LETTER Communicated by Rajesh P. N. Rao

ISO Learning Approximates a Solution to the Inverse-Controller Problem in an Unsupervised Behavioral Paradigm

Bernd Porr
bp1@cn.stir.ac.uk
Department of Psychology, University of Stirling, Stirling FK9 4LA, Scotland

Christian von Ferber
ferber@physik.uni-freiburg.de
Theoretical Polymer Physics, Freiburg University, 79104 Freiburg, Germany

Florentin Wörgötter
worgott@cn.stir.ac.uk
Department of Psychology, University of Stirling, Stirling FK9 4LA, Scotland

In "Isotropic Sequence Order Learning" (pp. 831–864 in this issue), we introduced a novel algorithm for temporal sequence learning (ISO learning). Here, we embed this algorithm into a formal nonevaluating (teacher free) environment, which establishes a sensor-motor feedback. The system is initially guided by a fixed reflex reaction, which has the objective disadvantage that it can react only after a disturbance has occurred. ISO learning eliminates this disadvantage by replacing the reflex-loop reactions with earlier anticipatory actions. In this article, we analytically demonstrate that this process can be understood in terms of control theory, showing that the system learns the inverse controller of its own reflex. Thereby, this system is able to learn a simple form of feedforward motor control.

1 Introduction

In our companion article in this issue, "Isotropic Sequence Order Learning," we introduced a novel, linear, and unsupervised algorithm for temporal sequence learning, which we called ISO learning. ISO learning has the special feature that all sensor inputs are completely isotropic, which means that any input can drive the learning behavior. We used the algorithm to generate robot behavior by means of sensor inputs and motor actions. While the organism transforms sensor events into motor actions, the environment passively performs the opposite and forms together with the organism a closed sensor-motor feedback loop system. Here, we explain the system-theoretical consequences of this.

ISO learning is completely unsupervised, and the output is self-organized. Unsupervised temporal sequence learning, however, usually leads—without additional measures taken—to rather undesired situations for the organism since it can learn arbitrary behavioral patterns. A fixed reflex loop prevents arbitrariness by defining an initial behavioral goal (Verschure & Voegtlin, 1998). A reflex, however, is a typical *reaction*, which will always occur only after its eliciting sensor event (Wolpert & Ghahramani, 2000). ISO learning leads to the functional elimination of the reflex loop in using predictive sensorial cues and generating appropriate anticipatory actions to prevent the triggering of the reflex. We will see that these qualitative observations can be embedded in a control theoretical framework.

In the field of control theory, a reflex loop is represented by a fixed feedback loop (McGillem & Cooper, 1984; D'Azzo, 1988; Nise, 1992; Palm, 2000). Feedback loops try to maintain a desired state by comparing the actual input values with a predefined state and adjusting the output so that the desired state is optimally maintained. The main advantage of a feedback loop is that the controller needs only very limited knowledge about the relation between input and output (the environment). Consider the typical example of a thermostat-controlled central heating system. There, it is necessary only to measure the temperature at the thermostat and use this to control the furnace; it is not necessary to know how much fuel needs to be burned to get a certain temperature increase. Even this is not enough to control the heating, because the temperature increase also depends on the existing inside-outside temperature gradient and maybe on more elusive parameters. In general, only in idealized situations does there exist sufficient prior knowledge to control a system without feedback, thus, by means of pure feedforward control. The central advantage of such an (ideal) feedforward controller, however, is that it acts without the feedback-induced delay. The sometimes fatally damaging sluggishness of feedback systems makes this a highly desirable feature. As a consequence, engineers try to replace feedback controllers with their equivalent feedforward controllers wherever possible, thereby trying to solve the famous inverse controller problem (Nise, 1992).

In this study, we analytically prove that ISO learning approximates the inverse controller of a reflex when embedded in a behavioral situation where the reflex represents the reference for self-organized predictive learning.

The article is organized in the following way. Very briefly we summarize the main equations from our other article in this issue. Then we introduce the necessary terminology from control theory by means of discussing the reflex-loop situation. After that, we show which shape a transfer function must take in order to approximate the inverse controller of the reflex. In the next step, we demonstrate that the set of functions used in ISO learning can indeed approximate this transfer function. Finally, we show why the actual learning process does converge into the correct solution.

2 The ISO Learning Algorithm: A Brief Summary

The system consists of $N+1$ linear filters h receiving inputs x and producing outputs u. The filters connect with corresponding weights ρ to one output unit v (see Figure 1). The output $v(t)$ in the time domain and its transformed equivalent $V(s)$ in the Laplace domain are given as

$$v(t) = \rho_0 u_0 + \sum_{k=1}^{N} \rho_k u_k \leftrightarrow V(s) = \rho_0 U_0 + \underbrace{\sum_{k=1}^{N} \rho_k U_k}_{H_v}. \tag{2.1}$$

The transfer functions h shall be those of bandpass filters, which transform a δ-pulse input into a damped oscillation. They are specified in the time and in the Laplace domain by

$$h(t) = \frac{1}{b} e^{at} \sin(bt) \leftrightarrow H(s) = \frac{1}{(s+p)(s+p^*)}, \tag{2.2}$$

where p^* represents the complex conjugate of the pole $p = a + ib$, with

$$a := \mathrm{Re}(p) = -\pi f/Q, \quad b := \mathrm{Im}(p) = \sqrt{(2\pi f)^2 - a^2}. \tag{2.3}$$

f is the frequency of the oscillation and Q the damping characteristic. Learning takes place according to

$$\frac{d}{dt}\rho_j = \mu u_j v' \qquad \mu \ll 1, \tag{2.4}$$

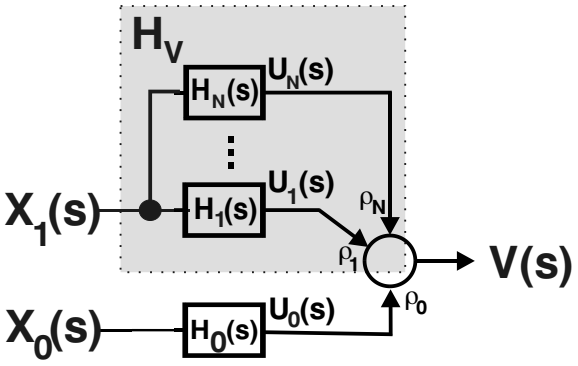

Figure 1: The neuronal circuit in the Laplace domain. The shaded area marks the connections of the weights ρ_k with $k \geq 1$ onto the neuron. The overall contribution from these input is called H_V.

where v' is the temporal derivative of v. (For a comparison of ISO learning with other models for temporal sequence learning, see appendix B in the companion article in this issue.) Note that μ is very small. The integral form of this learning rule is in the Laplace domain given by

$$\Delta\rho_j = \frac{\mu}{2\pi}\int_{-\infty}^{\infty} -i\omega V(-i\omega)U_j(i\omega)\,d\omega \quad \text{with } U = XH. \tag{2.5}$$

Note that we use indices k to denote outputs (e.g., when associated with v), while indices j denote inputs (e.g., associated with u). (See the companion article for a complete description of the ISO learning algorithm and its properties.)

3 Analytical Treatment of the Closed-Loop Condition

3.1 Reflex Loop Behavior. Every closed-loop control situation with negative feedback has a so-called desired state; the goal of the control mechanism is to maintain (or reach) this state as best and fast as possible. In our model, we assume that the desired state of the reflex feedback loop is unchanging and defined by the properties of the reflex loop. We define it as $X_0 = 0$. First, we discuss the system without learning. Figure 2a shows the situation of a learner embedded into a very simple but generic (i.e., unspecified) formal environment, which has a transfer function P_0. This learner is able to react to an input only by means of a reflex.

A possible set of signals that can occur in such a system is shown in Figure 2b. First, the disturbance signal d deviates from zero, then the input x_0 senses this change $x_0 \neq 0$, and finally the motor output v can generate a reaction in order to restore the desired state $x_0 = 0$. Thus, there is always a reaction delay in such a system.

3.2 Augmenting the Reflex by Temporal Sequence Learning. In this section, we show that the ISO learning algorithm can approximate the inverse controller of the reflex. Figure 3 shows how the same disturbance D elicits a sequence of sensor events: it enters the outer loop, arriving at X_1 filtered by the environment (P_1), while it arrives at X_0 only after a delay T. The goal of learning is to generate a transfer function H_v, which compensates for the disturbance. The inner structure of H_v given by the ISO learning setup is depicted by Figure 1. The environmental transfer function P_{01} closes the outer loop.

3.2.1 General Condition. The reflex loop defines the goal of the feedforward controller: that there should always be zero input at X_0. Thus, first we must show what shape the transfer function of the predictive pathway H_v (see Figures 1 and 3) takes when we assume that $X_0 = 0$ holds. This is the necessary condition, which needs to be obeyed in order to obtain an

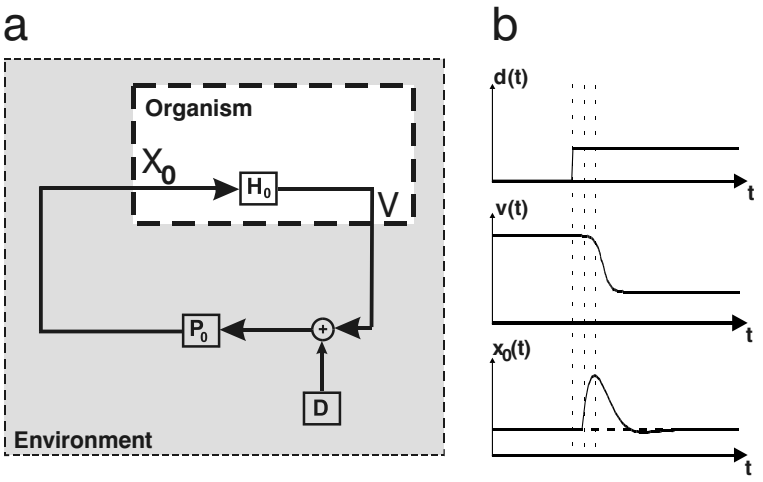

Figure 2: (a) Fixed reflex loop. The organism transfers a sensor event X_0 into a motor response V with the help of the transfer function H_0. The environment turns the motor response V again into a sensor event X_0 with the help of the transfer function P_0. In the environment, there exists the disturbance D, which adds its signal at $\oplus$ to the reflex loop. (b) Possible temporal signal shapes occurring in the reflex loop when a disturbance $d \neq 0$ happens. The desired state is $x_0 := 0$. The disturbance d is filtered by P_0 and appears at x_0 and is then transferred into a compensation signal at v, which eliminates the disturbance at $\oplus$.

appropriate H_v. It generally applies regardless of the learning algorithm used.

In the following, we omit the function argument s where possible. Then we can write

$$X_0 = P_0[V + De^{-sT}] \tag{3.1}$$

as the reflex pathway and

$$X_1 = \frac{P_1 D + P_1 P_{01} X_0 H_0}{1 - P_1 P_{01} H_V} \tag{3.2}$$

$$H_V = \sum_{k=1}^{N} \rho_k H_k \tag{3.3}$$

as the predictive pathway (see Figure 3). Eliminating X_1 and V, we get

$$X_0 = e^{-sT} D + H_V \frac{P_1 D + P_1 P_{01} X_0 H_0}{1 - P_1 P_{01} H_V}. \tag{3.4}$$

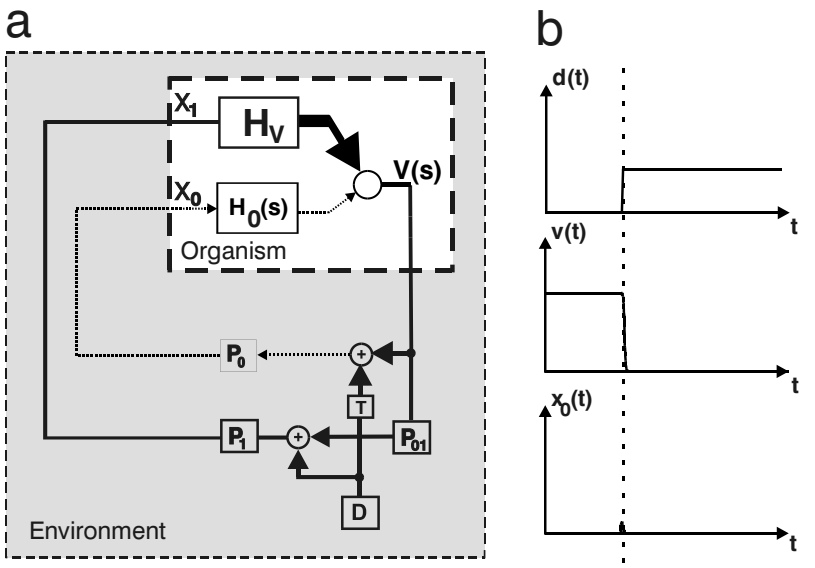

Figure 3: Schematic diagram of the augmented closed-loop feedback mechanism, which now contains a secondary loop representing ISO learning. (a) H_0 and P_0 form the inner feedback loop shown in Figure 2. The new aspect is the input line X_1, which gets its signal via transfer function P_1 from the disturbance D. The inner feedback loop receives a delayed version (T) of the disturbance D. The adaptive controller H_V has the task to use the signal X_1, which is earlier than x_0, and thus "predicts," the disturbance D at X_0, to generate an appropriate reaction at V to prevent a change at X_0. (b) A schematic timing diagram for the situation after successful learning when a disturbance has occurred. The output v sharply coincides with the disturbance d and prevents a major change at the input x_0.

Solving for $X_0 = 0$ leads to

$$H_V = \sum_{k=1}^{N} \rho_k H_k \tag{3.5}$$

$$= -\frac{P_1^{-1} e^{-sT}}{1 - P_{01} e^{-sT}}. \tag{3.6}$$

The transfer function H_V is the overall transfer function of the predictive pathway. Equation 3.5 demands that the weights ρ_k should be adjusted in such a way that equation 3.6 is obtained at the end of learning.

Equation 3.6 requires interpreting. First, we consider the numerator and remember that the learning goal is to achieve $X_0 = 0$. This requires com-

pensating the disturbance D. The disturbance, however, enters the organism only after having been filtered by the environmental transfer function P_1. Thus, compensation of D requires undoing this filtering by the term P_1^{-1}, which is the inverse transfer function of the environment (hence, "inverse controller"). The second term e^{-sT} in equation 3.6 compensates the delay between the signal in X_1 and that at X_0, when the disturbance actually enters the inner feedback loop.

Now we discuss the relevance of the denominator, showing that it can be generally neglected. Transfer functions are fully described by their poles and zero crossings. Poles very strongly affect the behavior of a system, while zero crossings are phase factors, which do not alter its general transfer characteristic (Stewart, 1960; Blinchikoff, 1976; McGillem & Cooper, 1984; Terrien, 1992; Palm, 2000). As a consequence, following methods from control theory, any transfer function may be reduced to only those terms that contain poles or zero crossing by neglecting all other components (Sollecito & Reque, 1981; Nise, 1992).

Thus, we rewrite equation 3.6 as

$$H_V = -P_1^{-1} e^{-sT} \frac{1}{1 - P_{01} e^{-sT}} \tag{3.7}$$

and analyze it to determine if the second term produces additional poles for H_V. This would happen if $1 - P_{01}e^{-sT} = 0$ holds, which is equivalent to $P_{01} = e^{sT}$. The term e^{sT}, however, is meaningless; it represents a time-inverted delay, and thus an entity that violates causality.

As a result, there are no additional poles for H_V, and in the following we are allowed to set $P_{01} := 0$ without loss of generality, thereby neglecting only possible changes in phase relationships. Thus, the behavior of H_V is apart from phase terms entirely determined by[1]

$$H_V = P_1^{-1} e^{-sT}. \tag{3.8}$$

Equation 3.8 represents the necessary condition for the learning, and we ask in the next two sections if our specific algorithm is sufficient to achieve this.

3.2.2 Solutions in the Steady-State Case $X_0 = 0$. Here we show by construction that for one resonator, there already exists a solution that approx-

[1] Readers who are less familiar with control theory may find it useful to think about P_{01} in a different way. P_{01} represents how the environmental transfer of the reaction of the system will influence the sensor X_1. Many times this influence is plainly zero from the beginning (or the connecting path can be decoupled by an appropriate system design). For example, for a predictively acting external temperature sensor X_1, the change of the temperature of the environment due to the heating of a room is totally insignificant.

imates equation 3.8 to the second order. Results for a fourth-order approximation have been numerically obtained, showing that the approximation continues to improve.

Thus, first we limit the discussion to the case of only two resonators, H_0 and H_1 (i.e., $N = 1$). The case with more resonators will be reintroduced at the end of this section. We specify which parameters the resonator H_1 in the outer loop has in order to satisfy the learning goal. At first, we set $P_1 = 1$, looking at the case when the environment does not alter the shape of the disturbance (but see below).

Considering equation 3.8, we have to solve

$$-e^{-sT} = \rho_1 H_1. \tag{3.9}$$

The resonator H_1 has two parameters, $f_1 = 1/T_1$ and Q_1, and together with its weight ρ_1, we are looking for three parameters to solve this equation.

The left-hand side of equation 3.9 can now be developed into a Taylor series,

$$-\frac{1}{e^{sT}} = \frac{-1}{1 + sT + \frac{1}{2}s^2T^2 + \cdots} \approx \frac{-2T^{-2}}{2T^{-2} + 2sT^{-1} + s^2}, \tag{3.10}$$

and the right-hand side of equation 3.9 has to be explicitly written out according to equations 2.2 and 2.3:

$$\rho_1 H_1(s) = \frac{\rho_1}{(s+p)(s+p^*)} = \frac{\rho_1}{\underbrace{pp^*}_{(2\pi f_1)^2} + s\underbrace{(p+p^*)}_{\frac{-2\pi f}{Q_1}} + s^2}. \tag{3.11}$$

We can now compare the coefficients of equation 3.10 with equation 3.11 and get for the parameters,

$$\rho_1 = -\frac{2}{T^2}, \quad f_1 = \pm\frac{1}{\pi T\sqrt{2}}, \quad Q_1 = \sqrt{\frac{1}{2}}. \tag{3.12}$$

This result shows that for all T, there exists a resonator H_1 with a weight ρ_1, which approximates e^{-sT} to the second order.

The result for f can be interpreted in the context of the companion article in this issue. We remember that $X_0 = 0$ and hence $V = X_1 H_1$. If we consider pure δ-pulse input at X_1, as in the simulations in the companion article, we receive the impulse response of the resonator $h_1(t)$ at the output; thus:

$$v(t) = \rho_1 \frac{1}{b_1}\sin(b_1 t)e^{-a_1 t} \tag{3.13}$$

$$= \rho_1 T \sin\left(\frac{t}{T}\right)e^{-\frac{t}{T}} \tag{3.14}$$

$$= -\frac{2}{T}\sin\left(\frac{t}{T}\right)e^{-\frac{t}{T}}. \tag{3.15}$$

This function has its maximum at $t^{(2)}_{\max} = T\text{atan}(1)$. We can assume that this is approximately equal to $t^{(2)}_{\max} \approx T$.[2] This, however, would be indicative of a response maximum that occurs exactly at the moment where the input x_0 is to be expected. We refer readers to Figure 5 in the companion article, where this type of behavior has been observed in the simulations. We found that during learning, the output always has its first maximum at the location where x_0 occurs (or would have occurred). The strength of the resonator response, equation 3.14, is determined by the weight ρ_1, which is adjusted in a way that the resulting integral (see equation 3.15) becomes $\int_0^\infty v(t)\,dt = -1$ so that it has the same energy as the δ-pulse of the disturbance D and therefore optimally counteracts it. The shape of the disturbance in the form of the δ-pulse obviously cannot be achieved by a single or two resonators, but the energy (or the effect) is preserved.

The final stable value for ρ_1 is the main difference between the open-loop case and the closed-loop case. While in the open-loop case, the weight ρ_1 grows endlessly due to the lack of feedback (see the simulations in the companion article), in the closed-loop condition, the weight ρ_1 converges to a specific value at the moment when $x_0 = 0$ has been achieved. As a consequence, the experimentally observed behavior of the algorithms leads to a function H_v with similar properties as that obtained from the second-order Taylor approximation.

For all practical purposes, N needs to be found in trying to resolve the trade-off between the actually needed precision for $t^{(\infty)}_{\max} \to T$ and hardware and software engineering constraints (costs). The robot experiment in the companion article in this issue demonstrates that in a real-world application, few resonators ($N = 10$) suffice to obtain the desired obstacle avoidance behavior after learning.

Now we have to consider more complex transfer functions for P_1. Up to this point, we have set $P_1 = 1$, which means that the disturbance basically reaches the input X_1 unfiltered, which is in general not the case. Due to specific sensor properties and due to properties in the environment, the disturbance reaches the input X_1 in a filtered form. All of these changes can be subsumed from the organism's point of view by the function P_1 (and the

[2] The relation $t^{(2)}_{\max} \approx T$ could be confirmed because we performed the same Taylor approximation with $N = 2$ (leading to a fourth-order Taylor approximation):

$$-e^{-sT} = \frac{-1}{1 - sT + \frac{1}{2}s^2T^2 - \frac{1}{6}s^3T^3 + \frac{1}{24}s^4T^4} \tag{3.16}$$

$$\rho_1 H_1(s) + \rho_2 H_2(s) = \frac{\rho_1}{(s + p_1)(s + p_1^*)} \frac{\rho_2}{(s + p_2)(s + p_2^*)} \tag{3.17}$$

The resulting set of equations (from comparing the coefficients) has been solved numerically, and we received a solution that leads to $t^{(4)}_{\max} = 0.978T$. This suggests that $t^{(\infty)}_{\max} = T$ is correct in the limit of $N \to \infty$.

same applies to P_0). We recall that we have used a Taylor approximation of equation 3.9 and matched it with the sum of resonators to obtain the coefficients. This, however, allows concluding that any transfer function P_1 of the shape

$$P_1 = \frac{(s+z_0)(s+z_0^*)\cdots(s+z_n)(s+z_n^*)}{(s+p_0)(s+p_0^*)\cdots(s+p_m)(s+p_m^*)} \tag{3.18}$$

can still (together with the delay term $-e^{-sT}$) be approximated by a sum of resonators, because this sum continues to take the shape of a broken rationale function similar to that in equation 3.18.[3] Such a shape of P_1, however, covers all generic combinations of high- and low-pass characteristics. Hence, it represents a standard passive transfer function. In addition, we can normally assume that the environment does not actively interfere with signal transmission in such a system and can therefore—with great likelihood—be represented by equation 3.18. Thus, we can argue that an appropriate approximation of the complete equation 3.8 will be found in almost all natural situations. The robot application shown in the companion article in this issue supports this notion experimentally.

3.2.3 Convergence Properties. The previous section has shown that it is possible to construct approximative solutions of equation 3.8 using resonators so that $X_0(s) \to 0$. Here, we address the problem of whether the learning rule will actually converge onto such a solution.

Conventional techniques used to derive a learning rule by calculating the partial derivatives of the weights and finding the minimum fail in our case because ISO learning is linear. As a consequence, the derivatives are constant, and a minimum cannot be found. An approach that leads to success, however, is to apply perturbation theory instead.

Let us first treat the system very generally without making a priori assumptions as to the characteristics of the H_k. In so doing, we can employ perturbation analysis with the nice aspect that we will not make any assumption as to the size of the perturbation. Thus, proof of stability against such a perturbation is equivalent to a proof of convergence. For real resonators, this will be a little bit different, though, as we will see.

Let us assume that we have found a set of weights $\rho_k,\ k > 0$ that solves equation 3.8, and we know that the development of the weights follows equation 2.5. Now we perturb the system, substituting ρ_j in equation 2.5 with $\rho_j + \delta\rho_j = \tilde{\rho}_j$. In order to ensure stability, we must prove that the

[3] Note that we are even able to approximate zero crossings of equation 3.18 since we have a sum of resonator responses. If we calculate the overall transfer function of a sum of resonators ($H_1 + H_2 + \cdots$), we automatically also get zero crossings, which can be used to identify them with the zero crossings in equation 3.18. Thus, the approximation, including the phase terms, is correct.

perturbation is counteracted by the weight change; thus, we must solve equation 3.8 hoping to find

$$\Delta\rho_j \sim -\delta\rho_j. \tag{3.19}$$

Note that this would guarantee convergence because we know that μ is small, which prevents oscillations.

After some calculations (see the appendix), we arrive at

$$\Delta\rho_j = \frac{\mu}{2\pi}\sum_{k=1}^{N}\delta\rho_k\int_{-\infty}^{\infty} -i\omega\frac{|X_1|^2H_k^-}{1-\rho_0P_0^-H_0^-}H_j^+\,d\omega, \tag{3.20}$$

where we use the superscripts $^+$ and $^-$ for the function arguments $+i\omega$ and $-i\omega$. This result is still general in the sense that we are not necessarily dealing with resonator functions, so at the moment we are still free to make some reasonable assumptions about the set of H_k. Let us thus assume orthogonality given by

$$0 = \int_{-\infty}^{\infty} -i\omega\frac{|X_1|^2H_j^+H_k^-}{1-\rho_0P_0^-H_0^-}\,d\omega \quad \text{for} \quad k \neq j, \tag{3.21}$$

and we get

$$\Delta\rho_j = \frac{\mu}{2\pi}\delta\rho_j\int_{-\infty}^{\infty}|X_1^+|^2|H_j^+|^2\frac{-i\omega}{1-\rho_0P_0^-H_0^-}\,d\omega. \tag{3.22}$$

In order to prove that the integral in equation 3.22 will be negative (ensuring convergence), the inner (reflex) loop, which is determined by $\rho_0H_0P_0$, needs to be considered. Note that this loop must at least be stable; otherwise, the system would not be functional to begin with. A theoretical result from the literature (Sollecito & Reque, 1981) supports the notion that the integral in question is negative as long as the stability of $\rho_0H_0P_0$ is guaranteed. Let us try to spell this rather general argument out more concretely. (In the appendix, we rigorously prove convergence for the important case of unity feedback.)

By the use of Plancherel's theorem (Stewart, 1960), we transfer the integral in equation 3.22 into the time domain and get

$$\Delta\rho_j = \mu\delta\rho_j\int_0^{\infty} a_{x*h}(t)f'(t)\,dt, \tag{3.23}$$

where we call $a_{x*h}(t)$ the autocorrelation function of $x_1(t)*h_j(t)$, which is the inverse transform of $|X_1^+H_j^+|^2$ (the asterisk denotes a convolution). We note

that the remaining term in equation 3.22, $\frac{-i\omega}{1-\rho_0 P_0^- H_0^-}$, contains the derivative operator $-i\omega$ in the numerator. Thus, $f'(t)$ in equation 3.23 is the temporal derivative of the impulse response of the inverse transform of $\frac{1}{1-\rho_0 P_0^- H_0^-}$.

Now we must ask what the most general condition for the reflex loop is (defined by $\rho_0 H_0 P_0$) to be stable. For a concrete stability analysis, knowledge of P_0 would be required, which normally cannot be obtained. We can, however, in general assume that P_0, being an environmental transfer function, should again behave passively and follow equation 3.18. Furthermore, we know that the environment delays the transmission from the motor output to the sensor input. Thus, P_0 must be dominated by a low-pass characteristic, without which it would be unstable.[4] As a consequence, we can in general state that the fraction $\frac{1}{1-\rho_0 P_0 H_0}$ is dominated by the characteristic of a (nonstandard) high pass. It follows that its derivative has a very high negative value for $t = 0$ (ideally $= -\infty$) and vanishes soon after. The autocorrelation a is positive around $t = 0$. Thus, the integral in question will remain negative as long as the duration of the disturbance D remains short. As an important special case, we find that this especially holds if we assume delta-pulse disturbance at $t = 0$, corresponding to $x_1(t) = \delta(t)$.

Thus, for an orthogonal set of H_k, we have found that ISO learning will converge if P_0 is dominated by a low-pass characteristic and if we use a disturbance D with a short duration.

Finally, we have to prove that equation 3.22 is zero in the equilibrium state case where the feedback loop is no longer needed. Thus, we have $0 = X_0 = \rho_0 H_0 P_0$, and the denominator becomes one. We get

$$\Delta\rho_j = \frac{\mu}{2\pi}\delta\rho_j \int_{-\infty}^{\infty} -i\omega |X_1^+|^2 |H_j^+|^2 \, d\omega. \tag{3.24}$$

This integral is antisymmetrical, and thus zero as required. In the companion article in this issue, we discussed that the synaptic weights in the open-loop condition stabilize as soon as we explicitly set $X_0 = 0$, arriving at the same equation (compare equation 2.21 in the companion article). In the closed-loop condition used here, this is obtained in a natural way, as the result of implicitly eliminating the reflex during the learning process.

3.2.4 Matching the Theoretical Convergence Properties to the Practical Approach. In this section, we now use real resonator functions for H_k and H_j (see equations 2.2 and 2.3). Normally, the transfer functions of the resonators are not orthogonal, but we will show by numerical integration that the system still behaves properly.

[4] Note that the unity feedback condition, treated in the appendix, represents the simplest possible stable reflex loop. Its low-pass characteristic is reduced to being a mere delay in this case.

Here, we use the unity feedback condition defined in the appendix in order to be able to work with a concrete example, which is initially stable, and we get for equation 3.20

$$\Delta\rho_j = \frac{\mu}{2\pi} \sum_{k=1}^{N} \delta\rho_k \int_{-\infty}^{\infty} \frac{-i\omega H_j^+ H_k^-}{1 - \rho_0 e^{i\omega\tau}}\, d\omega, \tag{3.25}$$

where we have set $D = 1$, which represents a δ-function as a disturbance.

Figure 4a shows the numerically obtained results for $\Delta\rho_j$ as defined in equation 3.25 in the case of a perturbation.

We note that the resonators are not orthogonal since we have nonzero contributions for nearly all $j \neq k$. The system, however, still compensates for perturbations and thus converges, for the following reason. First, consider Figure 4a, which represents the case of how the system reacts to a perturbation, and look at the diagonal. We find that the values of the integral (see equation 3.25) are negative on the diagonal. This means that any perturbation at ρ_j will lead to a counterforce onto itself and, consequently, to a compensation of the perturbation.

However, the nondiagonal elements $k \neq j$ are nonzero, so we have to discuss them and argue why this does not interfere with the compensation process. Thus, the question of stability must be rephrased into a question of how a perturbation at one given weight ρ_k will influence the other weight(s). Most important, we observe that the value of the integral (see Figure 4a) is substantially smaller than one everywhere else. This, however, shows that any perturbation at index k will reenter the system at index j only in a strongly damped way. This process leads to a decay of any perturbation through further iterations. This strictly holds for two paired indices j and k. However, even for the complete sum in equation 3.25, which describes all cross-interference terms, we can argue that perturbations will be eliminated. This is true as long as the sum remains below one, which is realistic, given the small and sign-alternating values of the integral surface.

From this, we realize that strict orthogonality as defined in equation 3.21 is not necessary to ensure convergence. This constraint can be relaxed to the constraint that the absolute value of the sum in equation 3.25 (or equation 3.20, respectively) should remain below one. Thus, for all practical purposes, we can concentrate on the behavior of the diagonal elements even without having to employ an orthogonal set of H.

Figure 4b shows the equilibrium case with $\rho_0 X_0 P_0 = 0$. We note that in this case, the integral is zero for $k = j$, which is in accordance with theory. Since we are in the equilibrium, we do not expect any weight changes.

4 Discussion

In this article, we have focused on finding a mathematically motivated interpretation of the results from feedback loop (self-referential)–based ISO

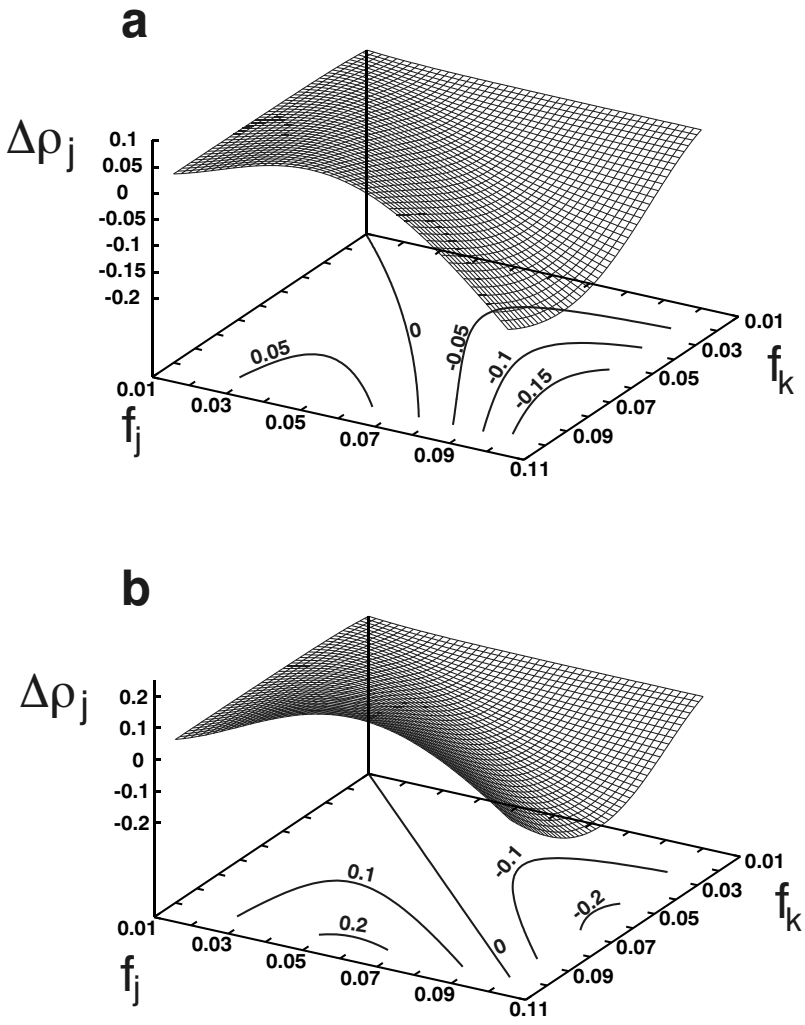

Figure 4: Numerical integration of equation 3.25. The disturbance D and the reflex loop delay τ were both set to one. The frequencies of the resonators H_k and H_j were varied from 0.01 to 0.1 in steps of 0.001. The value of Q was set to $Q = 0.9$ for both resonators. The weight of the reflex loop was $\rho_0 = -0.9$.

learning. We were able to show that such a system approximates the inverse controller of the reflex. The theoretical results are at some critical points rather nicely linked to the experimental findings shown in the companion article that support the validity of the theory.

Most of the technical aspects, like necessary assumptions (e.g., the orthogonality problem), have already been discussed in the previous sections. Therefore, we restrict the discussion here to more general problems.

The inverse controller problem belongs to the most famous problems in engineering. Typical solutions are always based on an intrinsic model (a so-called forward model) of the to-be-controlled system. In contrast, our approach is model free because it is based on learning. Furthermore, engineered forward models have the central disadvantage that they will fail if something unexpected happens. Thus, control engineers use their forward controllers only in conjunction with the feedback-loop controller on which the forward model was originally based. The same strategy is pursued in a natural way in our setup. The double-loop structure of Figure 3 clearly shows that the reflex will again take over if the outer loop fails. In contrast to engineered systems, however, this will lead to a continuation of the learning process such that the system will continue to improve throughout its lifetime.

A frequently addressed problem in biology is the control of voluntary limb movements, for example, in the arm movement models developed by Haruno, Wolpert, and Kawato (2001) and others. These authors also employ forward models (inverse controllers) to address problems of limb control in a mixed-model approach (Wolpert & Ghahramani, 2000). The idea that forward models are involved in motor control has been explored, for example, by Grüsser (1986), who tried to explain the stability of the visual percept during voluntary eye movements by means of an internal representation of the motor command, which is called "efferent copy" or "corollary discharge" (von Uexküll, 1926). By now, clear evidence exists for such a general mechanism, the details of how it is implemented, however, are still under debate. A discussion about this is beyond the scope of this article, but our theoretical results suggest that sequence order learning can provide a method by which forward models can be generically designed (i.e., learned). It is conceivable that this observation is not restricted to our specific algorithm but also holds for other temporal sequence learning algorithms like TD learning.

The models by Wolpert and Ghahramani (2000), Haruno et al. (2001), and others have in common that they use supervised learning schemes, usually TD learning, to learn the forward model. As we stated in section 1, the goal of our two letters in this issue is to provide an unsupervised temporal sequence learning algorithm for autonomous behavior. An organism that is autonomous cannot rely on external rewards. Internal rewards are possible, but if we treat autonomy seriously, then even an internal reward originates in the last instance from a sensor input. Verschure and Voegtlin (1998) used the same paradigm: a (Hebb-like) unsupervised learning algorithm together

with a reflex as a reference. A novel aspect of our work is that we have taken the environment explicitly into account and introduced it as a nonevaluative structure. Thereby, the organism reacts only to the relevant parts of the environment's structure, and in the theoretical treatment, only aspects of the environment have to be taken into account that either establish the reflex loop or can be used to supersede it. In that sense, the organism acquires not arbitrary sensorial information but useful information—useful in the sense of helping it to supersede the reflex (von Glasersfeld, 1996).

The definition of autonomy has been based on the aspect of (un-)predictability of behavior (Ford & Hayes, 1995). It is interesting to consider how our system fits into this framework. The acquisition of additional useful sensorial information enables the organism to predict unwanted changes in the environment. Thus, for the organism, predicting the reflex leads to more behavioral security as compared to the situations when it had to entirely rely on the reflex reaction. However, the gain of security for the organism will lead to an increase of uncertainty observed in the environment. What this means can be understood by reconsidering the robot experiment shown in the companion article. As long as the robot has only its reflex behavior, it is absolutely predictable for an observer. From the moment learning eliminates the reflex, the robot's behavior becomes more and more unpredictable. Although the robot solves its goal (obstacle avoidance), it cannot be predicted how the robot actually achieves this. It is specifically this duality of certainty versus uncertainty (depending on the point of view of actor versus observer) that is central to the definitions of autonomy. Such principles are also identified as the basis for the emergence of social behavior (Luhmann, 1995).

Our two articles in this issue are meant to provide an alternative framework for temporal sequence learning, which by its linear structure provides better access to analytical treatment than do existing techniques. In addition, we believe that the ISO learning algorithm could have significant commercial potential, because it can, in a model-free way, solve various inverse controller problems, which should be of relevance for different applied control situations.

Two questions immediately arise that should be addressed by future research: Which modifications have to be done to implement an "attraction" case opposed to the shown "avoidance" case? and Is there a way to implement ISO learning using spike trains and biophysically modeled neurons? These issues extend the scope of this article and are topics for further investigation.

Appendix

In this appendix, we give the detailed equations for the convergence proof and derive the proof in a rigorous way for the so-called unity feedback condition.

A.1 Detailed Equations. We continue after equation 3.19. We need to define U and V. U is easy:

$$U_j = X_j H_j = \begin{cases} X_0 H_0 & \text{for} \quad j = 0 \\ X_1 H_j & \text{for} \quad j > 0 \end{cases}. \tag{A.1}$$

V is more complicated. From the definition, we have

$$V = \rho_0 X_0 H_0 + X_1 \sum_{k=1}^{N} \rho_k H_k, \tag{A.2}$$

and from above, we know (see equation 3.1) that

$$X_0 = P_0[V + De^{-sT}]. \tag{A.3}$$

Thus, we get for V,

$$V = \rho_0 P_0 [V + De^{-sT}] H_0 + X_1 \sum_{k=1}^{N} \rho_k H_k \tag{A.4}$$

$$= \rho_0 P_0 H_0 V + \rho_0 P_0 H_0 De^{-sT} + X_1 \sum_{k=1}^{N} \rho_k H_k, \tag{A.5}$$

resulting in

$$V = \frac{\rho_0 P_0 H_0 De^{-sT} + X_1 \sum_{k=1}^{N} \rho_k H_k}{1 - \rho_0 P_0 H_0}. \tag{A.6}$$

Substituting $\rho_j \to \rho_j + \delta\rho_j$, we get

$$\tilde{V} = \frac{\rho_0 P_0 H_0 De^{-sT} + X_1 \sum_{k=1}^{N} \rho_k H_k + X_1 \sum_{k=1}^{N} \delta\rho_k H_k}{1 - \rho_0 P_0 H_0} \tag{A.7}$$

$$= V + \frac{X_1 \sum_{k=1}^{N} \delta\rho_k H_k}{1 - \rho_0 P_0 H_0}. \tag{A.8}$$

Calculating the weight change is done using equation 2.5:

$$\Delta\tilde{\rho}_j = \frac{\mu}{2\pi} \int_{-\infty}^{\infty} -i\omega \left[V^- + \frac{X_1^- \sum_{k=1}^{N} \delta\rho_k H_k^-}{1 - \rho_0 P_0^- H_0^-} \right] X_1^+ H_j^+ \, d\omega, \tag{A.9}$$

where we have introduced the abbreviations $^+$ and $^-$ for the function arguments $+i\omega$ and $-i\omega$.

We realize that the first part of this integral describes the equilibrium state condition and can be dropped; thus,

$$\Delta\rho_j = \frac{\mu}{2\pi}\sum_{k=1}^{N}\delta\rho_k\int_{-\infty}^{\infty} -i\omega\frac{|X_1|^2H_k^-}{1-\rho_0P_0^-H_0^-}H_j^+\,d\omega, \tag{A.10}$$

where for X_1 we have made use of the fact that for transfer functions in general, we can write $Y^+Y^- = |Y|^2$, and we have reached equation 3.20 of the main text.

A.2 Introducing the Unity Feedback Loop Restriction. The basic (critical) property of a reflex loop is its delay characteristic. This property underlies the conceptual necessity for temporal sequence learning and is essential for any relevant mathematical treatment. The specific characteristics of some of the transfer function, on the other hand, are secondary and can therefore be simplified.

Thus, we will use the so-called unity feedback loop assumption to capture this property. It is defined by

$$\rho_0 \in\ ]-1, 0[\tag{A.11}$$

$$H_0 := 1 \tag{A.12}$$

$$P_0 := e^{-s\tau}. \tag{A.13}$$

The reflex loop is thus entirely determined by its gain ρ_0 and by the delay τ (not to be confused with T), which is the delay between the motor output V and the sensor input X_0. The range of ρ_0 defined by equation A.11 results from the demand that the reflex should be a negative feedback loop and that it must be stable.

In addition, we assume that the transfer function P_1 of the predictive pathway represents unfiltered throughput given by

$$P_1 := 1. \tag{A.14}$$

Finally, we assume that the disturbance D should be short, with a duration that is shorter than τ (otherwise, the loop would become unstable) and that it can be developed into a product series of conjugate zeroes and poles (e.g., low-/band- or high-pass characteristics). Thereby, D also takes on the property of a typical transfer function.

A.3 Convergence for Unity Feedback. In the main text, we arrived at equation 3.22,

$$\Delta\rho_j = \frac{\mu}{2\pi}\delta\rho_j\int_{-\infty}^{\infty}|X_1^+|^2|H_j^+|^2\frac{-i\omega}{1-\rho_0P_0^-H_0^-}\,d\omega \tag{A.15}$$

and we have to prove that this integral is negative. This can be directly shown for unity feedback. Thus, equation A.15 turns into

$$\Delta\rho_j = \frac{\mu}{2\pi}\delta\rho_j \int_{-\infty}^{\infty} \underbrace{|DH_j|^2}_{A(i\omega)} \underbrace{\frac{-i\omega}{1-\rho_0 e^{i\omega\tau}}}_{-i\omega F(-i\omega)}\, d\omega. \tag{A.16}$$

As in the main text, we apply Plancherel's theorem to equation A.16 in order to transfer the integral back into the time domain and prove that it is negative. We get

$$\Delta\rho_j = \mu\delta\rho_j \int_0^{\infty} a(t) f'(t)\, dt. \tag{A.17}$$

The function $F(s)$ of equation A.16 is given by the transformation pair,

$$F(s) = \frac{1}{1-\rho_0 e^{-s\tau}} \leftrightarrow f(t) = (-1)^n \delta(t - n\tau), \quad n = 0, 1, 2, \ldots, \tag{A.18}$$

where f represents an alternating δ-function at $t = 0, \tau, 2\tau, \ldots$, which starts with a positive delta pulse (Doetsch, 1961). Thus, together with $-i\omega$, the complete term $(-i\omega\frac{1}{1-\rho_0 e^{i\omega\tau}})$ represents $f'(t)$, hence the temporal derivative of f.

The other term $A(s)$ of equation A.16 is given by

$$A(s) = |DH_j|^2 \leftrightarrow a(t) = \Phi[d(t) * h_j(t)], \tag{A.19}$$

where the asterisk denotes a convolution and Φ the autocorrelation function.

As a consequence of the above findings, we have to discuss the integral in equation A.17 specified by the time functions in equations A.18 and A.19. The integral should be negative to ensure stability. We know that D is short-lived with a duration shorter than τ, without which the loop system would be unstable to begin with. Thus, we can restrict the discussion of the integral to $t = 0$. We know that the autocorrelation function a has a positive maximum at $t = 0$ and that the derivative f' of a delta pulse at zero approaches $-\infty$ for $t \to 0; t > 0$. As a consequence, the integral is negative, as required for convergence.

Acknowledgments

We are grateful to Leslie Smith and to the members of the CCCN seminar for their helpful comments during various stages of this work. This study was supported by grants from SHEFC RDG INCITE and by the European funding ECOVISION. UK patent pending.

References

Blinchikoff, H. J. (1976). *Filtering in the time and frequency domain*. New York: Wiley.

D'Azzo, J. J. (1988). *Linear control system analysis and design*. New York: McGraw-Hill.

Doetsch, G. (1961). *Guide to the applications of the Laplace and z-transforms*. New York: Van Nostrand Reinhold.

Ford, K. M., & Hayes, P. J. (Eds.). (1995). *Android epistemology*. Cambridge, MA: MIT Press.

Grüsser, O. (1986). Interaction of efferent and afferent signals in visual perception: A history of ideas and experimental paradigms. *Acta Psychol., 63*, 3–21.

Haruno, M., Wolpert, D. M., & Kawato, M. (2001). Mosaic model for sensorimotor learning and control. *Neural Comp., 13*, 2201–2220.

Luhmann, N. (1995). *Social systems*. Stanford, CA: Stanford University Press.

McGillem, C. D., & Cooper, G. R. (1984). *Continuous and discrete signal and system analysis*. New York: CBS Publishing.

Nise, N. S. (1992). *Control systems engineering*. New York: Cummings.

Palm, W. J. (2000). *Modeling, analysis and control of dynamic systems*. New York: Wiley.

Sollecito, W., & Reque, S. (1981). Stability. In J. Fitzgerald (Ed.), *Fundamentals of system analysis*. New York: Wiley.

Stewart, J. L. (1960). *Fundamentals of signal theory*. New York: McGraw-Hill.

Terrien, C. (1992). *Discrete random signals and statistical signal processing*. Upper Saddle River, NJ: Prentice Hall.

Verschure, P., & Voegtlin, T. (1998). A bottom-up approach towards the acquisition, retention, and expression of sequential representations: Distributed adaptive control III. *Neural Networks, 11*, 1531–1549.

von Glasersfeld, E. (1996). Learning and adaptation in constructivism. In L. Smith (Ed.), *Critical readings on Piaget* (pp. 22–27). London: Routledge.

von Uexküll, B. J. J. (1926). *Theoretical biology*. London: Kegan Paul, Trubner.

Wolpert, D. M., & Ghahramani, Z. (2000). Computational principles of movement neuroscience. *Nature Neuroscience, 3* (Suppl.), 1212–1217.

Received April 12, 2002; accepted November 1, 2002.