Theory Biosci. (2012) 131:161–179
DOI 10.1007/s12064-011-0137-9

ORIGINAL PAPER

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

Nihat Ay · Holger Bernigau · Ralf Der ·
Mikhail Prokopenko

Received: 12 May 2011 / Accepted: 26 October 2011 / Published online: 29 November 2011

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 fundamental 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 article 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 article 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.

Keywords Autonomous systems · Predictive information · Self-organization · Sensorimotor loop · Embodiment · Hebbian learning · Intrinsic motivation

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 toward 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.

First results in that direction have already been obtained in Ay et al. (2008) for the case of one-dimensional systems. This article aims at generalizing those results in

N. Ay (✉) · H. Bernigau · R. Der · M. Prokopenko
Max Planck Institute for Mathematics in the Sciences, Leipzig, Germany
e-mail: nay@mis.mpg.de

N. Ay
Santa Fe Institute, Santa Fe, NM, USA

M. Prokopenko
CSIRO, Sydney, Australia

several respects, we consider (i) systems in arbitrary dimensions, get (ii) exact results for special model systems, (iii) derive explicit learning rules for the general case, and demonstrate that these learning rules are of a Hebbian like structure so that parallels to biological systems can be drawn. Last but not least we find, using our exact results, several surprising effects in stochastic oscillator systems that may help to better understand the efficiency of the information-theoretic approach in embodied robotic systems.

Information-theoretic aspects

The development of the information-theoretic approaches 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 randomness. Optimal in that sense would be a robot that behaves as random as possible. An alternative is Kolmogorov complexity, a measure on which Schmidhuber has based his approach to self-motivation and artificial curiosity, see Schmidhuber (2007) for an introduction. Moreover, in Lungarella et al. (2005) a set of univariate and multivariate statistical measures are introduced to quantify the information structure in sensory and motor channels. Generic information-theoretic criteria may vary in their emphasis: e.g., one may focus on maximization of empowerment (the perceived amount of influence or control that the agent has over the world; Anthony et al. 2009; Klyubin et al. 2005); minimization of heterogeneity across states of multiple agents, measured with either the variance of Shannon entropy of rule-space (Prokopenko et al. 2005) or Boltzmann entropy of swarm-bots' states (Baldassarre 2008); maximization of spatiotemporal coordination within a modular robot, measured via the excess entropy computed over a multivariate time series of modules' states (Prokopenko et al. 2006), etc.

What is common to these examples of information-driven self-organization is the characterization of the sensorimotor (or perception-action) loop in information-theoretic terms. For instance, empowerment measures the amount of Shannon information that the agent can, by executing the actions, inject into its sensors through the environment, affecting future actions and future perceptions. Technically, for a predefined agent's behavior, empowerment is defined as the capacity of the agent's actuation channel: the maximum mutual information (MI) for the channel over all possible distributions of the transmitted signal (i.e., actions) (Klyubin et al. 2005, 2007). On the other hand, the maximization of excess entropy during a time interval, used in Prokopenko et al. (2006), allows to change the controllers' logic (i.e., 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 by evolution 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).

Predictive information (PI)

This article studies in some detail the use of the predictive information of the sensorimotor process for the self-organization of robot behavior. 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 PI of a process quantifies the total information of past experience that can be used for predicting future events. Technically, it is defined as the MI between the future and the past, see Bialek et al. (2001). It has been argued that PI, also termed excess entropy (Crutchfield and Young 1989) and effective measure complexity (Grassberger 1986), is the most natural complexity measure for time series.

The behaviors emerging from maximizing the PI are qualified by the fact that PI 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 large variety of behaviors without becoming chaotic or purely random. In this working regime, somewhere between order and chaos, the robot may be expected to explore its behavioral possibilities in a most effective way. How and why this works is made more explicit in the concrete dynamical system investigated below.

Intrinsic motivation

The use of PI complements approaches that equip the robot with a motivation system producing intrinsic reward signals. Pioneering work has been done by Schmidhuber (1990) using the prediction error as a reward signal to make the robot curious for new experiences. The approach has been further developed in a number of papers, see e.g., Storck et al. (1995) and Schmidhuber (2009). Related ideas have been put forward in the so called play ground experiment by Kaplan and Oudeyer (2004; Oudeyer et al. 2007) using the learning progress as a reward signal. Steels (2004) proposes the Autotelic Principle, i.e., the balance of skill and challenge of behavioral components as the motivation for open-ended development, whereas Barto (2004) uses the prediction error of skill models to build hierarchical skill collections. PI 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. Theodorou et al. (2010), Kober and Peters (2009), and Engel (2010), because the PI is not restricted to discrete spaces at all. However, in this article, we will not follow this line but instead derive task-free learning rules directly from the gradient ascent on the PI.

PI and dynamical systems

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 approach 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 using recurrent neural networks as controller of the dynamical system consisting of body, brain, and environment, see Pfeifer and Bongard (2006) and Pfeifer et al. (2007) 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 article to present some first results in that direction. We start by answering the following question: Given a robot in its environment, how can we find an explicit learning rule for optimizing the behavior such that the PI 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. In this article, we approach that challenge by studying, in a first step, 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. Furthermore, the results are useful also in the non-linear case which is subject of our ongoing research.

In a recent article (Zahedi et al. 2010), a general learning rule has been derived from the PI using the natural gradient technique in a finite state-action space. This article complements that approach by the study of the case of continuous spaces and controllers realized by parameterized functions such as neural networks. The information-theoretic approach can also be considered as an alternative to the principle of homeokinesis (Der and Liebscher 2002; Der 2001), a systematic approach to the self-organization of behavior that has been applied successfully to a large number of complex robotic systems, cf. Der et al. (2005, 2006a, b), Der and Martius (2006, 2011). Moreover, this principle has also been extended to form a basis for a guided self-organization of behavior (Martius et al. 2007; Martius 2010; Martius and Herrmann 2010). We hope to benefit substantially from this parallel in future work.

Organization of the article

The general aim of the article is to establish PI as a systematic basis for the behavioral self-organization of autonomous robots. We start with formulating the sensorimotor loop in the language of probability theory as it is most appropriate for the information-theoretic approaches and give subsequently the formulation of specific model systems in the language of dynamical system theory. This part is of a more didactic nature providing the interested but not specialized reader with the relevant background. We introduce PI in "Predictive information" and study some basic properties without going much into detail since we find afterward explicit expressions for the PI and discuss general properties by the examples. The considered systems are of a structure well known from linear control theory so that several of the results are not new. We rederive them on an elementary basis to keep the article self-contained. New results are presented in "Example stochastic oscillator": we study stochastic oscillator systems in two dimensions and find some surprising effects. In our setting, the PI is shown to be maximal if the controller engenders a period 4 oscillation. Moreover, if the world is supporting a stochastic oscillation by itself, we find a resonance effect in maximizing the PI. This setting mimics specific embodied systems with latent oscillatory locomotion patterns that can be excited by the controller. Section "Learning rules—the self-referential robotic system" eventually introduces and discusses the explicit learning rules derived from a gradient ascent on the landscape of the PI over the controller parameters.

The sensorimotor loop

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

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 every specific situation but, to our

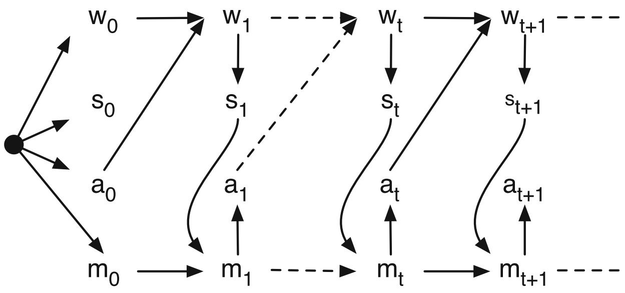

Fig. 1 Schematic representation of the sensorimotor loop. The state of the world at time t is w_t. The world is observed by the sensor values s_t which are then memorized by the internal state m_t. Actions are functions of the internal state m

experience, most of the situations encountered in robotics or biology are being covered. Quite generally, a kernel $p(y|x)$, where $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$, is a function $f : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^1$ assigning to each pair of vectors (x, y) a non-negative real number, such that $\int f(x, y)\,\mathrm{d}y = 1$ for every x. 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 $p(w_{t+1}|w_t, a_t)$ defining the probability density that the world state in the next time step is w_{t+1} given the world is now in state w_t and action a_t has been executed.
- The world state $w_t \in \mathbb{R}^{nw}$ observed at time t is mapped by the kernel $p(s_t|w_t)$ to the sensor state $s_t \in \mathbb{R}^{ns}$.
- However, the sensor process $(S_t)_{t\in\mathbb{N}}$ usually does not contain complete information. This situation is generally known as the problem of hidden variables. One way out of this is to introduce the memory m_t which includes earlier sensor values. In this way, part or all the missing information can be reintroduced by the memory. A theoretical foundation can be found in the embedding theorems well known form dynamical systems theory or time series analysis, cf. Kantz and Schreiber (2003). The time evolution of the memory m_t is given by the kernel $p(m_t|m_{t-1}, s_t)$.
- Actions are given in terms of the current memory by the kernel $p(a_t|m_t)$ which defines the policy of the agent.
- Internally, the agent constructs a world model represented by the kernel $p(s_{t+1}|m_t, a_t)$ defining the transition to the new sensor values in terms of the memory and the actions taken by the agent. Note that in general the world model will be valid only approximately.
- The aim of this article is to derive learning rules that allow the agent to adapt its behavior in the direction of higher PI. For this purpose, the agent has to make an estimate of the latter. This is possible in terms of the kernels $p(s_{t+1}|m_t, a_t), p(a_t|m_t),$ and $p(m_t|m_{t-1}, s_t)$.

In order to derive explicit expressions, we are going to consider the particularly simple situation where the memory is restricted to the last sensor vector s_t alone. In this situation, only the following two kernels are required:

$$p(s_{t+1}|s_t, a_t) \quad p(a_t|s_t) \tag{1}$$

In this specific case, the sensor process is more compactly defined by the kernel

$$p(s_{t+1}|s_t) = \int p(s_{t+1}|s_t, a)p(a|s_t)\,\mathrm{d}a \tag{2}$$

One of the aims of this paper is to use the information-theoretic measures, specifically the PI, in the dynamical systems theory of the sensorimotor loop. We, therefore, give in the following the relation between the probabilistic and the dynamical system formulation of the sensor process, restricting ourselves to the simple sensor process as defined in Eq. 1.

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. A general approach is given by the method of functional causal models (Pearl 2000). However, here we want to consider specific systems given by the kernels of the kind

$$p(s_{t+1}|a_t, s_t) = f(s_{t+1} - F(a_t, s_t)) \tag{3}$$

where $f : \mathbb{R}^n \rightarrow \mathbb{R}^1$ is a probability density function, i.e., a function with $f \geq 0$ and

$$\int f(u)\,\mathrm{d}u = 1$$

In general, the noise may be state-dependent. However, we will use only additive noise terms (no state dependence) in the following to get analytical results as a first step toward more general situations.

Consider the stochastic dynamical system

$$S_{t+1} = F(S_t, A_t) + \Pi_{t+1} \tag{4}$$

where $\{\Pi_t\}_{t\in\mathbb{N}}$ are independent, identically distributed random variables with values in $\mathbb{R}^n$ having the probability density function f. From a physical perspective the random variable Π_t is the noise at time t. Identical distribution of the noise implies time homogeneity of the noise process and independence is equivalent to the white noise property. Then S_t is a time-homogeneous Markov chain with transition kernel given by Eq. 3.

In a similar way, the action kernel is assumed to have the structure

$p(a|s) = g(a - K(s))$.

This is equivalent to a process A_t

$$A_t = K(S_t) + \Theta_t \tag{5}$$

where A_t has values in $\mathbb{R}^{n_A}$ and represents the vector of motor values at time t, whereas Θ_t has values in $\mathbb{R}^{n_A}$ and represents the actuator noise.

For most calculations, we will restrict ourselves to the case of Gaussian noise so that

$$f(u) = \frac{1}{\sqrt{(2\pi)^n |D_1|}} \exp\left(u^T D_1^{-1} u\right),$$

and

$$g(u) = \frac{1}{\sqrt{(2\pi)^n |D_2|}} \exp\left(u^T D_2^{-1} u\right),$$

where D_1 and D_2 are positive matrices. We use the notation

$$|L| = \det L$$

wherever this does not lead to ambiguities. Furthermore, Π_t and Θ_t are independent, Gaussian random variables with mean zero and covariance matrices given by

$$D_1 = E\left(\Pi_t \Pi_t^{\ T}\right) \quad \text{and} \quad D_2 = E\left(\Theta_t \Theta_t^{\ T}\right), \quad \text{respectively.}$$

Before going into a detailed discussion of such processes, let us first introduce the PI of a stochastic process.

Predictive information

The PI is the MI between the future and the past, relative to some instant of time t, of the sensor process $S = (S_t)_{t \in \mathbb{N}}$

$$\begin{aligned} I\left(S_{\text{past}}; S_{\text{future}}\right) &= \left\langle \ln \frac{p\left(S_{\text{past}}, S_{\text{future}}\right)}{p\left(S_{\text{past}}\right) p\left(S_{\text{future}}\right)} \right\rangle \\ &= H\left(S_{\text{future}}\right) - H\left(S_{\text{future}} | S_{\text{past}}\right) \end{aligned} \tag{6}$$

where the averaging is over the joint probability $p\left(S_{\text{past}}, S_{\text{future}}\right)$. Note that in the case of continuous variables we are dealing with differential entropies so that the individual entropy components $H(S_{\text{future}}), H\left(S_{\text{future}} | S_{\text{past}}\right)$ 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.

Equation 6 simplifies considerably if S is a Markov chain, see Ay et al. (2008), the case we want to consider in this article. In this case, the PI is given by the MI between two successive time steps, i.e., instead of Eq. 6 we consider

$$I(S_{t+1}; S_t) = \left\langle \ln \frac{p(S_{t+1}, S_t)}{p(S_{t+1}) p(S_t)} \right\rangle = H(S_{t+1}) - H(S_{t+1} | S_t) \tag{7}$$

which simplifies the sampling process considerably.

This expression also reveals the properties of the PI in most simple terms. The PI obviously is large if both $H(S_{t+1})$ is large and $H(S_{t+1}|S_t)$ is small. The first point means that the variability in the sensor values is high which is the case if the robot displays a high behavioral diversity. The second point means that S_t determines S_{t+1} very well. This is the case, if the behavioral diversity is a direct consequence of the actions of the robot in relation to the specific environmental conditions.

In experiments with a coupled chain of robots done earlier (Der et al. 2008), it was observed that the PI 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 effectively explore the arena. 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 the PI and other information measures for relatively simple sensor processes as is done in this article.

Example systems

Let us consider the PI for the case of a linear dynamics, i.e., we choose in Eqs. 4 and 5

$$K(s) = Cs \quad \text{and} \quad F(s, a) = Ts + Va \tag{8}$$

the matrix T representing the contribution to the sensor process due to some dynamics of the world alone and V represents the sensor response to the output of the controller. Note that many of the linear control systems studied in engineering are of this kind (with a different notation convention).

Under the assumptions made, any realization of the sensor process S_t is now given by

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

where

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

and $\xi_t = V\theta_t + \pi_t$ is the effective combination of controller and world noise. The kernel from Eq. 2 now becomes $p(s_{t+1}|s_t) = h(s_{t+1} - Rs_t)$, where h is the probability density of a Gaussian random variable with mean zero and covariance matrix $D = D_1 + VD_2V^T$.

PI in linear control systems

Equation 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 such as predictive information can be obtained in closed form. We rederive some of these results here in elementary ways to keep the article self-contained.

The process

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

$$\Xi_t \sim N(0, D) \quad \text{for all}\, t$$

The noise is assumed to be white so that $E\left(\Xi_t \Xi_{t'}^T\right) = 0$ for $t \neq t'$. Moreover, the spectral radius of R is assumed to be less than one so that $\lim_{t\to\infty} R^t s = 0$ for any vector s.

Equation 9 immediately implies:

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

Since the noise is white and since linear combinations of linearly transformed Gaussian random variables are Gaussian again, we get the following distribution of $S_{t+\tau}$ given S_t:

$$S_{t+\tau|t} \sim N(R^\tau S_t, \Sigma_{s_{t+\tau}|t}), \quad \text{where } \Sigma_{s_{t+\tau}|t} = \sum_{k=0}^{\tau-1} R^k D R^{kT} \tag{12}$$

Considering the limit $\tau \to \infty$, which exists since we assumed the spectral radius of R to be less than one, we find that, for any initial distribution, the sensor process converges strongly[1] to a centered Gaussian distribution with the covariance matrix

$$\Sigma_s = \sum_{k=0}^{\infty} R^k D R^{k^T} \tag{13}$$

Alternatively, Σ_s is easily shown to be the solution of the discrete Lyapunov equation

$$\Sigma_s = R \Sigma_s R^T + D \tag{14}$$

by simply rewriting Eq. 13.

These considerations show that the sensor process is ergodic and has a unique stationary distribution. From here on we assume that the initial distribution is that stationary distribution, such that $(S_t)_{t\in\mathbb{N}}$ becomes a stationary process.

Explicit expression

The conditional distribution of S_{t+1} with S_t given is a special case of Eq. 12 with $\tau = 1$:

$$P(S_{t+1}|S_t) = N(RS_t, D) \tag{15}$$

The value of the entropy of a Gaussian random vector is well known (see Cover and Thomas 2006 for example) as

$$H(S_{t+1}|S_t) = \frac{1}{2}\ln|D| + \frac{n}{2}\ln 2\pi e \tag{16}$$

Using the results from “The process,” we find:

$$H(S_t) = \frac{1}{2}\ln|\Sigma_s| + \frac{n}{2}\ln 2\pi e \tag{17}$$

Inserting the entropies from Eqs. 16 and 17 into 7 yields:

$$I(S_{t+1}; S_t) = \frac{1}{2}\ln|\Sigma_s| - \frac{1}{2}\ln|D|, \tag{18}$$

which is the entropy of the state minus that of the noise. Using the additive structure of the noise in Eq. 4, the formula

$$I(S_{t+1}; S_t) = H(S_t) - H(\Xi)$$

can be inferred in the same manner. Hence, this decomposition of the PI is a direct consequence of using additive noise.

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 $I(S_{t+1}; S_t) = H(S_t) - H(S_{t+1}|S_t)$ and the fact that entropies obey

$$H(QS_t) = H(S_t) + \ln|Q| \tag{19}$$

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 that the covariance matrix D commutes with R:

$$[D, R] = 0 \tag{20}$$

This property is always fulfilled if the noise is isotropic, i.e.,

$$D = \sigma^2 \mathbb{1} \tag{21}$$

where $\mathbb{1}$ is the unit matrix and σ^2 measures the overall strength of the noise. In this case, we get the covariance matrix of the state directly from Eq. 13 as

[1] Let P_t denote the distribution of S_t and let P denote the stationary distribution. Strong convergence means that $\int f\,\mathrm{d}P_t$ convergence to $\int f\,\mathrm{d}P$ for every bounded measurable function. This implies weak convergence (also known as convergence in distribution).

$$\Sigma_s = \sum_{k=0}^{\infty} R^k D R^{kT} = D\sum_{k=0}^{\infty} R^k R^{kT} = DM \tag{22}$$

where M is independent of the noise. Inserting this expression into Eq. 18, the PI is obtained as

$$I(S_{t+1}; S_t) = \frac{1}{2}\ln|M| = \frac{1}{2}\ln\left|\sum_{k=0}^{\infty} R^k \left(R^k\right)^T\right| \tag{23}$$

which is positive since $|M| > 1$. This is obvious since M is a sum of non-negative matrices, starting with the unit matrix.

If, moreover, the matrix R is normal (i.e., it commutes with its transposed R^T), we may rearrange the terms of the sum in Eq. 22 so that it is recognized as a geometric series, yielding

$$\sum_{k=0}^{\infty} R^k \left(R^k\right)^T = \sum_{k=0}^{\infty} \left(RR^T\right)^k = \frac{1}{\mathbb{1} - RR^T} \tag{24}$$

provided the spectral radius of R is less than one. In this special case, we obtain the explicit expression for the PI (as given also in Cover and Thomas 2006)

$$I(S_{t+1}; S_t) = -\frac{1}{2}\ln\left|\mathbb{1} - RR^T\right| \tag{25}$$

The results, Eqs. 23 and 25, show that the PI is independent of the noise in the isotropic noise case. It is defined entirely in terms of the dynamical operator R of the deterministic dynamical system although the PI does not make sense if there is no noise at all. As will be demonstrated below, if the noise is anisotropic, the PI displays a complicated interplay between dynamics and noise leading to interesting effects in the learning dynamics derived from maximizing the PI. In particular, we observe a self-organized frequency sweeping effect under specific conditions.

Arbitrary noise

The extension to general noise and arbitrary R is obtained in the following way. Using $|A|/|B| = \left|AB^{-1}\right|$ and the Lyapunov equation, Eq. 14, we write

$$\begin{aligned}\frac{|D|}{|\Sigma_s|} &= \left|\left(\Sigma_s - R\Sigma_s R^T\right)\left((\Sigma_s)^{-\frac{1}{2}}\right)^2\right| \\ &= \left|\left(\mathbb{1} - (\Sigma_s)^{-\frac{1}{2}} R\Sigma_s R^T (\Sigma_s)^{-\frac{1}{2}}\right)\right|\end{aligned}$$

Here, we used the invertibility of Σ_s. Introducing the whitening operation, see also DelSole (2004),

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

we obtain

$$\frac{|D|}{|\Sigma_s|} = \left|\mathbb{1} - WW^T\right|$$

where we used the symmetry of Σ_s. From Eq. 18 we obtain the PI also as

$$I(S_{t+1}; S_t) = -\frac{1}{2}\ln\left|\mathbb{1} - WW^T\right|. \tag{26}$$

The PI 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 obtained by means of the covariance matrix Σ of the stochastic process S_t (DelSole 2004). This generalizes the expression Eq. 25 to the case of anisotropic noise and arbitrary R in a straightforward way. However, the problem in this general case is that the matrix W is obtained only if the covariance matrix is already known. We will return to that point below when deriving the learning rules.

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 present them here 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. 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. In that case, the PI is defined entirely in terms of the dynamical operator R of the deterministic dynamical system, see Eq. 23. 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.

Example stochastic oscillator

Let us now consider a two-dimensional system 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

$$D = E\left(\Xi\Xi^T\right) = \sigma^2\begin{pmatrix} 1-m & 0 \\ 0 & 1+m \end{pmatrix}$$

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. We will also put $\sigma^2 = 1$ without loss of generality in the following derivations.

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 Eq. 10. The controller is linear and deterministic

$$a = Cs$$

with controller matrix C chosen as

$$C = cU(\phi)$$

where U is a rotation matrix

$$U(\phi) = \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix} \tag{27}$$

Moreover, we assume $V = \mathbb{1}$ so that the dynamical operator is

$$R = cU(\phi) \tag{28}$$

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.

Evaluating the PI

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

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

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

$$I(S_{t+1}; S_t) = \frac{1}{2}\ln\frac{|\Sigma_s|}{1-m^2} = \frac{1}{2}\ln\frac{1}{\left(c^2-1\right)^2} + \frac{1}{2}\ln\frac{1-m^{*2}}{1-m^2} \tag{29}$$

with

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

which can be considered as some kind of re-scaled noise asymmetry reflecting the interaction with the dynamics to reinterpret the result we rewrite this expression in the following way:

$$I(S_{t+1}; S_t) = \frac{1}{2}\ln\left|\frac{1}{1-RR^T}\right| + \frac{1}{2}\ln\frac{|D^*|}{|D|} \tag{30}$$

$$= I_{\text{iso}}(S_{t+1}; S_t) + \frac{1}{2}\ln\frac{|D^*|}{|D|} \tag{31}$$

where $I_{\text{iso}}(S_{t+1}; S_t)$ is the PI with isotropic noise as given in Eq. 25 and D^* is the noise matrix with m replaced by m^*. Equation 30 presents the PI as a term which depends only on the dynamics of the system, as in the isotropic noise case, plus a term reflecting the interplay of the dynamics with the anisotropy of the noise.

The parameter values ϕ that maximizes the PI given by Eq. 18 coincide with those values that maximize $|\Sigma_s|$, since the logarithm is a monotonic function and $|D|$ is a constant. Because there is no physical difference between two values of ϕ that differ by a multiple of 2π, there remain two essentially different values of ϕ that maximize the MI, namely $\phi_{1,2} = \pm\pi/2$. Both describe a period 4 damped oscillation, the positive value corresponding to a clockwise, the negative one to the counterclockwise rotation.

Gradient ascending the PI

The central idea of this article is to consider the PI as an intrinsic objective for the adaptation of the system towards increasing PI. A concrete realization of such a gradient ascent method is described in "Learning rules—the self-referential robotic system," below. However, we can already discuss the adaptation dynamics by gradient ascending the PI as given by Eq. 30. Let us start with the frequency parameter $\phi = 0$ and change the latter step by step into the direction of increasing PI. Obviously, this procedure drives ϕ to larger values until $\phi = \pi/2$ is reached, corresponding to the period 4 cycle of the system dynamics. We may consider this as a kind of frequency sweep, driven by the PI maximization dynamics, through the whole frequency space from frequency zero to the period 4 oscillation.

Resonance—a case for embodiment

A purely random world, i.e., $T = 0$, 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 special case of a two-dimensional system with a world dynamics given by an oscillatory system, i.e., put $T = wU(\omega)$ and $C = cU(\phi)$ so that (assuming $V = \mathbb{1}$ in this section)

$$R = cU(\phi) + wU(\omega) \tag{32}$$

where $U(\phi)$ was introduced in Eq. 27 above. We have to assume, moreover, that c and w are chosen such that the eigenvalues of RR^T are less than 1 so that the PI exists. The concrete conditions are given in Eq. 34 below.

Isotropic noise

The case of isotropic noise can be considered explicitly in the special case of two-dimensional systems. We consider again the series given by Eq. 22, and use the fact that in Eq. 32 R is a linear combination of unitary matrices and that all 2×2 unitary matrices commute and are normal. In higher dimensional spaces one can always find subspaces of commuting matrices so that the argument can partly be transferred also to the general case. Assuming normality of R, the PI is given in terms of the operator

$$\mathbb{1} - RR^T = (1 - \mu)\mathbb{1} \tag{33}$$

where using $\cos(\phi - \omega) = \cos\phi\cos\omega + \sin\phi\sin\omega$, we have

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

Equation 25 implies

$$I(S_{t+1}; S_t) = -\ln(1 - \mu) \tag{34}$$

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, i.e., $C = 0$ in Eq. 32) 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 as sketched in "Gradient ascending the predictive information" above. 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 in "The resonance effect" below.

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. Now, let us consider the case that the matrix $T = 0$ in the beginning of an experiment (no latent oscillation excited) and put $\phi = 0$. As discussed in "Gradient ascending the predictive information" above, with even the slightest anisotropy of the noise, the gradient ascent procedure increases the rotation angle ϕ (the frequency sweeping effect).

If, during this sweep, a mode in the world is excited and the world model, the matrix T in this case, 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 from the isotropic noise case, this scenario is not completely free of any sampling requirements. In the considered scenarios, we still need the covariance matrices D and Σ_s but in cases of practical interest, a coarse sampling of the latter might turn out sufficient. This is (weakly) supported by the fact that the discussed effects, in particular the sweeping and the resonance effect, set in as soon as there is any anisotropy of the noise at all. So, it would be sufficient to have a very coarse sampling and start the parameter adaptation process right from the outset. The sampling can continue during the information maximization so that, on the fly, the covariance matrices may be improved. A deeper reason may be seen in the dominance of the properties of the deterministic dynamical system so that the statistical properties are of secondary importance. But so far that is more or less speculation.

PI over several time steps

We may also consider the more general case of a larger step width, i.e., we investigate $I(S_{t+\tau}; S_t)$. The derivation for the PI is given in "Predictive information over several time steps". Most 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 $c = 0.9$ and $m = 0.8$ 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 $\tau > 1$.

Summary

This section has considered the application of the PI concept to a specific two-dimensional system, 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, and (ii) an oscillatory part of the same dynamical structure as the controller. 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 white 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.

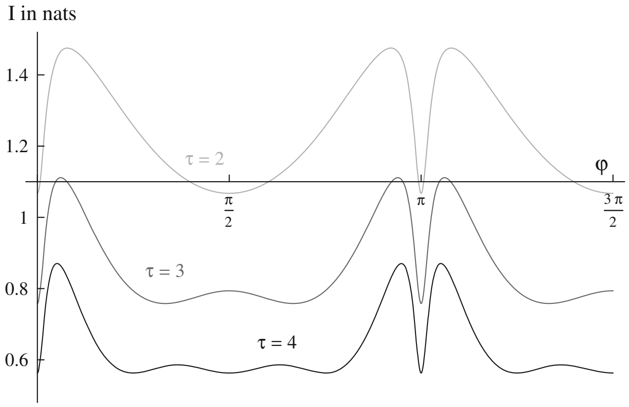

Fig. 2 The predictive information over τ time steps for the stochastic oscillator model as a function of the rotation angle ϕ for $\tau = 2$ (*upper*), $\tau = 3$ (*middle*), and $\tau = 4$ (*lower curve*). Instead of the maximum at $\pi/2$ observed for $\tau = 1$, there are two global maxima and one 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

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

$$\Delta p = \varepsilon \frac{\partial I(S_t; S_{t-1})}{\partial p} \tag{35}$$

where p is any parameter of the controller of the robot. In the present case, this are the matrix elements of the matrix C. The properties of the self-referential robotic system depend 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.

Explicit learning rules for the maximization of predictive information

Our aim now is the derivation of explicit rules for the gradient ascent dynamics over the parameter space. Let us start with the most simple case that the noise is isotropic and the matrix R is normal, i.e., the commutator $[R, R^T] = 0$, so that the explicit expression (25) is valid. We will start with the naive gradient as given in Eq. 35 and will discover that this is not appropriate in most cases since the emerging gradient dynamics does not conserve normality of R. Fortunately, this can be remedied by using the gradient with respect to a convenient metric. The idea is given in an intuitive way below and will be substantiated by introducing the metric in "Generalized gradient for obtaining a self-consistent update rule."

So, let us consider the naive gradient dynamics for the matrix C given by

$$\Delta C_{ij} = \varepsilon \frac{\partial I(S_t; S_{t-1})}{\partial C_{ij}}$$

leading to the explicit learning rule (see "Derivation of the learning rule")

$$\Delta C = \varepsilon V^T \frac{1}{\mathbb{1} - RR^T} R \tag{36}$$

Questions of the consistency of the learning rule are discussed below.

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 system dynamics is diverging whenever 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

$$K = I(S_{t+1}; S_t) - \lambda Tr(\Sigma_s) \tag{37}$$

as the new objective function to be maximized. The trace over Σ_s can be evaluated from Eq. 13 by elementary means. In the case of isotropic noise we get the rule, see "Derivation of the learning rule",

$$\Delta C = \varepsilon V^T \frac{1}{\left(\mathbb{1} - RR^T\right)^2}\left(\gamma \mathbb{1} - RR^T\right) R \tag{38}$$

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

$$RR^T = \gamma \mathbb{1} \quad \text{or} \quad \left(\gamma^{-\frac{1}{2}} R\right)\left(\gamma^{-\frac{1}{2}} R\right)^T = \mathbb{1} \tag{39}$$

hence $\left(\gamma^{-\frac{1}{2}} R\right)$ is an orthogonal matrix. In the stationary case, the PI becomes

$$I(S_{t+1}; S_t) = \frac{1}{2} \ln \frac{\varepsilon}{\lambda} \tag{40}$$

Hence, in the sequence of update steps, the dynamical matrix R converges towards an orthogonal matrix multiplied by $\sqrt{\gamma}$, the limit of the MI depending only on the ratio of λ and ε.

Consistency

The derivation of the learning rule (38) is based on two critical assumptions:

1. the noise is isotropic (more generally, D commutes with the matrix R, compare Eqs. (20, 21))
2. the matrix R is normal, i.e., $[R, R^T] = 0$ (see Eq. 25).

The learning rule is consistent, if the freshly learned policy matrix C leads to a transformation matrix R that is normal. Unfortunately, this is not the case if the update rule (38) is used, in other words

$$\left[(R + \Delta R), (R + \Delta R)^T\right] \neq 0 \text{ in general} \tag{41}$$

To see this, note that the update rule (38), together with $\Delta R = V \Delta C$, leads to a change ΔR of R which is given by

$$\Delta R = VV^T \frac{\varepsilon}{\mathbb{1} - RR^T} R$$

If $[VV^T, R] = 0$, the commutator in Eq. 41 vanishes indeed so that the normality of R is conserved by the learning dynamics under the given conditions. However, if $[VV^T, R] \neq 0$ consistency is not guaranteed. We will solve this problem by modifying the learning rule (38). As a first step we consider $I = I(S_t; S_{t-1})$ as a function of R and define

$$\widetilde{\Delta R} = \varepsilon \frac{\partial I}{\partial R} = \frac{\varepsilon}{\mathbb{1} - RR^T} R \tag{42}$$

This defines a dynamics in R space that conserves normality of R as is immediately proven by evaluating the commutator in Eq. 41. The essential point now is to find a rule for changing C such that we obtain $\widetilde{\Delta R}$ as given by Eq. 42. Using $R = VC + T$, this means to define the change $\widetilde{\Delta C}$ of C such that

$$\widetilde{\Delta R} = V \widetilde{\Delta C}$$

We want to ensure that each change $\widetilde{\Delta R}$ is feasible by changing C. Let us assume from here on that V^{-1} exists. Then we define $\widetilde{\Delta C}$ as $\widetilde{\Delta C} = V^{-1} \widetilde{\Delta R}$. As a result, we obtain the new learning rule

$$\widetilde{\Delta C} = \varepsilon V^{-1} \frac{1}{\mathbb{1} - RR^T} R \tag{43}$$

and correspondingly with damping

$$\widetilde{\Delta C} = \varepsilon V^{-1} \frac{1}{\left(\mathbb{1} - RR^T\right)^2}\left(\gamma \mathbb{1} - RR^T\right) R \tag{44}$$

With this consistent update rule normality of R is automatically conserved. Note that the consistent learning rule (44) involves a multiplication by a factor V^{-1} (unlike the naive learning rule (38) that involves a factor V^T). However, the stationarity condition

$$\gamma \mathbb{1} = RR^T \tag{45}$$

is immediately seen to agree with Eq. 39. As we will show in "Generalized gradient for obtaining a self-consistent update rule," the change $\widetilde{\Delta C}$ can be obtained as the gradient of the pertinent objective function (involving both the MI and the penalty term) with respect to some non-standard metric on the set of square matrices. As will be shown, this metric turns out to be the pull-back metric of the standard metric with respect to the map $C \mapsto VC + T$. This new metric also better reflects the symmetry of the problem. As can be inferred from Eq. 25, the PI is invariant with respect to conjugation of R with an orthogonal matrix O:

$$R \mapsto ORO^T$$

This symmetry is also represented by the standard metric $\langle \cdot, \cdot \rangle$ in R space (compare "Generalized gradient for obtaining a self-consistent update rule"). Using the pull-back of $\langle \cdot, \cdot \rangle$ as a metric on the space of controller matrices C ensures that whenever C_1 and C_2 are dynamically equivalent in the sense that they lead to conjugate values R_1 and R_2, they have the same norm.

The resonance effect

The result already reveals a specific feature of the predictive information maximization paradigm. As is obvious from the stationarity condition Eq. 45, the gradient dynamics 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. By virtue of Eq. 33 we find in this case, using $RR^T = \mu \mathbb{1}$ and $\mu = c^2 + w^2 + 2cw\cos(\phi - \omega)$ that the condition of stationarity can be written as

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

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 observed in "Resonance—a case for embodiment" was obtained with ϕ as the only parameter. We have seen there that maximizing the PI drives ϕ until the controller is in resonance with the external oscillation inherent in the matrix T. The present result shows that the strength and the frequency parameter, if driven both by the gradient ascent, heavily interfere with each other. Moreover, the resonance effect is less dominant than the drive for increasing c, so that the latter dominates the gradient dynamics in parameter space. As a results, the convergence of ϕ is stopped before it reaches the resonance frequency. Overall, this means that the resonance phenomenon does not disappear altogether but that it is not complete.

The general case

The above results have been obtained under the proviso that the matrix R obeys the commutation property $RR^T = R^TR$ and the noise is isotropic so that the PI is expressed by Eq. 25. If these conditions do not hold, we have to use the general expression given by Eq. 26 for the PI

$$I(S_{t+1}; S_t) = -\frac{1}{2}\ln\left|\mathbb{1} - WW^T\right| \tag{46}$$

The difference to the special case consists of the replacement of R by W. However, this does not mean that we can simply replace R by W also in the learning rules since

$$W = \Sigma_s^{-\frac{1}{2}} R \Sigma_s^{\frac{1}{2}} \tag{47}$$

with Σ_s depending on R, too. We now proceed as above, finding at first the gradient in the space of matrices R and afterward relate that to the C space.

The gradient $\partial I/\partial R$ is obtained by the chain rule in two steps. First of all we derive the (canonical) gradient of I with respect to W using Eq. 60 from "Derivation of the learning rule":

$$\frac{\partial I}{\partial W} = \frac{1}{\mathbb{1} - WW^T} W \tag{48}$$

The gradient of W with respect to R is tricky since Σ_s depends on R. It is possible to derive a power series in R but the resulting expression is rather complicated and unwieldy for computation. Therefore, using Eq. 47 and ignoring, in a rough approximation, the dependence of Σ_s on R, we get:

$$\frac{\partial W_{ij}}{\partial R_{kl}} \approx \sum_{mn} \left(\Sigma_s^{-\frac{1}{2}}\right)_{im} \frac{\partial R_{mn}}{\partial R_{kl}} \left(\Sigma_s^{\frac{1}{2}}\right)_{nj} = \left(\Sigma_s^{-\frac{1}{2}}\right)_{ik} \left(\Sigma_s^{\frac{1}{2}}\right)_{lj}$$

so that (using that Σ_s is symmetric)

$$\frac{\partial I}{\partial R_{kl}} = \sum_{ij} \frac{\partial I}{\partial W_{ij}} \frac{\partial W_{ij}}{\partial R_{kl}} \approx \left(\Sigma_s^{-\frac{1}{2}} \frac{1}{\mathbb{1} - WW^T} W \Sigma_s^{\frac{1}{2}}\right)_{kl}$$

Introducing

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

we also obtain

$$\frac{\partial I}{\partial R} \approx \frac{1}{\mathbb{1} - \hat{R}R^T} \hat{R}.$$

We again have to decide about the metric on the space of policy matrices. In order to achieve consistency with the anisotropic noise case considered in the last section, we follow the same arguments that lead from Eq. 42 to 43[2] and define

$$\Delta C := \varepsilon V^{-1} \frac{1}{\mathbb{1} - \hat{R}R^T} \hat{R}. \tag{49}$$

This expression generalizes Eq. 43 to the case of anisotropic noise and arbitrary R but is valid only approximately. We nevertheless include the result here with the hope that this might be a start for later developments coping with the missing terms.

In "Consistency," we argued that the standard metric on the space of transformation matrices R is a good choice

[2] This boils down to using the pull-back metric on the space of policy matrices again (compare with Eq. 71 from "Generalized gradient for obtaining a self-consistent update rule" in Appendix).

since it reflects the symmetry of the system. However, this symmetry is broken if the noise is anisotropic. In this case, the PI is invariant with respect to conjugation of both R and D with the same orthogonal matrix O:

$$R \mapsto ORO^T \quad \text{and} \quad D \mapsto ODO^T \tag{50}$$

Instead of using the standard metric on the space of dynamical matrices R as we did before, one could equip this space with a metric that represents the symmetry better, as for example

$$\tilde{g}(X, Y) = \frac{n}{TrD} Tr\left(X^T Y D\right).$$

The appropriate metric on the space of controller matrices would be $f^* \tilde{g}$ in this case (for notation and concepts see "Generalized gradient for obtaining a self-consistent update rule" in Appendix).

The Hebbian nature of the learning rule

The learning rule can be rewritten in many different forms. This section intends to show that there is a close relationship between the derived learning rules and Hebbian learning as we know it from neural networks. Moreover, we will also show how under specific conditions the matrix inversion can be avoided altogether.

Stochastic gradient ascent rule

Let us first consider the special case of normal matrices R and isotropic noise and assume that VV^T commutes with R. Using Eq. 22 and $D = \sigma^2 \mathbb{1}$, we have $\left(\mathbb{1} - RR^T\right)^{-1} = \Sigma_s D^{-1}$ so that the learning rule can be written as

$$\Delta C = \varepsilon V^T \Sigma_s R \tag{51}$$

where σ^2 is absorbed into ε. The covariance matrix 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. The covariance matrix Σ_s can approximately be obtained as a time average of $s_t s_t^T$ in a finite time window. This averaging procedure is done implicitly by the update rule

$$\Delta C_t = \frac{\varepsilon}{N} V^T s_t s_t^T R \tag{52}$$

where N defines the length of the time window (note that N update steps realize one update step in Eq. 51, i.e., $\Delta C \approx \sum_{1 \le t \le N} \Delta C_t$). This may be helpful in practical applications since it does not involve any matrix inversion, the update being fully determined by the current value of the state vector s_t.

The case of general matrices V can be treated as above by simply replacing V^T by V^{-1}.

Hebbian learning

The above rule can be still further modified to make a connection to neural network learning paradigms. Let us introduce the new vectors $\tilde{s}_t = R^T s_t$ and $\tilde{a}_t = V^T s_t$. In terms of these states, we write the learning rule (52) as (omitting time indices)

$$\Delta C_{ij} = \frac{\varepsilon}{N} \tilde{a}_i \tilde{s}_j \tag{53}$$

If VV^T does not commute with R, the consistent update rule should be used as explained in "Consistency." In this case $\tilde{a}_t$ has to be defined by $\tilde{a}_t = V^{-1} s_t$ instead.

We may consider C_{ij} as the synaptic strength of a linear neuron. Interpreting $\tilde{a}_i$ as signal at the output of the neuron i and $\tilde{s}_j$ 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 relate this Hebbian learning rules to the error back propagation methods 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 ($T = 0$)

$$s_{t+1} = V a_t + \xi_{t+1}$$

with $a_t = C s_t$, and interpret Va as the output of the top layer of the network so that

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

with a linear output function $g(z) = z$. The controller can also be represented as a neural network

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

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

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

The error backpropagation 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 s_t, according to that rule, from the output of the network (top layer) back to the output of the controller (bottom layer) yields

$$(\tilde{a}_t)_i = \left(V^T s_{t+1}\right)_i$$

which, in the learning step, Eq. 53, 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. 53.

Again, if the matrix $V\,V^T$ does not commute with R, we have to replace $(\widetilde{a}_t)$ by

$$(\widetilde{a}_t)_i = \left(V^{-1} s_{t+1}\right)_i$$

so that we do not have simple backpropagation. Instead, this operation defining $\widetilde{a}_t$ may be called a backprojection of the state s_{t+1} through the world model given by V, back to the output of the controller. It is interesting that the transition from backpropagation to backprojection is a result of modifying the metric in matrix space and we will come back to that result in a later article (Ay et al. 2011).

Summary

The aim of "Learning rules—the self-referential robotic system" is the derivation of explicit learning rules for the maximization of the PI. In our specific case of linear systems, we derive 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 is transformed into a purely local form, see Eq. 52, so that no matrix inversions are necessary. This is of much interest from the practical point of view.

We discuss several penalty terms (which are necessary in the case of linear systems) and demonstrate 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 is partly 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, corresponding, for example, to a locomotion pattern. Maximizing the PI via the learning mechanism leads to the recognition and amplification of that mode. This may also be understood as a kind of self-motivated exploration of bodily affordances of embodied robots.

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 question raises immediately further questions such as (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 PI 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 sensor responses 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.

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 (Ay et al. 2011). 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, such as intrinsic motivation and artificial curiosity.

Acknowledgments 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. Nihat Ay also acknowledges support by the Santa Fe Institute at the early stage of the paper. Mikhail Prokopenko thanks the Max Planck Institute of Mathematics in the Sciences in Leipzig, Germany, for support and hospitality at the Institute. The authors thank the anonymous reviewer for many important comments that helped to improve the paper substantially.

Appendix

Here, we derive some results used in the text.

PI over several time steps

In order to find the PI over τ time steps, we need the conditional entropy $H(s_{t+\tau}|s_t)$ of $s_{t+\tau}$ given s_t which is well known, see, for example, DelSol (2004). We rederive it here by elementary means from our previous results, starting with Eq. 12 to obtain

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

Hence, we find

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

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

$$H(S_{t+\tau}|S_t) = \frac{1}{2}\ln\left|\Sigma_{s_{t+\tau|t}}\right| + \frac{n}{2}\ln 2\pi e \tag{54}$$

so that

$$I(S_{t+\tau};S_t) = \frac{1}{2}\ln\frac{|\Sigma_s|}{\left|\Sigma_{s_{t+\tau|t}}\right|} \tag{55}$$

In analogy to the derivation of Eq. 26 we rewrite this as

$$I(S_{t+\tau};S_t) = -\frac{1}{2}\ln\left|\mathbb{1} - W_{\tau}W_{\tau}^T\right| \tag{56}$$

with the pre-whitened operator

$$W_{\tau} = \Sigma_s^{-\frac{1}{2}} R^{\tau} \Sigma_s^{\frac{1}{2}} \tag{57}$$

Derivation of the learning rule

We use the well-known formula for the derivative of the determinant $|X|$ of a regular matrix X (see for example Magnus and Neudecker 1988):

$$\frac{\partial}{\partial X_{ij}}|X| = |X|\left(X^{-1}\right)_{ji}$$

The chain rule implies

$$\frac{\partial}{\partial X_{ij}}\ln|X| = \frac{1}{|X|}\frac{\partial}{\partial X_{ij}}|X| = \left(X^{-1}\right)_{ji} \tag{58}$$

or more compactly

$$\frac{\partial}{\partial X}\ln|X| = \frac{1}{X^T}$$

If Y is a matrix-valued, differentiable function of X and f is a real-valued, differentiable function we get by the chain rule:

$$\frac{\partial}{\partial X_{ij}}f(Y(X)) = \sum_{k,l}\frac{\partial Y_{kl}}{\partial X_{ij}}\frac{\partial}{\partial Y_{kl}}f(Y(X))$$

Rewriting the contraction over k and l as a trace, the chain rule for matrices reads as:

$$\frac{\partial f(Y)}{\partial X_{ij}}(X) = Tr\left(\frac{\partial Y^T}{\partial X_{ij}}(X)\frac{\partial f}{\partial Y}(Y(X))\right)$$

or more compactly

$$\frac{\partial f(Y(X))}{\partial X} = Tr\left(\frac{\partial Y^T}{\partial X}\frac{\partial f(Y)}{\partial Y}\right) \tag{59}$$

It is also useful to remember that

$$Tr\left(A\frac{\partial X}{\partial X_{ij}}\right) = A_{ji}, Tr\left(A\frac{\partial X^T}{\partial X_{ij}}\right) = A_{ij}$$

or more symbolically

$$Tr\left(A\frac{\partial X}{\partial X}\right) = A^T, Tr\left(A\frac{\partial X^T}{\partial X}\right) = A$$

Using Eq. 58 and putting $Q = \mathbb{1} - RR^T$ we find (exploiting the symmetry of Q)

$$\begin{aligned}-\frac{\partial}{\partial R}\ln|Q| &= Tr\left(\frac{\partial(RR^T)}{\partial R}\frac{1}{Q^T}\right)\\ &= Tr\left(\frac{1}{Q^T}\frac{\partial R}{\partial R}R^T\right) + Tr\left(\frac{1}{Q^T}R\frac{\partial R^T}{\partial R}\right)\\ &= 2\frac{1}{Q}R\end{aligned}$$

and

$$-\frac{1}{2}\frac{\partial}{\partial R}\ln\left|\mathbb{1} - RR^T\right| = \frac{1}{\mathbb{1} - RR^T}R \tag{60}$$

Remembering $R(C) = VC + T$ and using the chain rule (59) again, we get for an arbitrary real-valued function f:

$$\frac{\partial f(R(C))}{\partial C} = Tr\left(\frac{\partial R^T}{\partial C}\frac{\partial f}{\partial R}\right) = Tr\left(\frac{\partial C^T}{\partial C}V^T\frac{\partial f}{\partial R}\right) = V^T\frac{\partial f}{\partial R} \tag{61}$$

Hence

$$-\frac{1}{2}\frac{\partial}{\partial C}\ln\left|\mathbb{1} - RR^T\right| = V^T\frac{1}{\mathbb{1} - RR^T}R \tag{62}$$

The gradient of the penalty term is obtained in the following way. With isotropic noise and assuming as above that $RR^T = R^TR$ we have to consider[3]

$$\frac{\partial}{\partial R}Tr\left(\frac{1}{\mathbb{1} - RR^T}\right) = 2Tr\left(\frac{1}{\mathbb{1} - RR^T}\frac{\partial R}{\partial R}R^T\frac{1}{\mathbb{1} - RR^T}\right) = 2\frac{1}{(\mathbb{1} - RR^T)^2}R \tag{63}$$

so that the gradient descent step yields (absorbing factors into λ)

$$\begin{aligned}\Delta C &= \varepsilon V^T\frac{1}{\mathbb{1} - RR^T}R - \lambda V^T\frac{1}{(\mathbb{1} - RR^T)^2}R \\ &= \varepsilon V^T\frac{1}{\mathbb{1} - RR^T}\left(\mathbb{1} - \frac{\lambda}{\varepsilon}\frac{1}{\mathbb{1} - RR^T}\right)R\end{aligned} \tag{64}$$

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

Generalized gradient for obtaining a self-consistent update rule

In this part of the appendix we will investigate the mathematical background of the consistent update rule (44) of the controller matrix C found in "Consistency." We will show that the consistent update rule is also a gradient ascent algorithm, where the gradient is taken with respect to some non-standard metric on the differentiable manifold of $n \times n$ matrices, denoted by $M(n)$. We will further characterize this metric as the pull-back of the standard metric under the map that links the value of the controller matrix C to the dynamical matrix R:

$$f : M(n) \to M(n); \quad C \mapsto R := VC + T. \tag{65}$$

As in "Consistency" we will consider systems with V being a non-singular square matrix only.

Furthermore, we will introduce a general class of metrics on matrix spaces that contains the standard metric, our pull-back metric as well as the right-invariant metric on the space of invertible matrices used for example by Amari (compare Amari 1998). These results can be used to modify gradients of matrix functions in various ways without changing the stationary points of the learning algorithms. We provide an explicit formula for the gradient with respect to a metric from this class. We hope that this might be useful to modify learning algorithms on matrix spaces.

In this section we assume some familiarity with basic differential geometric concepts (as can be found in any introductory book on differential geometry such as Spivak 1999, Willmor 1959, Kühnel 2006, or Kobayashi and Nomizu 1963).

As stated above we are considering the differentiable manifold $M(n)$ of all $n \times n$ matrices. The only chart we want to use here is the most obvious choice (in order to be consistent with the usual notation of differential geometry we write upper indices for the matrix entries here):

$$\phi^{(i,j)} : M(n) \to \mathbb{R}; \quad X \mapsto X^{i,j}$$

In the following summation will always be carried out over pairs of indices consisting of one upper and one lower index. In other cases the summation sign will be written down explicitly. A metric is a positive-definite, symmetric (differentiable) bilinear form

$$g_p : T_pM(n) \times T_pM(n) \to \mathbb{R}; \quad p \in M(n)$$

The coefficients of the metric tensor with respect to the chart $\phi^{(i,j)}$ are:

$$g_{p;(i,j),(k,l)} = g_p\left(e_{(i,j)}\big|_p, e_{(k,l)}\big|_p\right)$$

The metric gives rise to a gradient of a function h, denoted by $\mathrm{grad}_g[h](p) \in T_pM(n)$. The gradient points into the direction of the steepest ascent of the function h at this point and its length is equal to $\left|D_pf(p)[\hat{e}]\right|$, where $\hat{e}$ is the unit vector pointing into this direction. So the definition of the gradient involves metric structures on both spaces:

- on $\mathbb{R}$ (which is canonically given; even a change of metric does not influence the direction of the gradient since two metrics at a certain point $p \in \mathbb{R}$ differ by a constant multiple only)
- on M to specify the unit-sphere in the tangent space $T_p M$ over which the maximization is carried out.

An equivalent definition requires the gradient $\mathrm{grad}_g[f](p)$ to be the unique vector $v \in T_pM$ such that:

$$\forall w \in T_pM : (D_ph)[v] = g_p(v, w) \tag{66}$$

The components of the gradient are:
$\mathrm{grad}_g[h](p)^{(i,j)} = g_p{}^{(i,j),(k,l)}\partial_{(k,l)}h(p)$

where $g_p^{(i,j),(k,l)}$ denotes the inverse $n^2 \times n^2$ matrix of $\left(g_{p;(i,j),(k,l)}\right)_{(i,j),(k,l)}$. Since $M(n)$ is a linear space, it is most natural to identify the tangent space at a given point $p \in M(n)$ with $M(n)$ itself. The canonical scalar product is then given by

$$\langle X, Y\rangle_p := TrX^TY.$$

It implies the standard notion of a gradient in $\mathbb{R}^{(n^2)}$:

[3] Use the rule for the derivative of a matrix inverse with respect to some parameter p

$$\frac{\partial}{\partial p}B^{-1} = -B^{-1}\frac{\partial B}{\partial p}B^{-1}$$

$$\mathrm{grad}_{\langle\cdot,\cdot\rangle}[f](p)^{(i,j)} = \delta^{(i,j),(k,l)}\partial_{(k,l)}f(p) = \frac{\partial f}{\partial X^{(i,j)}}(p)$$

Consider the problem of consistency in "Consistency" again. In order to find the optimal parameter for the policy matrix C we would like to implement some learning algorithm of the form

$$C_{n+1} = C_n + \Delta C_n.$$

By changing C the transformation matrix R is changed indirectly so we have $R_n := f(C_n)$ (where f has been defined in equation 65). For consistency, ΔC_n has to be chosen such that the following two conditions hold:

1. $\mathbb{1} - R_n R_n^T$ is invertible for every n
2. the matrices R_n and R_n^T commute for every n, i.e., R_n is normal.

The first point is easily fulfilled, since the set of invertible matrices is open in $M(n)$. To see this, let R be an invertible matrix, and let ΔR be a matrix with $\|\Delta R\| < \left\|R^{-1}\right\|^{-1}$. Then an inversion in terms of the von-Neumann series shows that $R + \Delta R$ is also invertible, and:

$$(R + \Delta R)^{-1} = \sum_{k=0}^{\infty}\left(R^{-1}\Delta R\right)^k R^{-1}$$

Hence, a sufficiently small learning rate ensures the validity of point one.

The second point is more subtle. The set of normal matrices is the algebraic set $\{A \in M(n) | AA^T - A^T A = 0\}$.[4] Considering both the MI term and the penalty term (compare Eqs. 37 and 25), the objective function to be maximized is

$$K(R) := \epsilon \ln\left|\mathbb{1} - RR^T\right| - \lambda Tr\left(\frac{1}{\mathbb{1} - RR^T}\right)$$

for appropriate constants λ and ϵ. As shown in "Derivation of the learning rule" (see Eqs. 60 and 63), the gradient with respect to R becomes:

$$\mathrm{grad}_{\langle\cdot,\cdot\rangle}[K](R) = \epsilon\frac{1}{\mathbb{1} - RR^T}R - \lambda\frac{1}{\left(\mathbb{1} - RR^T\right)^2}R.$$

Notably the gradient commutes with R and R^T whenever R is normal. Hence an update rule of the form

$$R_{n+1} = R_n + \mathrm{grad}_{\langle\cdot,\cdot\rangle}[K](R_n) \tag{67}$$

preserves normality—R_{n+1} is normal whenever R_n is normal. However a naive update of C using the usual gradient might very well destroy normality of R_{n+1}. In order to overcome this problem we make use of the freedom to use another metric for the calculation of the gradient (compare Amari 1998). We summarize some well-known facts about the pull-back of a one-form. Let $h : M \to N$ be a differentiable map between manifolds and let g be a metric on N then the pull-back of g under h is defined by

$$(h^*g)_p(X, Y) = g_{f(p)}(D_p f[X], D_p f[Y]).$$

If h is a diffeomorphism, i.e., it is invertible with differentiable inverse, then the pull-back has the following properties:

- h^*g is a metric on M (i.e., it is a positive definite, symmetric two-form on M)
- Let $\phi : N \to \mathbb{R}$ be a differentiable function. According to the definition of concatenation and according to the chain rule the following two diagrams commute:

$$\begin{tikzcd} M \arrow[r, "h"] \arrow[rr, bend right, "\phi\circ h"'] & N \arrow[r, "\phi"] & \mathbb{R} \end{tikzcd}$$

$$\begin{tikzcd} T_p M \arrow[r, "D_p h"] \arrow[rr, bend right, "D_p \phi\circ h"'] & T_{h(p)} N \arrow[r, "D_{h(p)}\phi"] & \mathbb{R} \end{tikzcd}$$

Using the definition of the gradient Eq. 66, the following formula is valid for any $v \in T_p M$:

$$\begin{aligned} D_p(\phi \circ h)[v] &= h^*g_p\left(\mathrm{grad}_{h^*g}[\phi \circ h](p), v\right) \\ &= g_{h(p)}\left(D_p h\left[\mathrm{grad}_{h^*g}[\phi \circ h](p)\right], D_p h[v]\right), \end{aligned} \tag{68}$$

Using the relationship $D_p(\phi \circ h)[v] = D_{f(p)}(\phi)\left[D_p(h)[v]\right]$, the left-hand side of Eq. 68 can also be written in the following way:

$$D_p(\phi \circ h)[v] = g_{h(p)}\left(\mathrm{grad}_g[\phi](h(p)), D_p h[v]\right).$$

Since D_p h is non-singular this implies:

$$\mathrm{grad}_g[\phi](h(p)) = D_p h\left[\mathrm{grad}_{h^*g}[\phi \circ h](p)\right] \tag{69}$$

In our case, consider the map f defined in Eq. 65. Its differential is simply:

$$D_p f : M(n) \to M(n); \quad X \mapsto VX \tag{70}$$

Since f is affine, $D_p f$ even maps finite changes of C to the corresponding finite changes of R. The idea is to start with a matrix C_0 such that $R_0 = f(C_0)$ is normal and to update every C_n such that the change $\Delta C_n := C_{n+1} - C_n$ causes indirectly the desired change $\Delta R_n = \mathrm{grad}_{\langle\cdot,\cdot\rangle}[K](R_n)$ given

[4] The normal matrices are not a differentiable submanifold of (n^2) as one might think at first glance. Actually the dimension of the tangent space at a certain point (realized by the set of all matrices $X \in M(n)$ that do not change the commutator in first order, i.e., $[R + hX, R^T + hX^T] = \mathcal{O}(h^2)$ or $[X, R^T] = -[X, R^T]^T$) depends on R. To see that assume R to be symmetric, and assume that R has eigenvalues given by the n − tuple $(\lambda_1, \lambda_2, \ldots, \lambda_n)$. Then the dimension of the tangent space is n^2 minus the number of pairs of indices (i, j) with $i < j$ and $\lambda_i \neq \lambda_j$. The minimal dimension of the tangent space is $n \cdot (n + 1)/2$, achieved for matrices with n pairwise different eigenvalues, whereas the maximal dimension is n^2, achieved for multiples of the identity.

by Eq. 67. This can be achieved by using the pull-back metric:

$$g := f^*\langle\cdot,\cdot\rangle \qquad (71)$$

Indeed Eq. 69 gives in our case

$$\mathrm{grad}_{\langle\cdot,\cdot\rangle}[K](R_n) = (D_{C_n}f)\left[\mathrm{grad}_g[K \circ f](C_n)\right]$$

Inserting the explicit value of the differential into the definition of the pull-back metric gives:

$$g(X,Y) = Tr\left(X^T V^T V Y\right)$$

A short calculation yields the metric tensor

$$g_{(i,j),(k,l)} = (V^T V)^{i,k}\delta^{j,l}$$

And its inverse:

$$g^{(i,j),(k,l)} = \left((V^T V)^{-1}\right)^{i,k}\delta^{j,l}$$

Therefore, we have

$$\mathrm{grad}_g = \frac{1}{V^T V}\mathrm{grad}_{\langle\cdot,\cdot\rangle}$$

Using Eq. 64 the update rule for C becomes:

$$C_{n+1} = C_n + \epsilon V^{-1}\frac{1}{\mathbb{1} - R_n {R_n}^T}\left(\mathbb{1} - \frac{\lambda}{\epsilon}\frac{1}{\mathbb{1} - R_n {R_n}^T}\right)R_n$$

This is exactly the consistent update rule derived in "Consistency." Since the pull-back metric is the only metric that makes f an isometry, it is the natural choice to transfer metric properties from R space to C space. The pull-back metric lies in a certain class of metrics that we would like to present now. Note that the two-form

$$g'(X,Y)_p = Tr\left(G(p)X^T H(p)Y\right) \qquad (72)$$

is a scalar product if for each $p \in M(n)$ the matrices $G(p)$ and $H(p)$ are strictly positive (bilinearity is trivial, to see symmetry use the transposition invariance and the cyclic invariance of the trace, to see positivity and non-degeneracy write $G(p)$ and $H(p)$ as the square of a real symmetric matrix and notice that TrX^TX is zero if and only if $X = 0$.)[5]

A similar calculation as carried out for the pull-back metric before yields the following expression for the gradient:

$$\mathrm{grad}_{g'}[f](p) = H(p)^{-1}\left(\mathrm{grad}_{\langle\cdot,\cdot\rangle}[f](p)\right)G(p)^{-1}. \qquad (73)$$

Obviously the standard metric and our pull-back metric are members of this class (they are obviously flat since there is no point dependence of the metric coefficients in the standard chart). Another example is the right invariant metric on the set of invertible matrices, $GL(n)$, considered for example by Amari (1998):

$$h(X,Y)_W = Tr\left(W^{-1T}X^T Y W^{-1}\right)$$

Here, we have $H(W) = \mathbb{1}$, $G(W) = W^{-1} \quad W^{-1T}$ and therefore:

$$\mathrm{grad}_h[f] = \left(\mathrm{grad}_{\langle\cdot,\cdot\rangle}[f]\right)W^T W$$

Equations 72 and 73 are useful to modify the canonical gradient. As a consequence, a multiplication of the gradient by (possibly point-dependent) positive matrices from the left and from the right does not change the nature of the problem. Mathematically it is equivalent to a change of metric on the underlying space $M(n)$. This modification of the standard gradient can be done with several aims in mind, for example:

1. to simplify the standard gradient;
2. to eliminate unfeasible quantities that appear in the standard gradient;
3. to maintain some given constraints (such as normality of R in our case);
4. to make use of a further mathematical structure underlying the given problem, such as symmetries or invariance properties.

References

Amari S-I (1998) Natural gradient works efficiently in learning. Neural Comput 10:251–276

Anthony T, Polani D, Nehaniv CL (2009) Impoverished empowerment: 'meaningful' action sequence generation through bandwidth limitation. In: Kampis G, Szathmry E (eds), vol 2. Springer, Budapest, pp 294–301

Ay N, Bertschinger H, Der R, Güttler F, Olbrich E (2008) Predictive information and explorative behavior of autonomous robots. Eur Phys J B 63(3):329–339

Ay N, Bernigau H, Der R, Martius G (2011) Information-driven homeokinesis (in preparation)

Baldassarre G (2008) Self-organization as phase transition in decentralized groups of robots: a study based on Boltzmann entropy. In: Prokopenko M (ed) Advances in applied self-organizing systems. Springer, Berlin, pp 127–146

Barto AG (2004) Intrinsically motivated learning of hierarchical collections of skills. In: Proceedings of 3rd international conference development Learning, San Diego, CA, USA, pp 112–119

Bialek W, Nemenman I, Tishby N (2001) Predictability, complexity and learning. Neural Comput 13:2409

Cover TM, Thomas JA (2006) Elements of information theory. Wiley, New York

Crutchfield JP, Young K (1989) Inferring statistical complexity. Phys Rev Lett 63:105–108

[5] Note that this class of metrics is strongly related to the dynamical system considered in Theorem 1 of (Georgiev et al. 2001). If therein η is replaced by $A(W)$ and $F(W)^T F(W)$ is replaced by $B(W)$ where A and B are functions with values in the set of strictly positive matrices, then the considered dynamical system is just the gradient flow of J with respect to some metric from the class introduced above.

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

Der R (2001) Self-organized acquisition of situated behaviors. Theory Biosci 120:179–187

Der R, Liebscher R (2002) True autonomy from self-organized adaptivity. In: Proceedings of EPSRC/BBSRC international workshop on biologically inspired robotics. HP Labs, Bristol

Der R, Martius G (2006) From motor babbling to purposive actions: emerging self-exploration in a dynamical systems approach to early robot development. In: Nolfi S, Baldassarre G, Calabretta R, Hallam JCT, Marocco D, Meyer J-A, Miglino O, Parisi D (eds) Proceedings from animals to animats 9 (SAB 2006). LNCS, vol 4095. Springer, pp 406–421

Der R, Martius G (2011) The playful machine—theoretical foundation and practical realization of self-organizing robots. Springer, Berlin

Der R, Hesse F, Martius G (2005) Learning to feel the physics of a body. In: Proceedings of the international conference on computational intelligence for modelling, control and automation (CIMCA 06). IEEE Computer Society, Washington, DC, pp 252–257

Der R, Hesse F, Martius G (2006a) Rocking stamper and jumping snake from a dynamical system approach to artificial life. Adapt Behav 14(2):105–115

Der R, Martius G, Hesse F (2006b) Let it roll–emerging sensorimotor coordination in a spherical robot. In: Rocha LM, Yaeger LS, Bedau MA, Floreano D, Goldstone RL, Vespignani A (eds) Proceedings of the artificial life X, August. International Society for Artificial Life, MIT Press, pp 192–198

Der R, Güttler F, Ay N (2008) Predictive information and emergent cooperativity in a chain of mobile robots. In: Artificial Life XI. MIT Press, Cambridge

Engel Y (2010) Gaussian process reinforcement learning. In: Claude S, Geoffrey IW (eds) Encyclopedia of machine learning. Springer, pp 439–447

Georgiev P, Cichocki A, Amari S-I (2001) On some extensions of the natural gradient algorithm. In: Proceedings of the 3rd international conference on independent component analysis and blind signal separation, pp 581–585

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

Kantz H, Schreiber T (2003) Nonlinear time series analysis, 2nd ed. Cambridge University Press, Cambridge

Kaplan F, Oudeyer P-Y (2004) Maximizing learning progress: an internal reward system for development. In: Iida F, Pfeifer R, Steels L, Kuniyoshi Y (eds) Embodied artificial intelligence, Lecture Notes in Computer Science, vol 3139. Springer, pp 629–629

Klyubin AS, Polani D, Nehaniv CL (2005) Empowerment: a universal agent-centric measure of control. In: Congress on evolutionary computation, pp 128–135

Klyubin AS, Polani D, Nehaniv CL (2007) Representations of space and time in the maximization of information flow in the perception-action loop. Neural Comput 19:2387–2432

Kobayashi S, Nomizu K (1963) Foundations of differential geometry. Wiley, New York

Kober J, Peters J (2009) Policy search for motor primitives in robotics. In: Koller D, Schuurmans D, Bengio Y, Bottou L (eds) Twenty-Second annual conference on neural information processing systems, Red Hook, NY, USA, 06 2009, Curran, pp 849–856

Kühnel W (2006) Differential geometry, vol 16. American Mathematical Society Student Mathematical Library

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

Magnus J, Neudecker H (1988) Matrix differential calculus with applications in statistics and econometrics. Wiley, New York

Martius G (2010) Goal-oriented control of self-organizing behavior in autonomous robots. PhD thesis, Georg-August-Universität Göttingen

Martius G, Herrmann J (2010) Taming the beast: guided self-organization of behavior in autonomous robots. In: Doncieux S, Girard B, Guillot A, Hallam J, Meyer J-A, Mouret J-B (eds) From animals to animats 11. LNCS, vol 6226. Springer, pp 50–61

Martius G, Herrmann JM, Der R (2007) Guided self-organisation for autonomous robot development. In: Almeida e Costa F, Rocha L, Costa E, Harvey I, Coutinho A (eds) Proceedings of the advances in artificial life, 9th European conference (ECAL 2007). LNCSm, vol 4648. Springer, pp 766–775

Oudeyer P-Y, Kaplan F, Hafner V (2007) Intrinsic motivation systems for autonomous mental development. IEEE Trans Evol Comput 11(2):265–286

Pearl J (2000) Causality. Cambridge University Press, Cambridge

Pfeifer R, Bongard JC (2006) How the Body Shapes the Way We Think: A New View of Intelligence. MIT Press, Cambridge

Pfeifer R, Lungarella M, Iida F (2007) Self-organization, embodiment, and biologically inspired robotics. Science 318:1088–1093

Prokopenko M, Wang P, Price D, Valencia P, Foreman M, Farmer AJ (2005) Self-organizing hierarchies in sensor and communication networks. Artif Life 11(4):407–426

Prokopenko M, Gerasimov V, Tanev I (2006) Evolving spatiotemporal coordination in a modular robotic system. In: Nolfi S, Baldassarre G, Calabretta R, Hallam J, Marocco D, Meyer J-A, Parisi D (eds) From animals to animats 9: 9th international conference on the simulation of adaptive behavior (SAB 2006). Lecture Notes in Computer Science, vol 4095. Springer, pp 558–569

Schmidhuber J (1990) A possibility for implementing curiosity and boredom in model-building neural controllers. In: Proceedings of the first international conference on simulation of adaptive behavior. MIT Press, Cambridge, pp 222–227

Schmidhuber J (2007) Simple algorithmic principles of discovery, subjective beauty, selective attention, curiosity and creativity. Springer, Berlin

Schmidhuber J (2009) Driven by compression progress: a simple principle explains essential aspects of subjective beauty, novelty, surprise, interestingness, attention, curiosity, creativity, art, science, music, jokes. In: Pezzulo G, Butz MV, Sigaud O, Baldassarre G (eds) Anticipatory behavior in adaptive learning systems, Lecture Notes in Computer Science, vol 5499. Springer, pp 48–76

Spivak M (1999) Differential geometry, vol 1. Publish or Perish, Inc., Berkeley

Steels L (2004) The autotelic principle. In: Iida F, Pfeifer R, Steels L, Kuniyoshi Y (eds) Embodied artificial intelligence, Lecture Notes in Computer Science, vol 3139. Springer, pp 629–629

Storck J, Hochreiter S, Schmidhuber J (1995) Reinforcement driven information acquisition in non-deterministic environments. In: Proceedings of the international conference on artificial neural networks, pp 159–164

Theodorou EA, Buchli J, Schaal S (2010) Reinforcement learning of motor skills in high dimensions: a path integral approach. In: International conference of robotics and automation (ICRA 2010) (accepted)

Willmore T (1959) Differential geometry. Oxford University Press, Oxford

Zahedi K, Ay N, Der R (2010) Higher coordination with less control—a result of information maximization in the sensorimotor loop. Adapt Behav 18(3–4):338–355