, the algorithm can be described as “direct inverse modeling” [2]. The name “auto-imitation” used later on for this strategy of “self-extension” [4,7] refers to the procedure, through which the controller explores and applies the dynamical properties of a part of a machine, respectively of a limb of the own body: stimulating the limb with arbitrary proof loads (torques, forces), observing the reactive movements, and casting both into a law which predicts the proof loads from the reactions. The thus accumulated knowledge enables the controller to select that load which, when exerted to the limb, realizes a predefined (desired) movement. So far, a purposively performed movement looks like an imitation of a previously self-generated, but non-purposive movement.

The method, however, was generally considered as not practicable to be used in the organismic control of limb movements, because the re-wiring of the respective neural network when switching from learning to operating and vice versa was assessed as too difficult or even impossible to be applicable in biological systems [2,14]. As Fig. 2 shows, this objection has been overcome by introducing the learner-operator concept, in which the functions of learning and operating are assigned to different neural networks complementing each other [9], what makes re-wiring unnecessary.
Though applicable also to multi-jointed levers or arms, here we demonstrate the method using a one-segment lever. At first we give a description of the robot used for the demonstration, then we explicate inverse control and auto-imitation, followed by the description of the outcome if auto-imitation and inverse control are applied to the robot. Finally it is suggested to take auto-imitation together with inverse control as a model to explain also higher order types of behavior, for instance those in the social domain.

2. METHODS

Hardware

The ADAM (Adaptable Machine) device is a simple robot test bed built to compare the effectiveness of different forcing functions applied to a lever. The version used in the current investigation is illustrated in Fig. 1 and comprises a DC-motor (Maxon RE30 310007) without gear, the shaft being aligned horizontally, and a freely rotating lever of about 0.27m length fixed at the front end’s shaft, which is considered a pointer. The motor is driven by an analogue power amplifier (Mattke MAR 24/6z) operating in the voltage-to-current mode. Hence, the current indicates the total torque $Q$ exerted on the lever via the motor. SIMULINK Realtime software (fixed step sample time 1ms, solver Dormand-Prince) organizes through AD- and DA-conversion (interface: Meilhaus ME 2600) the control of the power amplifier. A tachometer (Mattke T505) fixed to the other end of the shaft measures the actual angular velocity. Derivative and integration of this signal via SIMULINK software provide the actual angular acceleration and position.

![Figure 1: Robotic arm used for the hardware test. Left: Front view. Right: Side view with adumbrated wiring. The “hand” is made of coarse cellular foam (usage with courtesy of Thomas Schinauer).](image)

Inverse control and auto-imitation

Fig. 2 depicts, how operating the arm via inverse control, and learning the arm’s inverse dynamics via auto-imitation, can be arranged simultaneously. The upper part of the Fig. 2 sketches the arm (plant) and its control: The arm is regarded as an ideal gravity pendulum with viscous damping and friction around the pivot, which is governed by the differential equation

$$J\ddot{\phi} + B_\phi \dot{\phi} + B_f \sin(\phi_o - \phi) = Q,$$

where $J$ denotes the arm’s inertia, $B_\phi$ the coefficient of viscous damping, and $B_f$ the coefficient of friction. $M = -m\cdot d\cdot g$ ($m$: mass; $d$: the distance of the centre of mass from the pivot, $g = 9.81$N/kg) and $\phi_o$ (offset, angular deviation of the rest position from the gravitational reference) relate to the gravitational induced torque. The arm’s input is the torque $Q$ around the pivot induced by control (torques $Q_{in}$ and $Q_{fb}$), its output the actual angular kinematics (acceleration $\dot{\phi}$, velocity $\phi$, and position $\phi$). So, considered physically, the arm is seen in the causal direction. The controller’s task is to find that torque which makes the arm’s output matching given desired values. For this reason, the pattern generator transforms in each step a discrete angular goal into desired angular kinematics (desired trajectories of angular position, velocity and acceleration) which smoothly interpolate...
between the previous and the current goal. These three streams of values are put into the arm’s inverse dynamics model which computes that trajectory of torques $Q_{ff}$ which, when applied to the arm, realizes the desired kinematics values. That is to say, the inverse dynamics model represents the arm in an anti-causal manner. In engineering, this type of control is also known as ‘computed torque’ [16] or ‘inverse control’ [17]. The inverse dynamics model gets its state input (angular velocity and position) as ‘state feedforward’ from the pattern generator [10,11] and not, as usual, as ‘state feedback’ from the periphery. Hence, the pattern generator acts through the inverse dynamics model as if it would impose its own movement imagination on the arm which is considered as a part of the environment. Torques unforeseen by the inverse dynamics model (perturbations) may be added upon $Q$ via the input $Q_X$. The thus provoked deviations between the desired and the actual angular positions are processed by proportional negative feedback control (gain factor $K$). The hereby generated torque $Q_{fb}$ vanishes if the inverse model is perfect and disturbances are absent. Systematic deviations, however, would demand an update of the inverse dynamics model.

**Figure 2**: Auto-imitation embedded into a learner-operator scheme. The inverse model is split into two segregated networks, named the ‘learner’ and the ‘operator’, which have equal architecture, but different functions. The “learner” acquires the parameters (that are the synaptic weights of the underlying artificial neural network) of the inverse model through auto-imitation, but does not have an output to control the plant (here the arm). Training of the network only requires arbitrary teaching inputs $Q$ provided anyhow, for instance even by a “blind teacher” [3,5]. The “operator” gets the parameters from the learner and exerts control, but does not have a teaching input to update the parameters. Notice that the diagram is functionally oriented. Therefore, the physiological realization of the concept is left open.

The lower part of Fig. 2 sketches auto-imitative learning and re-adaptation of the arm’s inverse dynamics model. The basic idea is that learner and operator rely on the same formula to physically describe the arm’s dynamics (see Eq.1). If angular positions $\phi$ are small enough, $\sin \phi$ can be replaced by $\phi - \phi^3/6$. In so doing and omitting terms expected to get negligible, after some rearranging Eq. (1) becomes

$$M \ddot{\phi} + J \dot{\phi} + B J \phi + B D \text{ sign } \phi - M \dot{\phi} + \frac{M}{6} \phi^3 = Q.$$  

This means, that the right side of Eq. (2) is considered as a function $f$ in four variables which is approximated by an abbreviated power series in these variables,

$$f(x_1, x_2, x_3, x_4) = \sum w_m \cdot x_{1i} \cdot x_{2j} \cdot x_{3k} \cdot x_{4l},$$  

whereby $x_1 = \dot{\phi}$, $x_2 = \ddot{\phi}$, $x_3 = \text{sign}\dot{\phi}$, $x_4 = \phi$. $i, j, k, l$ denote non-negative integers, and the $w_m$ (m=0,1,2,...) represent coefficients (weights) to be determined. Regarding Eq. (3) as power series truncated by the condition $i+j+k+l \leq 3$ (which makes Eq. (3) a power series of order 3), and additionally discarding a number of terms, Eq.(3) may be written as
\[ w_0 + w_1 \cdot x_1 + w_2 \cdot x_2 + w_3 \cdot x_3 + w_4 \cdot x_4 + w_5 \cdot x_5^3 = Q. \]  

(4)

Compared to Eq. (2), this suggests to interpret the weights as \( w_0 = M \phi_e, \) \( w_1 = J, \) \( w_2 = B_d, \) \( w_3 = B_f, \) \( w_4 = M, \) and \( w_5 = M/6, \) provided Eq. (2) governs the arm’s movements precisely. The appearance of Eq. (4) is that of an inverse model of the arm’s dynamics, if \( x_1 = \ddot{\phi}, x_2 = \dot{\phi}, x_3 = \text{sign}\phi, \) \( x_4 = \phi \) are considered as inputs and \( Q \) as the output.

So, the inverse model implemented in the operator can tentatively be written as

\[ w_0 + w_1 \ddot{\phi} + w_2 \dot{\phi} + w_3 \text{sign}\phi + w_4 \phi + w_5 \phi^3 + w_6 \phi \dot{\phi} + w_7 \phi \ddot{\phi} \approx Q, \]

(5)

whereby the last two terms are added to check the effect of surplus terms on the value of the weights.

The computations described in Eq. (5) can be mimicked by a simple artificial neural network, the Power Net [6], using the form adapted to the present paper, which is depicted in Fig. 3.

![Figure 3: Power Net of order 3 representing the inverse dynamics model given in Eq. (5). The 1st layer nodes convey the input variables to the nodes of the 2nd (middle) layer, where they, if required, are first raised to the power of \( i \), and then multiplied. The power is indicated at the respective synapses marked by open triangles. The node of the 3rd (output) layer then sums the weighted outputs of the 2nd layer nodes. The weights \( w_m \) which are to be determined are marked with filled triangles. The sign-function is provided by an additional node in the 2nd layer.](image)

The weights \( w_m \) are determined through the “auto-imitative process” outlined in Fig. 2: Firstly, at \( n \) points of time in an ongoing movement, \( n \) samples, each consisting of the actual acceleration \( \ddot{\phi} \), velocity \( \dot{\phi} \), position \( \phi \), and torque \( Q \), are collected. Then, as shown in Fig. 4, the values \( a_1, \ldots, a_7 \) of the seven variables occurring in Eq. (5) are inserted into matrix \( A \), and the measurements of \( Q \) into \( B \). \( a_0 \) is always set to 1, and \( B \) has one column. The Matlab command \( W = B \cdot A \) then gets the solution via the least squares method.

![Figure 4: Arrangement of \( n \) samples of measurements \( a_1, \ldots, a_7 \) of the left-side variables of Eq. (5), and the related right-side torque values \( Q \) into matrices \( A \) and \( B \). \( a_0 \) is always set to 1, and \( B \) has one column. The MATLAB command \( W = B \cdot A \) then provides a least squares approximation of the weights.](image)
Application of the concept to the robot arm and results

In Fig. 5, the sequence of discrete angular goals put into the pattern generator (see Fig. 2) is indicated by the broken gray (yellow) line, and the corresponding continuous desired angular positional trajectory which interpolates between the steps of the goals by the thick gray (cyan) curve. The actual angular trajectory produced by the robot arm is indicated by the thin black line. The gain $K$ in the positional negative feedback loop was set to 0.5Nm/rad. In a first movement of about 10s, all weights were set to zero, so control was limited to pure negative feedback control. As can be seen in Fig. 5a, the actual angular positions then deviate from the desired ones considerably. To yield the weights, at the end of this period the recorded kinematics and torque values were, according to Fig. 4, processed by the least squares method. The weights were then inserted into the operator’s inverse dynamics model. This results in a nearly perfect matching of the actual with the desired positional angular trajectory (see Fig. 5b) in the next run.

![Figure 5a](image1.png)

![Figure 5b](image2.png)

**Figure 5:** The robot arm tested in the framework of the learner-operator concept. **a:** first run with negative feedback control only. **b:** second run additionally using the inverse model, updated by the parameters which are auto-imitatively determined in the first run. See also text.

In each ongoing movement the weights of the inverse model can be updated to improve the controller’s performance in the next movement. The weights thus obtained in seven successive runs and their means are listed in tab.1. After the first run, essential further improvements are not achievable, but obviously also not necessary.

Though the weights in tab.1 exhibit some fluctuations, the values seem physically plausible. In any case, the weights referring to a run produce, if inserted into the operator’s network, an actual positional curve like that shown in Fig. 5b. From $w_0$ and $w_4$ one gets the angular offset $\phi_1=0.0258\text{rad} \cong 1.5^\circ$ which is probably due to an imperfect offset adjustment of the power amplifier. $w_2$ reflecting the coefficient of viscous damping $B_D$ is zero, whereas $w_3$ representing the coefficient $B_F$ of friction force is greater than zero. This is in line with the construction of the motor which uses laminated brushes for commutation, what produces friction, but no damping. The small values of $w_6$ and $w_7$ further indicate that the corresponding terms in Eq. (5) are indeed dispensable.
Table 1: Weights of the inverse model determined by auto-imitation in seven subsequent runs.

<table>
<thead>
<tr>
<th>weight</th>
<th>run 1</th>
<th>run 2</th>
<th>run 3</th>
<th>run 4</th>
<th>run 5</th>
<th>run 6</th>
<th>run 7</th>
<th>mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>( w_0 = B q_0 )</td>
<td>0.0033</td>
<td>0.0032</td>
<td>0.0031</td>
<td>0.0031</td>
<td>0.0032</td>
<td>0.0029</td>
<td>0.0021</td>
<td>0.0030</td>
</tr>
<tr>
<td>( w_1 = J )</td>
<td>0.0019</td>
<td>0.0020</td>
<td>0.0019</td>
<td>0.0019</td>
<td>0.0019</td>
<td>0.0019</td>
<td>0.0019</td>
<td>0.0019</td>
</tr>
<tr>
<td>( w_2 = B \beta )</td>
<td>0.0000</td>
<td>-0.0001</td>
<td>-0.0001</td>
<td>-0.0002</td>
<td>-0.0001</td>
<td>0.0002</td>
<td>0.0001</td>
<td>-0.0000</td>
</tr>
<tr>
<td>( w_3 = B \gamma )</td>
<td>0.0155</td>
<td>0.0159</td>
<td>0.0174</td>
<td>0.0178</td>
<td>0.0178</td>
<td>0.0137</td>
<td>0.0147</td>
<td>0.0161</td>
</tr>
<tr>
<td>( w_4 = M )</td>
<td>0.1150</td>
<td>0.1167</td>
<td>0.1153</td>
<td>0.1173</td>
<td>0.1156</td>
<td>0.1166</td>
<td>0.1165</td>
<td>0.1161</td>
</tr>
<tr>
<td>( w_5 = M/6 )</td>
<td>-0.0168</td>
<td>-0.0193</td>
<td>-0.0171</td>
<td>-0.0246</td>
<td>-0.0226</td>
<td>-0.0196</td>
<td>-0.0195</td>
<td>-0.0199</td>
</tr>
<tr>
<td>( w_6 )</td>
<td>-0.0004</td>
<td>-0.0005</td>
<td>-0.0004</td>
<td>-0.0006</td>
<td>-0.0004</td>
<td>-0.0004</td>
<td>-0.0004</td>
<td>-0.0004</td>
</tr>
<tr>
<td>( w_7 )</td>
<td>-0.0001</td>
<td>-0.0001</td>
<td>-0.0001</td>
<td>-0.0001</td>
<td>-0.0001</td>
<td>-0.0001</td>
<td>-0.0001</td>
<td>-0.0001</td>
</tr>
</tbody>
</table>

3. DISCUSSION

Auto-imitation reveals as a technique for unsupervised learning of the inverse dynamics of a robot arm. The technique is quick, 10s suffice to program the respective network. Though demonstrated using an one-segmented lever, the algorithm is expected to be applicable at least also to a two-segmented real arm, a conclusion we extrapolate from the results of simulations [5].

In the present paper we use a simultaneous rule to acquire the weights of the artificial neural network representing the inverse dynamics. This rule is based on the least squares method applied to a great many of observed kinematics and torque data stored in memory. For auto-imitation, however, also a sequential rule is forthcoming, as could be demonstrated with a simulated two-segment arm, but a hardware test for this instance is yet lacking. The sequential rule is describable as modified Hebbian learning [8], what links auto-imitation with organismic learning. This leads to the motivation underlying the presented investigation, namely to test, whether the envisaged type of acquisition of an inverse dynamics model is applied also in human motor control. Regarding the test trilogy [11] proposed to tackle such questions, simulation tests have already been performed, and a hardware test has been presented in this paper. The behavioral comparison test is pending; its results will be published elsewhere.

Humans explore the physical properties not only of their own limbs, but also of the external environment. Here, the same technique as described above can be applied: Arbitrary, but scaled, test signals have to be put into the part of the environment to be explored, the sensory signals encoding the related environmental changes have to be observed, and in an auto-imitative process both types of signals have to be connected and condensed into a general law. The law then can be used to actively shape the environment towards a desired state. In everyday language, we call this ‘learning to handle a tool’ and ‘using a tool purposively’, in science and engineering ‘making an experiment’ and ‘solving a technical problem’. Considered psychologically, in interpersonal contacts each person is the other person’s social environment. In the framework of the so emerging intertwined auto-imitative social loops, each partner in a pair gets knowing and manipulating the other. So, the rules governing sensory motor behavior can be consulted how to investigate and explain also higher order human behavior.

Acknowledgements: This study was supported by grants SE1042/2 and KA417/18 of the German Research Foundation (DFG).

---

1 In this test trilogy, the simulation test probes whether a concept is logically sound. The hardware test employs a machine (robot) to check whether a successfully simulated theoretical concept is practicable also under real world conditions. The behavioral comparison test compares human data statistically with simulated data generated by checked models. With this test procedure it is possible to assess, which of several competing models is nearest to human data.
4. REFERENCES


