Information driven self-organization: The dynamical system approach to autonomous robot behavior

Nihat Ay[1,3], Ralf Der[1], Mikhail Prokopenko[2]
[1]Max Planck Institute for Mathematics, Leipzig, Germany
[2]CSIRO, Sydney, Australia
[3]Santa Fe Institute, Santa Fe, USA

June 30, 2010

Abstract

In recent years, information theory has come into the focus of researchers interested in the sensorimotor dynamics of both robots and living beings. One root for these approaches is the idea that living beings are information processing systems and that the optimization of these processes should be an evolutionary advantage. Apart from these more principal questions, there is much interest recently in the question how a robot can be equipped with an internal drive for innovation or curiosity that may serve as a drive for an open ended, self-determined development of the robot. The success of these approaches depends essentially on the choice of a convenient measure for the information. This paper studies in some detail the use of the predictive information (PI), also called excess entropy or effective measure complexity, of the sensorimotor process. The PI of a process quantifies the total information of past experience that can be used for predicting future events. However, the application of information theoretic measures in robotics mostly is restricted to the case of a finite, discrete state-action space. This paper aims at applying the PI in the dynamical systems approach to robot control. We study linear systems as a first step and derive exact results for the PI together with explicit learning rules for the parameters of the controller. Interestingly, these learning rules are of Hebbian nature and local in the sense that the synaptic update is given by the product of activities available directly at the pertinent synaptic ports. The general findings are exemplified by a number of case studies. In particular, in a two-dimensional system, designed at mimicking embodied systems with latent oscillatory locomotion patterns, it is shown that maximizing the PI means to recognize and amplify the latent modes of the robotic system. This and many other examples show that the learning rules derived from the maximum PI principle are a versatile tool for the self-organization of behavior in complex robotic systems.

1 Introduction

In recent years, information theory has come into the focus of researchers interested in the self-organization of robot behavior. One root for these approaches is the idea that living beings are information processing systems and that the optimization of these processes might be an evolutionary advantage. Apart from these more speculative ideas there is much interest recently in the question how a general principle can be found for equipping a robot with an internal drive for innovation or curiosity. This leads away from the pure task dependent paradigms of robotics towards a robot that is driven solely by the desire to get more and more information about itself and the environment. Eventually, a strategy for an open ended, self-determined development of the robot might emerge.

The development of these ideas has soon made clear that one has to use a convenient measure for the information. Maximizing Shannon's information is not directly feasible since it favors processes, like noise, of maximum complexity. Optimal in that sense would be a robot that behaves as random as possible. Alternatively, in [19] a set of univariate and multivariate statistical measures are introduced in order to quantify the information structure in sensory and motor channels. Generic information theoretic criteria may vary in their emphasis: for example, one may focus on maximization of empowerment (the perceived amount of influence or control that the agent has over the world) [18] and [17]; minimization of heterogeneity across states of multiple agents, measured with either the variance of Shannon entropy of rule-space [27], or Boltzmann entropy of swarm-bots' states [2]; maximization of spatiotemporal coordination within a modular robot, measured via the excess entropy computed over a multivariate time series of modules' states [26], etc. What is common to these examples of information-driven self-organization is the characterization of sensorimotor (or perception-action) loop in information theoretic terms. For instance, the empowerment measures the amount of Shannon information that the agent can "inject into" its sensor (i.e., received signal) through the environment, affecting future actions and future perceptions. Technically, for a pre-defined agent's behavior, empowerment is defined as the capacity of the agent's actuation channel: the maximum mutual information for the channel over all possible distributions of the transmitted signal (i.e., actions) [18] and [17]. On the other hand, the maximization of excess entropy during a time interval, used in [26], allows to change the controllers' logic (that is, change the agent's behavior) within a modular robot in such a way that its actuators become coordinated. In this example, the adaptation of controllers occurs evolutionarily with the fitness function rewarding the regularity and richness of the actuators' multivariate series. The same adaptation can also be achieved during the agent's lifetime — in other words, the time interval over which the excess entropy is computed may be interpreted either as the full lifetime of the individual (leading to an evolutionary representation), or as a finite period within such lifetime (leading to an online learning representation).

This paper studies in some detail the use of the predictive information of

the sensorimotor process. It differs from the above mentioned studies in the following way: on the one hand, the approach does not aim at maximizing the sensorimotor channel capacity for a fixed pre-defined behavior (that was the case with empowerment), but rather attempts to produce learning rules for the agent optimizing its behavior; on the other hand, the approach aims to produce such learning rules explicitly, rather than leaving the optimization of the behavior to evolutionary or other implicit operators (that was the case with excess entropy in [26]). Moreover, an essential objective is to make the approach independent of any discretization of the state and/or the action space so that it can be immediately useful in the dynamical systems approach to robotics.

The predictive information of a process quantifies the total information of past experience that can be used for predicting future events. Technically, it is defined as the mutual information between the future and the past, see [4]. It has been argued that predictive information, also termed excess entropy [6] and effective measure complexity [15], is the most natural complexity measure for time series.

The behaviors emerging from maximizing the PI are qualified by the fact that predictive information is high if – by its behavior – the robot manages to produce a stream of sensor values with high information content (in the Shannon sense) under the constraint, however, that the consequences of the actions of the robot remain still predictable. A robot maximizing PI therefore is expected to show a high variety of behavior without becoming chaotic or purely random. In this working regime, somewhere between order and chaos, the robot may be expected to explore its possibilities of actions in a most effective way.

This is why the PI may serve as a drive for a self-determined development of a robot. This complements approaches that equip the robot with a motivation system producing internal reward signal for reinforcement learning a pre-specified task. Pioneering work has been done by Schmidhuber using the prediction error as a reward signal in order to make the robot curious for new experiences, [28]. The approach has been further developed in a number of papers, see e.g. [31], [29]. Related ideas have been put forward in the so called play ground experiment by Kaplan and Oudeyer [16], [22] by using the learning progress as a reward signal. Steels[30] proposes the Autotelic Principle, i.e. the balance of skill and challenge of behavioral components as the motivation for open ended development whereas Barto [3] uses the prediction error of skill models to build hierarchical skill collections. Predictive information could be used alternatively as a reward signal in reinforcement learning. This would be of special interest also in connection with recent developments in reinforcement learning in continuous state action spaces, cf. [32], because the PI is not restricted to discrete spaces at all. However in the present paper we will not follow this line but instead derive task-free learning rules directly from the gradient ascent on the PI.

The application of information theoretic measures in robotics mostly is restricted to the case of a finite state-action space with discrete actions and sensor values. The past two decades in robotics have seen the emergence of a new trend of control in robotics which is rooted more deeply in the dynamical systems ap-

proach to robotics using continuous sensor and action variables. This approach is very appealing since it yields more natural movements of the robots and allows to exploit embodiment effects in a most effective way. For instance many successful realizations of the so called morphological computation are realized by using recurrent neural networks as controller of the dynamical system body, brain, and environment, see [24] and [25] for an excellent survey.

The information theoretic approach in the dynamical systems representation has still to be worked out in detail and it is the main motivation of this paper to present some first results in this direction. We start by answering the following question: Given a robot in its environment, how can we find the dynamical system describing its dynamics together with an explicit learning rule optimizing behavior so that the predictive information of the sensor process is maximized. This approach has to work from scratch, i.e. without any knowledge about the robot, so that everything has to be inferred from the sensor values alone. The question can not be answered in full generality so that we restrict ourselves in this paper to the case of linear systems. This is a restriction of generality but has the advantage that we get analytical results, general statements, and last but not least explicit learning rules. However, the results are useful also in the nonlinear case as will be demonstrated in a later paper.

In a recent paper a general learning rule [33] has been derived from the predictive information by using the natural gradient technique in a finite state-action space. This paper complements that approach for the case of continuous spaces and controllers realized by parameterized functions like neural networks. The information theoretical approach can also be considered as an alternative to the principle of homeokinesis [12],[8], a systematic approach to the self-organization of behavior that has been applied successfully to a large number of complex robotic systems, cf. [11], [10], [14], [13]. Moreover, this principle has also been extended to form a basis for a guided self-organization of behavior [21], [20]. We hope to benefit substantially from this parallel in future work.

The paper is organized as follows: Section 2.2 is devoted to a rather general sensorimotor dynamics. We start in Section 2 by formulating the general dynamic system, and give in 2.1 the corresponding translation into the probabilistic picture. This representation is necessary for the evaluation of the predictive information introduced and studied in some detail in Section 3. Section 4 exemplifies these general results for the case of a stochastic oscillator system, mimicking in particular embodied systems with latent oscillatory locomotion patterns. Considering the special case of linear systems, general learning rules are derived in Sec. 6. Although restricted to linear systems, these results are useful since they show a way how the sampling problem can be overcome in practical applications. The extension to the nonlinear case will be given in a later paper.

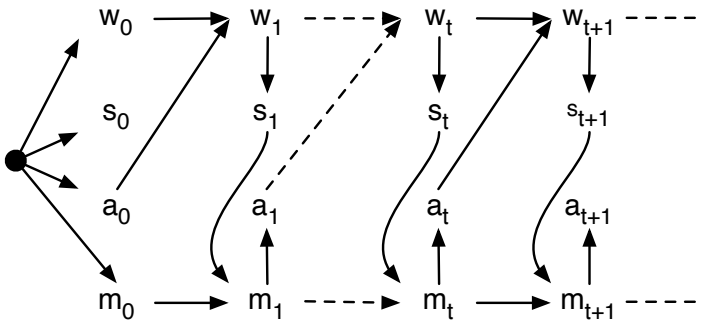

Figure 1: Schematic representation of the sensorimotor loop. The state of the world at time t is w_t. The world is observed by the sensor values s_t which are then memorized by the internal state m_t. Actions are functions of the internal state m.

2 The sensorimotor loop

The sensorimotor loop introduced by the Figure 1 can be formalized either in terms of the kernels which define the processes or by specifying the corresponding time discrete stochastic dynamical system.

2.1 Probabilistic formulation

We are now going to give a brief sketch of the representation of the sensorimotor loop, cf. Fig. 1 by formulating the relevant transition kernels. We do not claim that the diagram represents any situation but, to our experience, most of the situations encountered in robotics or biology are being covered. Quite generally, a kernel $p(y|x)$, where $x \in R^n$ and $y \in R^m$, is a function $f : R^n \times R^m \to R^1$ assigning each combination of vectors x, y a probability, so that $0 \leq f \leq 1$ and $\int f(s, w)\, ds = 1$. The pertinent kernels specifying the sensorimotor loop are given as

- The dynamics of the world (which is usually not known explicitly) is assumed to be described by the kernel $p(w_{t+1}|w_t)$ defining the probability that the world state in the next time step is w_{t+1} given the world is now in state w_t. If the process is non-stationary, the kernel will carry an extra time index t so that the kernel is written as $p(w_{t+1}|w_t; t)$. In the case of non-stationary states, all kernels introduced here and in the following are explicitly time dependent but we will omit this in the notation in the following.

- The world state $w_t \in R^{nw}$ observed at time t is mapped by the kernel $p(s_t|w_t)$ to the sensor state $s_t \in R^{ns}$.

- However the information in the sensor process S is usually not complete which is the problem of hidden variables. One way out of this is to introduce the memory m_t which includes earlier sensor values so that, hopefully, the missing information is reintroduced by the memory. The time evolution of the latter is given by the kernel $p(m_t|m_{t-1}, s_t)$.

- Actions are given in terms of the history by the kernel $p(a_t|m_t)$ which defines the policy of the agent.

- Given the action and the history of the sensor process we may define the kernel $p(s_{t+1}|m_t, a_t)$ which determines the sensorimotor dynamics. This kernel is essentially what we need in order to determine the predictive information studied in this paper.

One of the aims of this paper is to use the information theoretic measures like predictive information in the dynamical systems theory of the sensorimotor loop. We therefore give in the following the relation between these two approaches.

2.2 Dynamical systems formulation

The translation of the sensorimotor dynamics as given by the above kernels into the dynamical systems language can be done in different ways, we use the following. Let us consider first a generic kernel $p(y|x)$ defining a vector of random variables Y with statistical properties depending on the state x. Let us call this a state dependent random vector and denote it by $Y(x)$. For instance, the average $\langle Y(x)\rangle$ of $Y(x)$ is

$$\langle Y(x)\rangle = \int y\ p(y|x)\ dy$$

The trasition to the dynamical systems formulation is made by using the method of functional causal models [23] where any kernel describing a transition (like those in Fig. 1) from a state $x \in R^n$ into a state $y \in R^m$ can be defined by a function $f : R^n \times R^{n_u} \to R^m$ so that

$$y = f(x, u)$$

where $u \in R^{n_u}$ is a vector of perturbances.

In particular, any realization of the sensor process as represented by the kernel $p(s_{t+1}|s_t, a_t)$ can be written as

$$s_{t+1} = \phi(s_t, a_t, u_{t+1}) \tag{1}$$

where u_t is a vector of perturbances representing both the influence of purely random processes and the influences of hidden variables. The latter, although actually caused by a deterministic dynamics, may be corrupted by random influences on the starting conditions at least so that we may qualify the perturbances quite generally as noise of a general nature.

Special cases are easily obtained for the case that perturbances are weak. By Taylor expanding with respect to u we obtain in leading order

$$s_{t+1} = \phi(s_t, a_t) + \rho_{t+1} \tag{2}$$

where $\phi(s_t, a_t) = \phi(s_t, a_t, 0)$, $\rho_{t+1} = Qu_{t+1}$, and the matrix Q is given by

$$Q(s_t, a_t) = \left.\frac{\partial}{\partial u}\phi(s_t, a_t, u)\right|_{u=0}$$

Q depending both on the state and the action in a deterministic way. In many cases of practical interest, the state dependence of the noise can be neglected and this is what we are dealing with in the present paper as a first step towards a general approach to predicitive information maximization in stochastic dynamical systems.

Using the policy given by the kernel $p(a_t|s_t)$ we may introduce the kernel

$$p(s_{t+1}|s_t) = \int p(s_{t+1}|s_t, a_t)\, p(a_t|s_t)\, da_t$$

corresponding to a structural model

$$s_{t+1} = \psi(s_t, \mu_{t+1})$$

which, in the case of low noise μ, is equivalent to the functional model

$$s_{t+1} = \psi(s_t) + \xi_{t+1} \tag{3}$$

(where now $\psi(s) = \psi(s, 0)$, $Q = \partial\psi/\partial\mu$, and $\xi = Q\mu$) in the way described above. In linear systems, this decomposition is exact even for arbitrary noise so that the results derived below for linear systems are exact as well.

3 Predictive information

The predictive information (PI) is the mutual information between the future and the past, relative to some instant of time t, of the time series S

$$I(S_{\text{past}}; S_{\text{future}}) = \left\langle \ln \frac{p(S_{\text{past}}, S_{\text{future}})}{p(S_{\text{past}})\, p(S_{\text{future}})} \right\rangle = H(S_{\text{future}}) - H(S_{\text{future}}|S_{\text{past}}) \tag{4}$$

where the averaging is over the joint probability $p(S_{\text{past}}, S_{\text{future}})$. Note that in the case of continuous variables, the individual entropy components $H(S_{\text{future}})$, $H(S_{\text{future}}|S_{\text{past}})$ may well be negative whereas the PI is always positive and may exist even in cases where the individual entropies diverge. This is a very favorable property deriving from the explicit scale invariance of the PI, see below. Eq.4 simplifies considerably if S is a Markov process, see [1]. In this case

the PI is given by the mutual information (MI) between two successive time steps, i.e. instead of Eq. (4) we consider

$$I\left(S_{t+1};S_t\right)=\left\langle \ln\frac{p\left(S_{t+1},S_t\right)}{p\left(S_{t+1}\right)p\left(S_t\right)}\right\rangle =H\left(S_{t+1}\right)-H\left(S_{t+1}|S_t\right) \tag{5}$$

which simplifies the sampling process considerably. In experiments with a coupled chain of robots done earlier [9] it was observed that the MI of just a single sensor, one of the wheel counters of an individual robot, already yields essential information on the behavior of the robot chain. It proved to be maximal if the individual robots managed to cooperate so that the chain as a whole could navigate effectively. This is remarkable in that a one-dimensional sensor process can already give essential information on the behavior of a very complex physical object under real world conditions. These results give us some encouragement to study the role of PI and other information measures for relatively simple sensor processes as is done in the present paper.

3.1 Example systems

In order to get some feeling on the behavior of the PI in simple models of the sensorimotor loop we use Eq. (2) again

$$s_{t+1}=\phi\left(s_t,a_t\right)+\rho_{t+1} \tag{6}$$

ρ being state dependent in general and the controller is assumed to be given by a kernel $p\left(a_t|s_t\right)$ corresponding specifically to a structural model

$$a_t=K\left(s_t\right)+\kappa_t \tag{7}$$

where $a_t\in R^m$ is the vector of motor values the controller outputs at time t and the actuator noise $\kappa\in R^m$ is a realization of the stochastic vector $\varkappa$. In order to get some analytical results we assume both processes to be linear so that

$$K\left(s\right)=Cs \text{ and } \phi\left(s,a\right)=Ts+Va \tag{8}$$

the matrix T representing the contribution to the sensor process due to some dynamics of the world alone and V is the sensorial response to the output of the controller. This might seem an oversimplified system but it should be remembered that many of the control systems studied in engineering are of this kind.

Under the assumptions made, the sensor process is now

$$s_{t+1}=Rs_t+\xi_{t+1} \tag{9}$$

where

$$R=T+VC \tag{10}$$

and $\xi=V\kappa+\rho$ is the effective combination of controller and world noise.

In general, as explained above, the random parts of these expressions will be state dependent. However we will use only additive noise terms (no state dependence) in the following in order to get analytical results as a first step towards more general situations.

3.2 Predictive information in linear control systems

Eq. (9) can be considered as an AR(1) process. Autoregressive models play an important role in many branches of science and engineering so that there is a large body of available results. In particular with Gaussian noise, measures like predictive information, can be obtained in closed forms.

3.2.1 The process

Let us consider therefore the case of the process defined by Eq. (9) with

$$\xi \sim N(0, D)$$

In the stationary state S is a Gaussian process as well

$$S \sim N(0, \Sigma_s)$$

with mean zero and covariance matrix Σ_s (no state dependence for simplicity) which can be obtained simply from the solution of Eq. (9)

$$s_t = R^t s_0 + \sum_{k=0}^{t-1} R^k \xi_{t+1-k} \tag{11}$$

as

$$\Sigma_s = \sum_{k=0}^{\infty} R^k D {R^k}^T \tag{12}$$

where (E_p means averaging over the density distribution p of the noise)

$$D = E_p\ \xi\xi^T$$

and both the white noise property and $E_p\ \xi = 0$ was used. Alternatively, Σ is shown easily to be the solution of the discrete Lyapunov equation

$$\Sigma_s = R\Sigma_s R^T + D \tag{13}$$

3.2.2 Explicit expression

The conditional distribution for s_{t+1} with s_t given is obtained directly as

$$p(s_{t+1}|s_t) = N(Rs_t, D) \tag{14}$$

because of the additivity of the noise. Noting that the entropy does not depend on the mean, we find

$$H(S_{t+1}|S_t) = \frac{1}{2} \ln |D| + \frac{n}{2} \ln 2\pi e^2 \tag{15}$$

where we use the notation

$$\det M = |M|$$

here and in the following wherever this does not lead to ambiguities.

The entropy of a Gaussian random vector $S \sim N(\mu, \Sigma_s)$ is well known, see [5]

$$H(S) = \frac{1}{2} \ln |\Sigma_s| + \frac{n}{2} \ln 2\pi e^2$$

so that using Eq. (15)

$$I(S_{t+1}; S_t) = \frac{1}{2} \ln |\Sigma_s| - \frac{1}{2} \ln |D| \tag{16}$$

which is the entropy of the state minus that of the noise. This additive decomposition of the PI is a direct consequence of the linear dynamics plus additive noise.

3.2.3 Properties

The PI displays a number of interesting properties. Well known but especially noteworthy for robotics is the invariance of the PI against coordinate transformations so that the PI of a process S_t is the same as that of a process QS_t for any regular matrix Q. This follows immediately from $I(S_{t+1}; S_t) = H(S) - H(S_{t+1}|S_t)$ and the fact that entropies obey

$$H(QS) = H(S) + \ln |\det Q| \tag{17}$$

for any regular matrix Q. This also shows that the PI is independent of the scaling of the variables which is very convenient for robotics applications.

More specifically, in the system considered one of the striking properties of the PI is its preferentially dynamic nature. This is seen best by considering the special case of isotropic noise, i.e.

$$D = \sigma^2 \mathbf{I} \tag{18}$$

where $\mathbf{I}$ is the unit matrix and σ^2 measures the overall strength of the noise. In this case we get the variance of the state directly from Eq. (12) as

$$\Sigma_s = \sum_{k=0}^{\infty} R^k D R^{kT} = \sigma^2 \sum_{k=0}^{\infty} R^k R^{kT} = \frac{\sigma^2}{1 - RR^T} \tag{19}$$

so that the PI is

$$I(S_{t+1}; S_t) = -\frac{1}{2} \ln \det \left(\mathbf{I} - RR^T \right) \tag{20}$$

which depends on the dynamical operator R only. We will give further below an example where the dependence of the PI on the anisotropy of the noise is made explicit.

3.2.4 Further expressions

The PI can be rewritten in a number of different forms. In particular, using $|A|/|B| = \left|AB^{-1}\right|$ and the Lyapunov equation we write

$$\frac{|D|}{|\Sigma_s|} = \left|\left(\Sigma_s - R\Sigma_s R^T\right)(\Sigma_s)^{-1}\right| = \left|\left(\mathbf{I} - R\Sigma_s R^T(\Sigma_s)^{-1}\right)\right|$$

Introducing

$$W = \Sigma_s^{-\frac{1}{2}} R \Sigma_s^{\frac{1}{2}}$$

we obtain

$$\frac{|D|}{|\Sigma_s|} = \left|\left(\mathbf{I} - WW^T\right)\right|$$

where $\left|I + AMA^{-1}\right| = |I + M|$ and $\Sigma_s = \Sigma_s^T$ was used. Hence we obtain the predictive information also as

$$I(S_{t+1}; S_t) = -\frac{1}{2}\ln\left|\mathbf{I} - WW^T\right| \tag{21}$$

The predictive information is now expressed in terms of the so called pre-whitened dynamical operator W which is a similarity transform of the bare dynamical operator R by means of the covariance matrix Σ of the stochastic process [7]. This generalizes the expression Eq. (20) to the case of anisotropic noise in a straightforward way.

3.2.5 Approximations

From the computational point of view, the evaluation of the determinant may be annoying in high dimensional systems. This can be avoided if the eigenvalues of W are sufficiently small. Using

$$|\mathbf{I} - \varepsilon M| = 1 + \varepsilon Tr(M) + O\left(\varepsilon^2\right)$$

we obtain approximately

$$\left|\mathbf{I} - WW^T\right| \approx 1 - Tr\left(WW^T\right)$$

so that by means of the cyclic invariance of the trace

$$I(S_{t+1}; S_t) \approx \frac{1}{2} Tr\left(WW^T\right) = \frac{1}{2} Tr\left(RR^T\right)$$

Obviously the noise does not play any role even in the anisotropic case if the dynamics is strongly damped.

3.3 Summary

The present section has given explicit expressions for the PI of linear dynamical systems with additive noise. These results are partly known already but we worked them out again from the perspective of the sensorimotor loop. Remarkable features of the PI are seen in its invariance against scale transformations of the state variables which is very convenient for robotics applications. The more interesting point is the preferentially dynamic nature of the PI. With additive noise, the PI splits additively into a dynamical and a pure noise part, the latter being irrelevant for the maximization task. The dynamical part is essentially the entropy of the state variables which is seen to decouple completely from the noise if the latter is isotropic. The general case is covered in the same way by pre-whitening the dynamical operator. These results are encouraging for the use of the PI in the dynamical systems approach to robotics.

4 Example stochastic oscillator

Let us now consider a two-dimensional system in order to study pertinent properties of the PI, in particular the interplay between the controller and the dynamics of the world. By way of example we consider a system with a damped oscillation perturbed by noise, i.e. we consider Eq. (9)

$$s_{t+1} = Rs_t + \xi_{t+1}$$

with specific expressions of the dynamical operator R. Moreover we put the covariance matrix of the noise as

$$D = E\ \xi\xi^T = \sigma^2 \begin{pmatrix} 1-m & 0 \\ 0 & 1+m \end{pmatrix}$$

This is sufficiently general since D can always be brought into a diagonal form by using an orthogonal transformation of the state vector s. The specific way of writing the diagonal elements has proven to simplify the expressions to be derived in the following.

4.1 Controlling a random world

Let us start with the case that the deterministic part of the dynamics is determined by the controller alone, i.e. the intrinsic world dynamics is pure noise so that $T = 0$ in (10). The controller is

$$a = Cs$$

with controller matrix C

$$C = cU(\phi)$$

where U is a rotation matrix

$$U(\phi) = \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix}$$

so that the dynamical operator is

$$R = cU(\phi) \tag{22}$$

The system executes with $0 < c < 1$ a damped harmonic oscillation since the state vector is rotated in each time step by the angle ϕ and compressed by the factor c.

We find Σ_s from the solution of the discrete Lyapunov equation (Maple) as

$$\Sigma_s = \begin{pmatrix} \frac{2c^2\left(\left(c^2-1\right)m+2\right)\cos^2\phi+\left(1-c^4\right)m-\left(1+c^2\right)^2}{(c^4-1)(1+c^2)+4c^2(1-c^2)\cos^2\phi} & 2\left(\cos\phi\sin\phi\right)m\frac{c^2}{(c^2+1)^2-4c^2\cos^2\phi} \\ 2\left(\cos\phi\sin\phi\right)m\frac{c^2}{(c^2+1)^2-4c^2\cos^2\phi} & -\frac{\left(c^2+1\right)^2-m\left(c^4-1\right)+2c^2\left(\left(c^2-1\right)m-2\right)\cos^2\phi}{(c^4+c^6-c^2-1+(4c^2(1-c^2)\cos^2\phi))} \end{pmatrix}$$

The determinant can be written as

$$|\Sigma_s| = \left(1 - \frac{m^2}{1+4c_\phi^2}\right)\frac{1}{\left(c^2-1\right)^2}$$

where

$$c_\phi^2 = \frac{c^2\sin^2\phi}{\left(c^2-1\right)^2}$$

$|\Sigma_s|$ is seen to have minima at $\phi = 0, \pi, \ldots$ and maxima at $\pi/2, 3\pi/2, \ldots$independently of the values of m and c. Using $|D| = 1 - m^2$ we write the PI as

$$I(S_{t+1}; S_t) = \frac{1}{2}\ln\frac{|\Sigma_s|}{1-m^2} = \frac{1}{2}\ln\frac{1}{\left(c^2-1\right)^2} + \frac{1}{2}\ln\frac{1-\ m^{*2}}{1-m^2}$$

or with the specific setting for R, cf. Eq. (22)

$$I(S_{t+1}; S_t) = \frac{1}{2}\ln\frac{1}{1-RR^T} + \frac{1}{2}\ln\frac{|D^*|}{|D|} \tag{23}$$

$$= I_{iso}(S_{t+1}; S_t) + \frac{1}{2}\ln\frac{|D^*|}{|D|} \tag{24}$$

where $I_{iso}(S_{t+1}; S_t)$ is the PI with isotropic noise and D^* is the noise matrix with m replaced by

$$m^* = \frac{m}{\sqrt{1+4c_\phi^2}}$$

which can be considered as a kind of re-scaled noise asymmetry reflecting the interaction with the dynamics. Eq. (23) presents the PI as a term which depends only on the dynamics of the system, like in the isotropic noise case plus a term due to the interplay of the dynamics with the anisotropy of the noise.

The PI has the same extrema as $|\Sigma_s|$ so that it is maximal if the deterministic dynamics is a period 4 damped oscillation. A gradient ascent on the PI will drive the system towards this frequency. Moreover, the PI is a monotonously increasing function of c^2 so that the gradient ascent will not only increase the frequency up to the period 4 oscillation but also increase the noise amplification so that the noise is increasingly amplified (Σ_s increases) by the dynamics of the system.

4.2 Resonance – A case for embodiment

A purely random and therefore void world is not the most interesting or typical case. Instead the world will have a dynamics of its own and the question is how the PI depends on the interplay between the controller and the world. Let us consider the very simple case of a world dynamics given by an oscillatory system, i.e. put $T = wU(\omega)$ and $C = cU(\phi)$ so that

$$R = cVU(\phi) + wU(\omega)$$

(we will use $V = \mathbf{I}$ for the sake of simplicity for a while) with c and w chosen such that the eigenvalues of RR^T are less than 1.

4.2.1 Isotropic noise

The case of isotropic noise can be considered explicitly since

$$\mathbf{I} - RR^T = (1 - \mu)\,\mathbf{I} \tag{25}$$

where

$$\mu = c^2 + w^2 + 2cw\cos(\phi - \omega)$$

so that, cf. Eq. (20)

$$I(S_{t+1}; S_t) = -\frac{1}{2}\ln(1 - \mu) \tag{26}$$

where both $0 < c < 1$ and $0 < w < 1$. Assuming these values are such that $1 - (w + c)^2 > 0$, I exists and is maximal if $\phi = \omega$ i.e. if the controller is in resonance with the dynamics of the world.

This can be connected to the idea of embodiment. Our system (without controller) is driven by the noise into damped oscillations. Now assume that we switch on the controller and let the latter adapt its frequency by gradient ascending the PI. Then, the controller will bring gradually its frequency in resonance with the intrinsic oscillation of the world without doing any frequency analysis or the like. This is valid as long as we keep the strength factor c fixed. The more general case is considered below.

It is important to remember, that these processes are of a completely dynamic nature so that there is no sampling of any probability kernels involved, provided the world matrix T is known. The latter can be learned on-line what, so to say, is the price to pay for not having to sample. The learning however is done much more easily than the sampling, since learning here is a simple supervised task.

4.2.2 Anisotropic noise

The preceding scenario requires that the mode is already active so that it is represented explicitly in the world matrix T. In many cases of practical interest, modes get excited only if the controller already insinuates a near-resonance stimulation. However, if the noise is isotropic there is no dependence of the

PI on the frequency of the controller, see above, so that the frequency space would have to be searched by hand. But, as shown above, even the slightest inhomogeneity of the noise leads to a frequency sweep through the whole frequency space from frequency zero to the period 4 oscillation. If, during this sweep, a mode in the world is excited and the world model is adapted to cover this emerging feature sufficiently quickly, the resonance mechanism described above will start dominating the adaptation so that the controller is driven towards the intrinsic mode and amplifies the latter to maximum amplitude.

This mechanism shows not only how the PI maximization can lead to an active search of the behavior space but also how, in this procedure, latent modes of the world can be brought out. This is an even stronger point for the use of PI maximization in embodied AI.

Of course, different to the isotropic noise case, this scenario is not completely free of any sampling requirements. Nevertheless, since the sweeping effect sets in as soon as there is any anisotropy of the noise at all, it would be sufficient to have a very coarse sampling and start the adaptation process right from the outset. The sampling can continue during the information maximization so that, on the fly, the kernels may be improved.

4.3 The magic circle world

The resonance phenomenon is also present if, different to the preceding case, the matrices for the controller and the world are not of the same structure. Let us consider in particular the magic circle oscillator realized by the matrix

$$M(\omega) = \begin{pmatrix} 1 & \omega \\ -\omega & 1-\omega^2 \end{pmatrix}$$

which, if used instead of $U(\omega)$, produces an (anharmonic) oscillatory system. Its frequency is defined by the eigenvalues $\lambda_{1/2} = 1 - \frac{1}{2}\omega^2 \pm i\omega\sqrt{1-\frac{1}{4}\omega^2}$. The eigenvalues are identical to those of the orthogonal matrix $U(\omega)$ in lowest order of ω. Together with a controller defined by $U(\phi)$ we get the dynamical operator

$$R = cU(\phi) + wM(\omega)$$

which can again be analyzed in simple ways if the noise is assumed isotropic, cf. Eq. (20). We note without going into the details that with small values of ω we get a very precise resonance behavior. This is not surprising since with these rotation angles the matrices $U(\omega)$ and $M(\omega)$ are nearly identical. However one gets also for quite large values of ω a very good agreement in the frequencies. The comparison is made via the imaginary parts of the eigenvalues of $M(\omega)$ and the normalized dynamical operator $P = R/\sqrt{\det R}$. Using $c = .2$ and $w = .1$ we find for instance with $\omega = 0.5$ the maximum of the PI at $\phi = 0.52$. The corresponding eigenvalues of $M(\omega)$ are $0.875 \pm 0.48i$ and those of P are $0.87 \pm 0.49i$. This means that the frequency of the controlled system is practically identical to that of the intrinsic mode alone, i.e. information maximization tunes

the controller to nearly complete resonance. The corresponding values for $\omega = 1$ are $\phi = 1.15$ with eigenvalues $M(\omega)$ at $0.5 \pm 0.87i$ and those of P at $0.43 \pm 0.90i$ which means a deviation in frequencies of less than 5 percent in the region $-\pi/3 < \omega < \pi/3$.

We may conclude from this that, at least in the case considered, the PI is maximal if the controller is (nearly) in resonance with an intrinsic mode of the world even in the case that the world and controller are structurally different.

4.4 Summary

This section has considered the application of the PI concept to a specific two-dimensional systems mimicking a sensorimotor loop with a controller that can excite oscillatory motions. The world part (essentially the body of the robot) of the sensorimotor loop consisted of (i) a pure noise, (ii) an oscillatory part of the same dynamical structure as the controller, and (iii) an oscillator of different structure (magic circle world). We have demonstrated that in the pure noise case the anisotropy of the noise produces a frequency sweeping effect, driving the system towards a period 4 oscillation which is the dynamics with the highest predictive information. An interesting effect is observed if the world is not just pure noise but is capable of an oscillatory dynamics of its own. In that case, the PI is maximal if the controller is (nearly) in resonance with this intrinsic mode of the world even in the case that the world and controller are structurally different. This is encouraging, since maximizing the PI means (at least in this simple example) to recognize and amplify the latent modes of the robotic system. This is essentially what we need for the self-organization of behavior by the maximization of the PI in the sensorimotor loop.

5 PI over several time steps

We may also consider the more general case of a larger step width τ, the derivation for the PI being given in the Appendix. The PI as a function of the lead time τ is known to be a Lyapunov function for the process so that it decays with increasing τ. More interestingly for our purpose, the landscape is seen to become more and more complex with increasing τ. Let us consider the special case of the purely random world and consider the landscape of the PI for a special case, see Fig.2 which depicts for the case of $c = .8$ and $m = 0.05$ the dependence on the rotation angle ϕ.The picture shows that instead of the period 4 oscillation observed in the single time step case, the oscillations with maximum PI are now of much lower frequency, the frequency decreasing systematically with increasing τ. Moreover, there are also local maxima at high frequency oscillations but with a much lower value of the PI for $\tau > 1$.

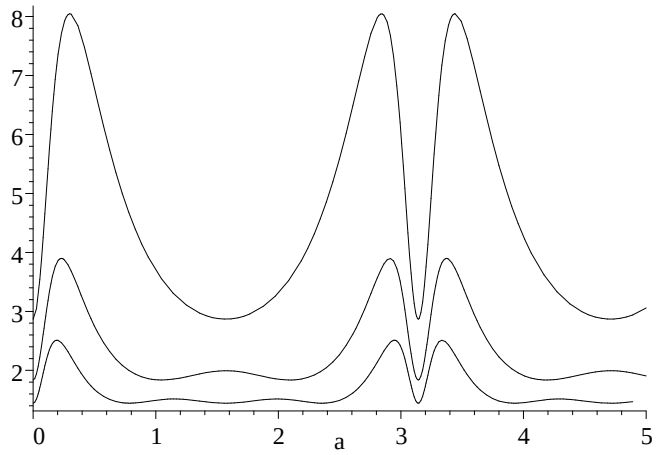

Figure 2: The predictive information over τ time steps for the stochastic oscillator model as a function of the rotation angle ϕ for $\tau = 2$ (upper), $\tau = 3$ (middle), and $\tau = 4$ (lower curve). Instead of the maximum at $\pi/2$ observed for $\tau = 1$ in the region $0 < \phi < \pi$, there are two global maxima and one or two local maxima for the case of $\tau = 3$ or $\tau = 4$, respectively. The frequency of the oscillations in the global maxima decreases with increasing τ and depends also in a very intricate way on the damping constant α and the asymmetry of the noise measured by m.

6 Learning rules. The self-referential robotic system

The PI is given in terms of the sensor values the robot produces in the course of time. There is no domain specific knowledge invoked into this function. We obtain a self referential robotic system when using the PI as the objective function for the adaptation of the parameters of the controller. In particular we may consider the gradient ascent on the MI

$$\Delta p = \varepsilon \frac{\partial I\left(S_t; S_{t-1}\right)}{\partial p}$$

where p is any parameter of the controller of the robot. The properties of the self-referential robotic system depends also on the choice of the learning rate ε which actually has to be chosen small enough so that the time scales are well separated. This, however, is a serious problem if the PI has to be sampled.

6.1 Explicit learning rules for the maximization of predictive information

The results obtained above may show a way how to avoid or at least to smooth the sampling problem. We have seen, that the PI in specific cases is a purely dynamical quantity so that actually there is no sampling necessary at all. This is the case if either the noise in the linear dynamical system is isotropic or, more generally, if the dynamical operator describes a highly damped system. The explicit learning rules in both cases can be obtained directly from the gradient ascent on $I\left(S_{t+1}|S_t\right)$ using Eq. (20) as

$$\Delta C = \varepsilon V^T \frac{1}{\mathbf{I} - RR^T} R \tag{27}$$

The rule can be used as long as all eigenvalues of RR^T are less than one. This is an immediate consequence of the fact that the dynamics is diverging if ever one of those eigenvalues exceeds one. When using the rule in practice a damping term should be added. There are several choices possible, in the present case it seems appropriate to keep the variances of the state variables at finite values. This amounts to using

$$K = I\left(S_{t+1}; S_t\right) - \lambda Tr\left(\Sigma_s\right) \tag{28}$$

as the new objective function to be maximized. The trace over Σ_s can be evaluated from Eq. (12) by elementary means. In the case of isotropic noise we get the rule, see the Appendix,

$$\Delta C = \varepsilon V^T \frac{1}{\left(\mathbf{I} - RR^T\right)^2} \left(\gamma \mathbf{I} - RR^T\right) R \tag{29}$$

where $\gamma = \left(1 - \frac{\lambda}{\varepsilon}\right)$ and $0 < \lambda < \varepsilon$. The learning process is stationary if

$$RR^T = \gamma \mathbf{I}$$

with PI

$$I\left(S_{t+1}|S_t\right) = \frac{1}{2} \ln \frac{\varepsilon}{\lambda}$$

meaning that all eigenvalues of R are equal to $\sqrt{\gamma}$ in absolute value and RR^T is diagonal. When expressing R in terms of a singular value decomposition, $R = UMV^T$ where U and V are rotation matrices and M is diagonal, we easily conclude that R converges towards an orthogonal matrix multiplied by $\sqrt{\gamma}$. The convergence process depends on the relation of λ and ε and on the initial condition for R.

6.2 The Hebbian nature of the learning rule

The learning rule can be rewritten in many different forms. Useful for the practical applications is the avoidance of the matrix inversion or the sampling necessary for the evaluation of Σ_s.

6.2.1 Stochastic gradient ascent rule

This can be done in the following way. Using Eq. (19) and $D = \sigma^2 \mathbf{I}$ we have $\left(\mathbf{I} - RR^T\right) = \Sigma_s D^{-1}$so that the learning rule is

$$\Delta C \ = \varepsilon V^T \Sigma_s R$$

where σ^2 was absorbed into ε. The variance Σ_s of the state variables can, in the sense of a stochastic gradient procedure, be invoked into the algorithm if the learning rate ε is chosen small enough, so that the update of C in the time step t is

$$\Delta C = \varepsilon V^T s_t s_t^T R \tag{30}$$

This may be helpful in practical applications since it does not involve any matrix inversion, the update is fully determined by the current value of the state vector s_t. Adding of penalty terms needs some care since for instance $\left(\mathbf{I} - RR^T\right)^{-2}$ is not given by $E_{p(s)} \ \left(ss^T\right)^2$. On the other hand, penalty terms like the one given in Eq. (32) do not cause any problems.

6.2.2 Hebbian learning

The above rule can be still further modified in order to make a connection to neural network learning paradigms. Let us introduce the new vectors $\widetilde{a}_t = V^T s_t$ and $\widetilde{s}_t = R^T s_t$. In terms of these states we write the learning rule as (omitting time indices)

$$\Delta C_{ij} = \varepsilon \widetilde{a}_i \widetilde{s}_j \tag{31}$$

We may consider C_{ij} as the synaptic strength of a linear neuron. Interpreting $\widetilde{a}_i$ as signal at the output of the neuron i and $\widetilde{s}_j$ as an input into the synapse, the learning rule is now clearly Hebbian, since the update for the synapse is given by the product of activities being available directly at the corresponding ports. The analogy can be made even closer if we make contact with the error back propagation rules which are central in the learning theory of layered feed forward neural networks. For this purpose we consider the combination of the controller matrix C and the world matrix V as a two-layer neural network of linear neurons. We consider the dynamics

$$s_{t+1} = Va_t + \xi_{t+1}$$

and interpret Va as the output of the top layer of the network so that (sum over repeated indices)

$$(Va)_i = g\left(V_{ij}a_j\right)$$

with a linear output function $g(z) = z$. The controller on its hand can be written as

$$a_j = g\left(C_{jk}s_k\right)$$

so that the deterministic part Rs_t of the full dynamics $s_{t+1} = Rs_t + \xi_{t+1}$ can be written as a two-layer neural network

$$(Rs)_i = g\left(V_{ij}a_j\right) = g\left(V_{ij}g\left(C_{jk}s_k\right)\right)$$

The error back-propagation rule allows to propagate a signal at the output of the network back to the lower layers and finally to the input of the network. Propagating according to that rule s_t from the output of the network (top layer) back to the output of the controller (bottom layer) yields

$$(\widetilde{a}_t)_i = \left(V^T s_t\right)_i$$

which, in the learning step, features as the output signal at the synapse. Propagating this activity further down to the input of the network yields

$$(\widetilde{s}_t)_j = \left(C^T \widetilde{a}_t\right)_j = \left(R^T s_t\right)$$

which is the input signal into the synapse C_{ij} in the learning step, see Eq. (31).

6.3 The resonance effect

The result already reveals a specific feature of the predictive information maximization paradigm. Since the PI is invariant with respect to an arbitrary orthogonal transformation of the state space, the learning will converge towards some orthogonal matrix depending on the initial conditions and the values of the parameters ε and λ.

This can be made more explicit in our specific resonance example. Using Eq. (25) we find in this case, using $RR^T = \mu\mathbf{I}$ and

$$\mu = c^2 + w^2 + 2cw\cos\left(\phi - \omega\right)$$

that the condition of stationarity reads now

$$c^2 + w^2 + 2cw\cos(\phi - \omega) = 1 - \frac{\lambda}{\varepsilon}$$

with infinitely many solutions for c and ϕ realizing the same value of the PI.

This is a little disappointing since there is no pronounced resonance behavior any more (the resonance was obtained with c fixed). A more detailed analysis shows that there is an approach of ϕ towards ω. However since c increases very rapidly the penalty term comes soon into play so that the convergence of ϕ is stopped before it reaches the resonance frequency. This means that the resonance phenomenon does not disappear altogether but that it is not complete.

The resonance can be reestablished by using appropriate penalty terms. For instance, if using the typical weight decay term, i.e.

$$K = I(S_{t+1}; S_t) - \lambda Tr\left(C^T C\right) \tag{32}$$

instead of Eq. (28) we find in our special case that $Tr\left(C^T C\right) = c^2$ so that

$$K = -\frac{1}{2}\ln(1-\mu) - \lambda c^2$$

We note without going into detail here that this will indeed drive ϕ towards ω. Moreover, with this damping term, the resonance effect is observed also in the case of an arbitrary parameterization of the controller matrix. However this does work only if λ is chosen sufficiently large. If not so, the logarithmic singularity at $\mu = 0$ dominates the learning dynamics so that the penalty term is overrun and the learning dynamics diverges.

This shows that there is a serious problem in using penalty terms in order to keep the linear system in bounds. Of course, one could solve the problem for instance by using a normalization of the controller matrix after each learning step. The main problem, however, is a conceptual one. The point here is that defining an appropriate penalty term is a domain specific task. This might work for many specific applications but does not meet the challenge of finding a general approach to the self-organization of behavior. Fortunately, these problems disappear more or less if the sensorimotor loop is confined by nonlinearities like the saturation properties of the involved neurons or a nonlinear sensor characteristics. This has been proven already in the case of one-dimensional systems, see [1], and will be discussed in the general case in a later paper. In particular, it will be shown that the nonlinearities support the resonance phenomenon so that nonlinearities are essential for the emergence of embodiment effects.

6.4 Summary

The aim of this section was the derivation of learning rules for the maximization of the PI. We restricted ourselves to the case of linear systems and derived an explicit update rule for the matrix C of the controller. The essential point is that, in the case of isotropic noise at least, the rule is of a completely dynamical

nature so that no sampling is necessary at all. Instead, the response matrix V and the world matrix T have to be learned, but this is a supervised learning task which is easy to achieve. Moreover, the learning rule has been transformed into a purely local form, see Eq. (30) so that no matrix inversions are necessary. This is of much interest from the practical point of view.

We have discussed several penalty terms (which are necessary in the case of linear systems) and demonstrated that the inherent contingency of behaviors emerging from PI maximization gives the opportunity to influence the course of the learning process by appropriate penalty terms. In particular, the resonance effect could be reestablished for more general parameterizations of the controller. This supports our point of view that the PI maximization makes the robot "feel" latent behavioral modes, in the special case the existence of an oscillatory regime, like a locomotion pattern. Maximizing the PI via the learning mechanism leads to the recognition and amplification of this mode. This may also be understood as a kind of self-motivated exploration of bodily affordances of embodied robots.

7 Conclusions

Can a robot develop its skills completely on its own, driven by the sole objective to gain more and more information about its body and its interaction with the world? This aim raises immediately further questions like (i) what is the relevant information for the robot and (ii) how can one find a convenient learning rule that realizes the gradient ascent on this information measure. We have studied the predictive information contained in the stream of sensor values as a tentative answer to the first question and, based on that, could give exact answers to the second question for simple cases. We had to limit the investigation to the case of linear controllers and sensorial responses in order to get exact analytical results. Nevertheless, already in such a linear world there are several effects which demonstrate the value of the information maximization principle. In particular, we could show that the (anisotropic) noise makes the system to explore its behavior space in a systematic manner, in the present case the PI maximization made the controller of a stochastic oscillator system to sweep through the space of available frequencies. More importantly, if the world the controller is interacting with is hosting a latent oscillation, the controller will learn by PI maximization to go into resonance with this intrinsic mode of the world. This is encouraging, since maximizing the PI means (at least in this simple example) to recognize and amplify the latent modes of the robotic system. In a sense, by PI maximization the robot is able to detect its bodily affordances and this may be interpreted as a tentative mathematical foundation of morphological computation.

In the special case of isotropic noise the PI maximization principle lead to simple learning rules which can be given a purely local formulation. In fact, it needs only standard backpropagation together with a Hebbian learning step. There is no need for sampling or doing any non-local operations. Of course, this

is a result of the linearity of the system and the isotropy of the noise. However, our preliminary results with non-linear systems indicate that a similar structure can be achieved also in the general case, at least in approximations. This may help to bridge the gap between standard neural network realizations (with supervised learning) which are so successful in robotics and the information theoretic methods which so far are based on discretization and burdened with high sampling efforts and involved learning rules. Hopefully our results will help to pave the way for the application of information theoretic methods as a reliable tool for the self-determined development of the behavior of complex autonomous robots. Moreover, the approach may lead to concrete realizations of concepts relevant for truly autonomous robots like internal motivation and artificial curiosity.

Acknowledgement 1 *Part of this work was completed during a stay of Nihat Ay and Ralf Der at the CSIRO in Sydney, Australia. Hospitality and financial support are gratefully acknowledged. Mikhail Prokopenko thanks the Max Planck Institute of Mathematics in the Sciences in Leipzig, Germany, for support and hospitality at the Institute.*

8 Appendix

We derive here some results used in the text.

8.1 Predictive information over several time steps

The conditional distribution for $s_{t+\tau}$ given s_t is obtained by means of Eq. (11) as

$$p\left(s_{t+\tau}|s_t\right) = N\left(R^\tau s_t, \Sigma_{s_{t+\tau|t}}\right) \tag{33}$$

where ($\tau > 0$)

$$\Sigma_{s_{t+\tau|t}} = \sum_{k=0}^{\tau-1} R^k D {R^k}^T \tag{34}$$

Writing

$$\begin{aligned}\Sigma_{s_{t+\tau|t}} = \sum_{k=0}^{\tau-1} R^k D {R^k}^T &= \sum_{k=0}^{\infty} R^k D {R^k}^T - \sum_{k=\tau}^{\infty} R^k D {R^k}^T \\ &= \sum_{k=0}^{\infty} R^k D {R^k}^T - R^\tau \sum_{k=0}^{\infty} R^k D {R^k}^T {R^\tau}^T\end{aligned}$$

we obtain the discrete Lyapunov equation for τ steps as

$$\Sigma_{s_{t+\tau|t}} = \Sigma_s - R^\tau \Sigma_s {R^\tau}^T$$

Noting that the entropy does not depend on the mean, we find

$$H\left(S_{t+\tau}|S_t\right) = \frac{1}{2} \ln \left|\Sigma_{s_{t+\tau|t}}\right| + \frac{n}{2} \ln 2\pi e^2 \tag{35}$$

so that

$$I\left(S_{t+\tau};S_t\right)=\frac{1}{2}\ln\frac{\left|\Sigma_s\right|}{\left|\Sigma_{s_{t+\tau|t}}\right|} \tag{36}$$

In analogy to the derivation of Eq. (21) we rewrite this as

$$I\left(S_{t+\tau};S_t\right)=-\frac{1}{2}\ln\left|\mathbf{I}-W_\tau W_\tau^T\right| \tag{37}$$

with the pre-whitened operator

$$W_\tau=\Sigma_s^{-\frac{1}{2}}R^\tau\Sigma_s^{\frac{1}{2}} \tag{38}$$

8.2 Derivation of the learning rule

We use

$$\frac{\partial}{\partial Q_{ij}}\ln\det Q=\frac{1}{\det Q}\frac{\partial}{\partial Q_{ij}}\det Q=\left(Q^{-1}\right)_{ji}$$

or more compactly

$$\frac{\partial}{\partial Q}\ln\det Q=\frac{1}{Q^T}$$

Putting $Q=\mathbf{I}-RR^T$ we find

$$\frac{\partial}{\partial Q}\ln\det\left(\mathbf{I}-RR^T\right)=\frac{1}{\mathbf{I}-RR^T} \tag{39}$$

and by means of (sum over repeated indices)

$$\frac{\partial}{\partial R_{ij}}\ln\det\left(\mathbf{I}-RR^T\right)=\frac{\partial Q_{kl}}{\partial R_{ij}}Q_{kl}^{-1}=-\frac{\partial R_{km}R_{lm}}{\partial R_{ij}}Q_{kl}^{-1}=Q_{il}^{-1}R_{lj}+Q_{ki}^{-1}R_{kj}$$

so that by the symmetry of Q we obtain

$$\frac{1}{2}\frac{\partial}{\partial R}\ln\det\left(\mathbf{I}-RR^T\right)=\frac{1}{\mathbf{I}-RR^T}R$$

Using $R=VC+T$ we find eventually

$$\frac{1}{2}\frac{\partial}{\partial C}\ln\det\left(\mathbf{I}-RR^T\right)=V^T\frac{1}{\mathbf{I}-RR^T}R$$

The derivation of the penalty term is obtained in the following way. Using the cyclic invariance of the trace we get

$$Tr\left(\Sigma_s\right)=Tr\left(\sum_{k=0}^{\infty}R^kDR^{k^T}\right)=Tr\left(\frac{1}{\mathbf{I}-R^TR}D\right)$$

which is valid for any kind of noise. With isotropic noise we get

$$Tr\left(\Sigma_s\right)=\sigma^2Tr\left(\frac{1}{\mathbf{I}-R^TR}\right)=\sigma^2Tr\left(\frac{1}{\mathbf{I}-RR^T}\right)$$

and

$$\frac{\partial}{\partial C} Tr\left(\frac{1}{\mathbf{I} - RR^T}\right) = -2Tr\left(\frac{1}{\mathbf{I} - RR^T} V \frac{\partial C}{\partial C} R^T \frac{1}{\mathbf{I} - RR^T}\right)$$
$$= -2V^T \frac{1}{\left(\mathbf{I} - RR^T\right)^2} R$$

so that we get (absorbing factors into λ)

$$\Delta C = \varepsilon V^T \frac{1}{\mathbf{I} - RR^T} R - \lambda V^T \frac{1}{\left(\mathbf{I} - RR^T\right)^2} R$$
$$= \varepsilon V^T \frac{1}{\mathbf{I} - RR^T} \left(\mathbf{I} - \frac{\lambda}{\varepsilon} \frac{1}{\mathbf{I} - RR^T}\right) R$$

which is easily transformed into that of the text by using $\gamma = 1 - \frac{\lambda}{\varepsilon}$.

References

[1] N. Ay, N. Bertschinger, R. Der, F. Güttler, and E. Olbrich. Predictive information and explorative behavior of autonomous robots. *The European Physical Journal B - Condensed Matter and Complex Systems*, 63(3):329–339, 2008.

[2] G. Baldassarre. Self-organization as phase transition in decentralized groups of robots: A study based on boltzmann entropy. In M. Prokopenko, editor, *Advances in Applied Self-Organizing Systems*, pages 127–146. Springer Verlag, 2008.

[3] A. G. Barto. Intrinsically motivated learning of hierarchical collections of skills. In *Proceedings of 3rd Int. Conference Development Learn.*, pages 112–119, San Diego, CA, USA, 2004.

[4] W. Bialek, I. Nemenman, and N. Tishby. Predictability, complexity and learning. *Neural Computation*, 13:2409, 2001.

[5] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley, 2006.

[6] J. P. Crutchfield and K. Young. Inferring statistical complexity. *Phys. Rev. Lett.*, 63:105–108, 1989.

[7] T. DelSole. Predictability and information theory. Part I: Measures of predictability. *J. Atmos Sci.*, 61(3):2425–2440, 2004.

[8] R. Der. Self-organized acquisition of situated behavior. *Theory in Biosciences*, 120:179–187, 2001.

[9] R. Der, F. Güttler, and N. Ay. Predictive information and emergent cooperativity in a chain of mobile robots. In *Artificial Life XI*. MIT Press, 2008.

[10] R. Der, F. Hesse, and G. Martius. Learning to feel the physics of a body. In *CIMCA '05: Proceedings of the International Conference on Computational Intelligence for Modelling, Control and Automation Vol-2 (CIMCA-IAWTIC'06)*, pages 252–257, Washington, DC, USA, 2005. IEEE Computer Society.

[11] R. Der, F. Hesse, and G. Martius. Rocking stamper and jumping snake from a dynamical system approach to artificial life. *J. Adaptive Behavior*, 14:105 – 116, 2005.

[12] R. Der and R. Liebscher. True autonomy from self-organized adaptivity. In *Proc. Workshop Biologically Inspired Robotics. The Legacy of Grey Walter 14-16 August 2002, Bristol Labs*, Bristol, 2002.

[13] R. Der and G. Martius. From motor babbling to purposive actions: Emerging self-exploration in a dynamical systems approach to early robot development. In S. Nolfi, editor, *From Animals to Animats*, volume 4095 of *Lecture Notes in Computer Science*, pages 406–421. Springer, 2006.

[14] R. Der, G. Martius, and F. Hesse. Let it roll – emerging sensorimotor coordination in a spherical robot. In L. M. Rocha, editor, *Artificial Life X*, pages 192–198. MIT Press, August 2006.

[15] P. Grassberger. Toward a quantitative theory of self-generated complexity. *Int. J. Theor. Phys.*, 25(9):907–938, 1986.

[16] F. Kaplan and P.-Y. Oudeyer. Maximizing learning progress: An internal reward system for development. *Embodied Artificial Intelligence*, pages 629–629, 2004.

[17] A. S. Klyubin, D. Polani, and C. L. Nehaniv. Empowerment: A universal agent-centric measure of control. In *Proc. CEC*. IEEE, 2005.

[18] A. S. Klyubin, D. Polani, and C. L. Nehaniv. Representations of space and time in the maximization of information flow in the perception-action loop. *Neural Computation*, 19:2387–2432, 2007.

[19] M. Lungarella, T. Pegors, D. Bulwinkle, and O. Sporns. Methods for quantifying the informational structure of sensory and motor data. *Neuroinformatics*, 3(3):243–262, 2005.

[20] G. Martius. *Goal-Oriented Control of Self-Organizing Behavior in Autonomous Robots*. PhD thesis, Georg-August-Universität Göttingen, 2010.

[21] G. Martius, J. M. Herrmann, and R. Der. Advances in Artificial Life, 9th European Conference, ECAL 2007, Lisbon, Portugal, September 10-14, 2007. volume 4648 of *Lecture Notes in Computer Science*, pages 766–775. Springer, 2007.

[22] P.-Y. Oudeyer, F. Kaplan, and V. Hafner. Intrinsic motivation systems for autonomous mental development. *Evolutionary Computation, IEEE Transactions on*, 11(2):265–286, April 2007.

[23] J. Pearl. *Causality*. Cambridge University Press, 2000.

[24] R. Pfeifer and J. C. Bongard. *How the Body Shapes the Way We Think: A New View of Intelligence*. MIT Press, Cambridge, MA, November 2006.

[25] R. Pfeifer, M. Lungarella, and F. Iida. Self-organization, embodiment, and biologically inspired robotics. *Science*, 318:1088–1093, 2007.

[26] M. Prokopenko, V. Gerasimov, and I. Tanev. Evolving spatiotemporal coordination in a modular robotic system. In S. Nolfi, G. Baldassarre, R. Calabretta, J. Hallam, D. Marocco, J.-A. Meyer, and D. Parisi, editors, *From Animals to Animats 9: 9th International Conference on the Simulation of Adaptive Behavior (SAB 2006)*, volume 4095 of *Lecture Notes in Computer Science*, pages 558–569. Springer, 2006.

[27] M. Prokopenko, P. Wang, D. Price, P. Valencia, M. Foreman, and F. A.J. Self-organizing hierarchies in sensor and communication networks. *Artificial Life, Special Issue on Dynamic Hierarchies*, 11(4):407–426, 2005.

[28] J. Schmidhuber. A possibility for implementing curiosity and boredom in model-building neural controllers. In *Proceedings of the first international conference on simulation of adaptive behavior on From animals to animats*, pages 222–227, Cambridge, MA, USA, 1990. MIT Press.

[29] J. Schmidhuber. Driven by compression progress: A simple principle explains essential aspects of subjective beauty, novelty, surprise, interestingness, attention, curiosity, creativity, art, science, music, jokes. *Anticipatory Behavior in Adaptive Learning Systems*, pages 48–76, 2009.

[30] L. Steels. The autotelic principle. *Embodied Artificial Intelligence*, pages 629–629, 2004.

[31] J. Storck, S. Hochreiter, and J. Schmidhuber. Reinforcement driven information acquisition in non-deterministic environments. In *Proceedings of the International Conference on Artificial Neural Networks*, pages 159–164, 1995.

[32] E. A. Theodorou, J. Buchli, and S. Schaal. Reinforcement learning of motor skills in high dimensions: A path integral approach. In *international conference of robotics and automation (icra 2010) - accepted*, 2010.

[33] K. Zahedi, N. Ay, and R. Der. Higher coordination with less control – a result of information maximization in the sensori-motor loop. *Adaptive Behavior*, to appear 2010.