Neurocomputing 400 (2020) 285–293

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Creating functionally favorable neural dynamics by maximizing information capacity

Elham Ghazizadeh [a,*], ShiNung Ching [a,b]

[a] Electrical and Systems Engineering Department, Washington University in St. Louis, 1 Brookings Drive, St. Louis, MO 63130, USA
[b] Biomedical Engineering and the Division of Biology and Biomedical Sciences, Washington University in St. Louis, 1 Brookings Drive, St. Louis, MO 63130, USA

ARTICLE INFO

Article history:
Received 5 June 2019
Revised 17 January 2020
Accepted 2 March 2020
Available online 10 March 2020

Communicated by Dr. Derui Ding

Keywords:
Information capacity
Neural dynamics
Empowerment

ABSTRACT

A ubiquitous problem in optimization and machine learning pertains to the design of systems that enact a desired behavior in dynamical environments. For example, the classical example of a control system that stabilizes an inverted pendulum. In this paper, we consider a complementary and less well-studied problem: the design of the environment itself. That is, can we create a dynamical system that in a general but mathematically rigorous way, is readily 'usable' by an unknown agent. We are especially interested in the synthesis of neuronal dynamics that are maximally labile with respect to afferent inputs. That is, can we create neural dynamics that propagate information well. To do so, we blend ideas from control and information theories, by turning specifically to the notion of empowerment, or the information capacity of a dynamical system in an input-to-state sense. We devise a strategy to optimize the dynamics of a system using empowerment over its state space as an objective function. This results in dynamics that are generically conducive to information propagation. For example, the optimized environment would be expected to perform well as an encoder (of afferent input distributions). We outline the key technical innovations needed in order to perform the optimization and, by means of example, discuss emergent dynamical characteristics of systems optimized according to this principle.

1. Introduction

A wide range of learning problems deal with the design of an optimal policy for an agent interacting with a dynamical environment, such as a controller attempting to balance an inverted pendulum. Here, a policy is a set of rules that determines the actions or inputs that the agent imparts on the environment based on observations. The fundamental goal of an optimal policy is to maximize a pre-specified objective function, such as a cumulative reward accrued over time [1,2].

In this paper we deviate from the classical notion of policy learning. Instead of seeking a policy for a fixed environment, we consider the scenario where the environment itself is variable, parameterized along a continuum of possible configurations. Our question of interest is then to find the particular environment that is maximally 'functional' according to a general usability criterion. In other words, we seek an environment that is most favorable, or perhaps usable, to an unknown agent. This problem is interesting in the context of computational neuroscience, because it relates to the fundamental problem of neural coding, or how circuits in the brain represent afferent information. For example, if we consider the environment as a set of biophysiological neuronal dynamics, we could study the sorts of kinetics that make a neuron a good encoder of afferent information. This could help us to derive general insights into the functionality of neural circuits. We are especially curious as to whether, there is a general principle that links the notion of maximal functionality to particular forms of dynamics.

In particular, we define the notion of functionality in terms of information theoretic quantities, since such quantities allow us to address the concept of optimality in a general way that is well-defined, task-independent and applicable universally to any agent-environment interaction. Specifically, we use as our objective the notion of empowerment, a measure of information processing capacity of a dynamical system [3–5]. At a conceptual level, empowerment can be understood as the maximum potential effect that an agent can impart on an environment through exogenous input. Thus, empowerment is a property of the input-to-output (or, input-to-state) transfer function of a system and is fundamentally related to the notions of channel capacity and reachability, from communication and control theories, respectively [5]. The

* Corresponding author.
E-mail addresses: elham@wustl.edu (E. Ghazizadeh), shinung@wustl.edu (S. Ching).

https://doi.org/10.1016/j.neucom.2020.03.008
0925-2312/

greater the empowerment of a system, the greater the diversity of states/outputs an agent can induce. As such, if empowerment is high, then the actions of the agent are well-encoded in the states/outputs of the system in question.

1.1. Related works:

The concept of empowerment was first introduced in [4] as a hypothetical, information-based utility function that might be considered as a formal definition of *intrinsic motivation* for learning and adaptation. Subsequently, empowerment has been widely used in reinforcement learning as a general framework for obtaining agent policies based on the value of information rather than manually-designed, task-specific utility functions [6–8]. The fundamental algorithm for evaluating empowerment is the Blahut–Arimoto (BA) algorithm [9], which is an enumeration-based approach with exponential computational complexity. However, the intensive computational complexity of empowerment calculation necessitates devising an effective method for its approximation. The authors in [10] proposed variational inference method as a scalable approach for empowerment approximation with the aim of using empowerment as a proxy for interinsically-motivated reinforcement learning. Similarly, the authors in [6] addressed the empowerment approximation over continuous systems for learning optimal policies via empowerment maximization.

The mathematical definition of empowerment is highly related to the 'Infomax' objective used in theoretical studies of neural coding [11–13].

1.2. Contributions:

The primary contribution of our work is a fundamental information-theoretic paradigm for system design. In particular, we seek to manipulate an environment so that it attains maximum empowerment. Concretely, we consider environments that can be mathematically represented as controlled dynamical systems. A primary motivation for this problem is a desire to understand the sorts of dynamics that are most advantageous for information encoding. We are most interested in revealing dynamical substrate of information processing in neurons and neuronal networks. That is, how do the dynamics of neurons facilitate information propagation and representation?

Our principle contributions are in terms of:

(i) the mathematical formulation of the above problem;
(ii) the use of the system impulse response to reduce the computational burden associated with empowerment calculation and optimization;
(iii) a method to find optimal environments, i.e., to solve (i). Particularly, our method involves considering a variable environment parameterized along a continuum of possible configurations;
(iv) illustration of the novelty of our ideas for typical and well-understood classical systems including linear/nonlinear dynamical systems and a well-known neural mass model, i.e., the Wilson–Cowan model.

As described, the main facet of our approach underlying these contributions involves adapting the information-theoretic definition of *empowerment* into the context of dynamical systems. Specifically, we use dynamical systems as (abstract) mathematical models of environments, for which we seek to optimize information capacity. Thus, all analysis and design is done at the level of the dynamical system, using the formalism of channel capacity/empowerment. In this sense, the terms *dynamical system, channel* and *environment* are interchangeable (i.e., dynamical systems are mathematical models of environments, in our context).

The remainder of the paper is structured as follows. In Section 2, we formalize our optimization problem. Sections 3 and 4 provide our proposed method for technical simplification of empowerment maximization and our method for designing usable environment, respectively. In Section 5, we provide simulation results illustrating the efficacy of empowerment maximization and the emergent dynamics. We conclude the paper and provide a brief discussion in sections 6.

2. Problem formulation and the environment model

In this section, we present the mathematical definition of empowerment as our objective function. Then, we provide the dynamical system model under consideration and specifically note the parameterization over which empowerment will be maximized. This problem setup allows the desirable dynamics of the environment to emerge only via the empowerment maximization criterion so as to form the environment maximally usable without having any a priori knowledge about the optimal dynamics. Throughout the paper the terms *dynamical system* and *environment* are used interchangeably.

2.1. Empowerment

Mutual Information is a core information theoretic quantity that measures the dependency between two random variables. In our formulation, the environment can be perceived as a communication channel from actions (its afferent inputs) to states. Given the current state $\boldsymbol{s}$, the mutual information between the action $\boldsymbol{a}$ and the final state $\boldsymbol{s'}$ is:

$$\mathcal{I}(\boldsymbol{a}, \boldsymbol{s'}|\boldsymbol{s}) = \mathbb{E}_{p(\boldsymbol{s'}|\boldsymbol{a},\boldsymbol{s})\omega(\boldsymbol{a}|\boldsymbol{s})} \log \frac{p(\boldsymbol{s'}, \boldsymbol{a}|\boldsymbol{s})}{p(\boldsymbol{s'}|\boldsymbol{s})\omega(\boldsymbol{a}|\boldsymbol{s})} \tag{1}$$

where we denote $\omega(\boldsymbol{a}|\boldsymbol{s})$ as the *action distribution* and $p(\boldsymbol{s'}|\boldsymbol{a}, \boldsymbol{s})$ as the *transition distribution*.

With this formulation of a dynamical system, empowerment is simply the *channel capacity*, i.e., the maximum information that an agent can potentially emit to the environment by manipulating the states [4,14] via actions/inputs. Particularly, for the given current state, the empowerment is:

$$\mathcal{E}(\boldsymbol{s}) = \max_{\omega(\boldsymbol{a}|\boldsymbol{s})} \mathcal{I}(\boldsymbol{a}, \boldsymbol{s'}|\boldsymbol{s}) \tag{2}$$

Here, $\boldsymbol{a}$ can be seen as an exogenous input (that is uncorrelated with the dynamics of the environment), referred to in the communication theory literature as a 'source'. Consequently, empowerment quantifies the extent to which the source is encoded in the system states.

2.2. Dynamical system model

We consider environments that can be described as dynamical systems of the affine form:

$$\dot{\boldsymbol{s}}(t) = f_{\boldsymbol{K}}(\boldsymbol{s}(t)) + \boldsymbol{a}(t) \tag{3}$$

where $f(\cdot) : \mathbb{R}^d \to \mathbb{R}^d$ denotes the vector field, which is in this case parameterized by $\boldsymbol{K}$. Here, $\boldsymbol{s}(t)$ and $\boldsymbol{a}(t)$ are the states and input actions at time t, respectively. We discretize the system using a fixed time-step dt, leading to:

$$\boldsymbol{s}_{t+1} = \boldsymbol{s}_t + (f_{\boldsymbol{K}}(\boldsymbol{s}_t) + \boldsymbol{a}_t)\, dt. \tag{4}$$

3. Simplified empowerment calculation via impulse response

The discrete model in Eq. (4) is a one-step autoregressive equation that describes the effect of an input action at the current time

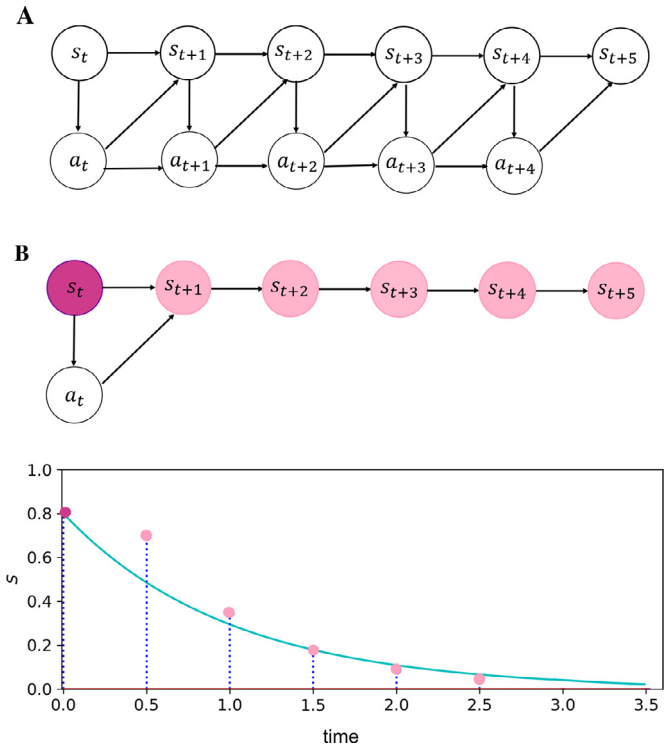

Fig. 1. Panel A shows the evolution of states over 5 sequence of actions. Panel B presents our proposed method of increasing the horizon of empowerment maximization. There is a one-to-one correspondence between the colors of the plots in panel B. The top plot shows the simplified method for evolution of states and the bottom one clarifies how the action is propagated through the dynamics. In this plot the continuous time response of the system is discretized for $dt = 0.5$.

on the state of the system at the subsequent time. When characterizing empowerment it is desirable to capture the effect of input actions over a prolonged temporal horizon [10,15,16]. In such a procedure, one obtains an optimal input action sequence leading to a trajectory within the state space (see, e.g., Fig 1. A where a sequence of 5 actions evolves the state according to (4)). However, calculating empowerment in this way is computationally expensive because there is exponential explosion in the search space (of actions/inputs) as a function of time steps.

Because of the dynamical nature of the environment, we posit a simplification of this idea wherein we instead calculate empowerment based on the system impulse response. We consider only a single step action $\boldsymbol{a}_t$ at time t, then allow the states to evolve over a horizon of n time steps i.e. $\boldsymbol{s}_{t+n}$ in (5). This constrains the search space (to a single input step), while nonetheless allowing the system dynamics to evolve the state over larger swaths of the state space. That is, we provide the dynamical system with enough time to nontrivially react to the input. This idea simplifies the schematic of the optimization problem, as shown in Fig. 1.B; comparing the graphs in panel A with B, we can see that using the scheme in panel A, the computational complexity of calculating empowerment is linearly proportional to the number of actions. However, using the impulse response simplification, we can reduce the computational complexity to a constant with respect to a single action.

For n time steps, we can obtain the final state as follows:

$$\boldsymbol{s}_{t+n} = \boldsymbol{s}_t + f_{\boldsymbol{K}}(\dots f_{\boldsymbol{K}}((f_{\boldsymbol{K}}(\boldsymbol{s}_t) + \boldsymbol{a}_t)))dt \tag{5}$$

where the number of recursions of $f_{\boldsymbol{K}}$ is n. To make the notion of empowerment meaningful it is necessary to introduce stochasticity to the above dynamics (in order to create nontrivial probability distributions). We do this by assuming additive noise in the readout of the system states, i.e.,

$$\hat{\boldsymbol{s}}_t = \boldsymbol{s}_t + \boldsymbol{w}_t, \tag{6}$$

where $\boldsymbol{w}_t$ is an uncorrelated noise process.

4. Empowerment maximization for creating usable environment dynamics

4.1. Environment parameterization

Thus far, we have set up the problem of calculating the empowerment of a dynamical system, as per Eq. (2). Our goal at this point is to treat the question of *maximizing* the empowerment of the system at hand with respect to its parameterization, i.e. $\boldsymbol{K}$. In particular, suppose that $\boldsymbol{K}$ in Eq. (5) represents a degree of freedom to alter the dynamics (e.g., in the case of a pendulum, altering the mass or length of the rod). We denote the number of discretized states of our considered system as M. For a given current state of the system, i.e. $\boldsymbol{s}_t^{(m)} \in \{\boldsymbol{s}_t^{(1)}, \dots, \boldsymbol{s}_t^{(M)}\}$, we posit the problem of empowerment maximization with respect to the parameterization $\boldsymbol{K}$ as follows:

$$\max_{\boldsymbol{K}} \mathcal{E}(\boldsymbol{s}_t) = \max_{\boldsymbol{K}} \max_{\omega} \frac{1}{M} \sum_m \mathcal{I}(\boldsymbol{a}_t, \boldsymbol{s}_{t+n} | \boldsymbol{s}_t^{(m)}) \tag{7}$$

4.2. Variational empowerment maximization

The problem (7) is challenging since it involves nested maximization over a continuous state space and nontrivial probability distributions. *Prima facie*, this is near intractable. Thus, in order to proceed, an approximation is required and for this we turn to the popular approach of using a variational lower bound for the key quantities [6,10,17,18]. We will specifically adopt the variational lower bound for empowerment approximation in [6], which enables evaluation of mutual information for continuous variables. However, we will use this computational approach in a different context (i.e., to design environments) and with an additional level of approximation to aid tractability. We proceed to discuss these contributions.

Specifically the major computational challenge alluded to above arises from the intractability of the probability terms in Eq. (1) and the integration over the continuous domain of all actions and states. To circumvent the mentioned issues, we can rewrite Eq. (1) as:

$$\mathbb{E}_{p(\boldsymbol{s}_{t+n}|\boldsymbol{a}_t,\boldsymbol{s}_t)\omega(\boldsymbol{a}_t|\boldsymbol{s}_t)} \log \frac{p(\boldsymbol{a}_t|\boldsymbol{s}_{t+n}, \boldsymbol{s}_t)}{\omega(\boldsymbol{a}_t|\boldsymbol{s}_t)} \tag{8}$$

where $p(\boldsymbol{a}_t|\boldsymbol{s}_{t+n})$ is the *posterior distribution* of actions. In a Bayesian sense, the action distribution, $\omega(\boldsymbol{a}_t|\boldsymbol{s}_t)$, is the prior.

Due to the intractability of the true posterior distribution, we use the mutual information variational bound, introduced in [11], to approximate the above equation as follows:

$$\hat{\mathcal{I}}(\boldsymbol{s}_{t+n}, \boldsymbol{a}_t|\boldsymbol{s}_t) = \mathbb{E}_{p(\boldsymbol{s}_{t+n},\boldsymbol{a}_t|\boldsymbol{s}_t)} \log \frac{q(\boldsymbol{a}_t|\boldsymbol{s}_{t+n}, \boldsymbol{s}_t)}{\omega(\boldsymbol{a}_t|\boldsymbol{s}_t)} \tag{9}$$

where $q(\boldsymbol{a}_t|\boldsymbol{s}_{t+n}, \boldsymbol{s}_t)$ is the *variational distribution* that approximates the true posterior. Using the variational approximation method, obtaining the variational distribution can be considered as an optimization problem where $q_\xi(\boldsymbol{a}_t|\boldsymbol{s}_{t+n}, \boldsymbol{s}_t)$ is a variational family of distributions with parameter ξ [19]. Given that the variational distribution expressively represents the true posterior distribution, a tight variational lower bound can be achieved [6]. Hence, we can obtain a variational lower bound on the empowerment as follow:

$$\hat{\mathcal{E}}(\boldsymbol{s}_t) = \max_{\omega,q} \hat{\mathcal{I}}(\boldsymbol{s}_{t+n}, \boldsymbol{a}_t|\boldsymbol{s}_t) \tag{10}$$

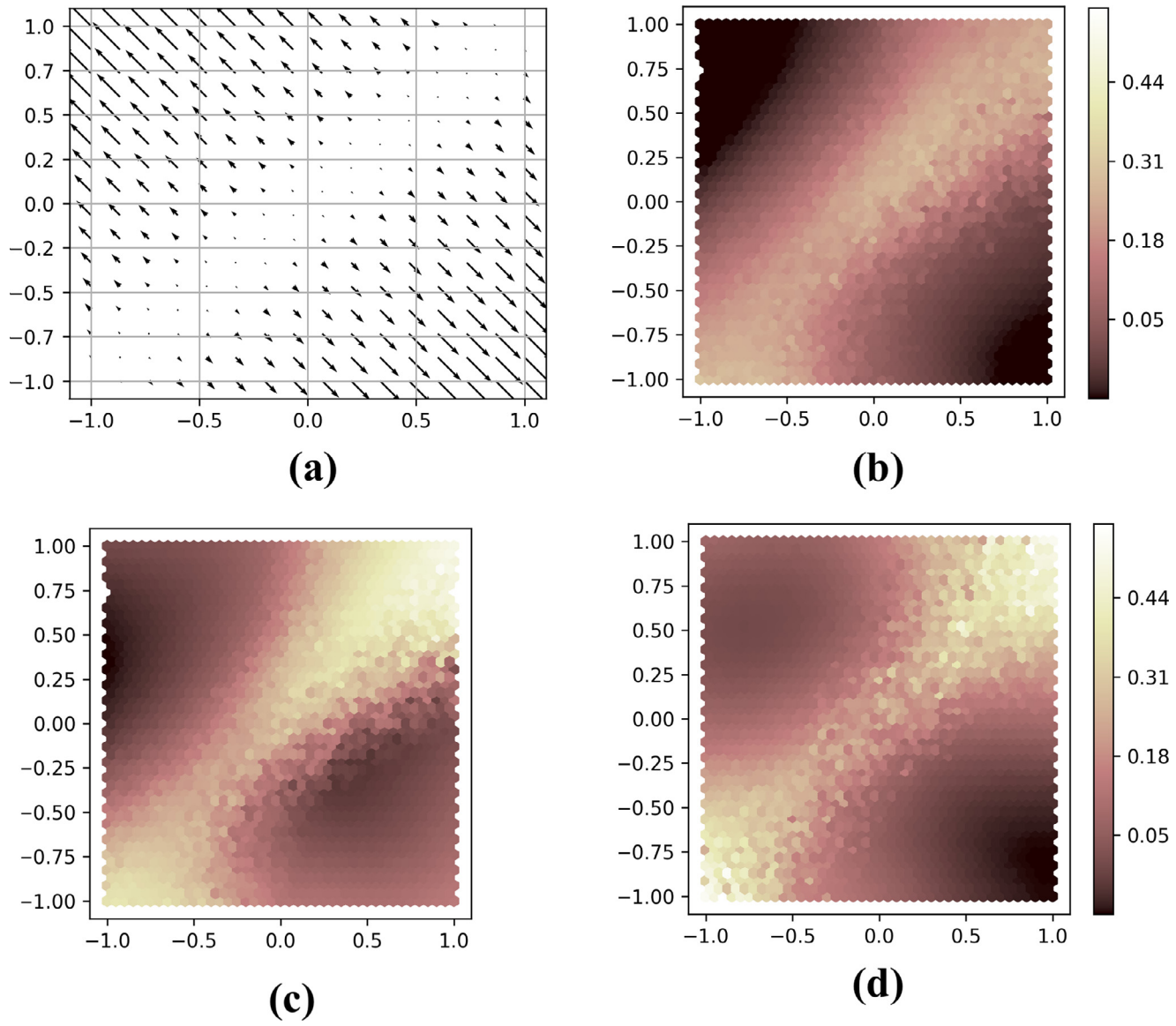

Fig. 2. The vector field and empowerment landscape result from n step empowerment calculation. Plot (a) shows the state space and its corresponding vector field of a 2 dimensional linear system and (b) depicts the corresponding empowerment landscape (in *nats*) optimized over two time steps of states. Plot (c) and (d) depict the comparison of empowerment (in *nats*) landscape over 5 time steps of states (i.e., the proposed impulse response method) versus 5 time steps of actions, respectively.

To perform the optimization of the variational bound, we can obtain the action distribution and the variational distribution via deep neural networks parameterized by ϕ and ξ, respectively. Sharing the same ideas with amortized variational inference [20,21], using neural networks(NNs) as a mapping from states to the distribution parameters enables us to only deal with finite number of neural network parameters (i.e. weights and biases) instead of learning a separate action distribution and variational distribution for each state. In this work, we choose the action and variational distributions from the Gaussian family as follows:

$$\begin{aligned} \omega(\boldsymbol{a}_t|\boldsymbol{s}_t^{(m)}) &= \mathcal{N}(\mu_\phi(\boldsymbol{s}_t^{(m)}), \sigma_\phi^2(\boldsymbol{s}_t^{(m)})I) \\ q(\boldsymbol{a}_t|\boldsymbol{s}_t^{(m)}, \boldsymbol{s}_{t+n}) &= \mathcal{N}(\mu_\xi(\boldsymbol{s}_t^{(m)}, \boldsymbol{s}_{t+n}), \sigma_\xi^2(\boldsymbol{s}_t^{(m)}, \boldsymbol{s}_{t+n})I) \end{aligned} \tag{11}$$

where mean μ and variance σ are also parameterized by NNs. $\theta = \{\phi, \xi\}$ are the joint parameter set of Eq. (10). Via joint optimization of variational bound w.r.t. parameters of ω and q, we can achieve the action distribution that maximizes the mutual information and the variational distribution that is responsible for the tightness of the variational lower bound. Exploiting the variational lower bound on empowerment, we pose our optimization problem as follows:

$$\max_{\boldsymbol{K}} \max_{\theta} J(\boldsymbol{K}, \theta), \tag{12}$$

where

$$J(\boldsymbol{K}, \theta) = \frac{1}{M} \sum_m \hat{\mathcal{I}}(\boldsymbol{s}_{t+n}, \boldsymbol{a}_t|\boldsymbol{s}_t^{(m)}) \tag{13}$$

and

$$\hat{\mathcal{I}}(\boldsymbol{s}_{t+n}, \boldsymbol{a}_t|\boldsymbol{s}_t^{(m)}) = \mathbb{E}_{p(\boldsymbol{s}_{t+n}, \boldsymbol{a}_t|\boldsymbol{s}_t^{(m)})}[\log q(\boldsymbol{a}_t|\boldsymbol{s}_{t+n}, \boldsymbol{s}_t^{(m)}) - \log \omega(\boldsymbol{a}_t|\boldsymbol{s}_t^{(m)})] \tag{14}$$

We can perform the joint Monte-Carlo sampling method to obtain samples from the joint distribution, $p(\boldsymbol{s}_{t+1}, \boldsymbol{a}_t|\boldsymbol{s}_t^{(m)})$, and use the reparameterization trick [22,23] to evaluate the stochastic

gradients of the objective function with respect to $\boldsymbol{K}$ and θ as:

$$\frac{\partial}{\partial\theta}\hat{\mathcal{I}}(\boldsymbol{s}_{t+n},\boldsymbol{a}_t|\boldsymbol{s}_t^{(m)}) \approx \frac{1}{L}\sum_l \frac{\partial}{\partial\theta}[\log q(\boldsymbol{a}_t^{(l)}|\boldsymbol{s}_{t+n}^{(l)},\boldsymbol{s}_t^{(m)}) - \log\omega(\boldsymbol{a}_t^{(l)}|\boldsymbol{s}_t^{(m)})] \tag{15}$$

and

$$\frac{\partial}{\partial\boldsymbol{K}}\hat{\mathcal{I}}(\boldsymbol{s}_{t+n},\boldsymbol{a}_t|\boldsymbol{s}_t^{(m)}) \approx \frac{1}{L}\sum_l \frac{\partial}{\partial\boldsymbol{K}}[\log q(\boldsymbol{a}_t^{(l)}|\boldsymbol{s}_{t+n}^{(l)},\boldsymbol{s}_t^{(m)}) - \log\omega(\boldsymbol{a}_t^{(l)}|\boldsymbol{s}_t^{(m)})] \tag{16}$$

where L is the number of samples. Algorithm 1 summarizes the optimization procedure. It is worth mentioning that if f_K is differentiable, a gradient-based approach can be used to update the decision variables. When examining convergence, we look at the magnitude of the gradients with respect to the decision variables and the magnitude of the cost function, J. However, we do not set any formal thresholds on these quantities, since they depend on the dynamics and parameters of the system under consideration.

Algorithm 1 Maximization of Empowerment Variational Lower Bound, w.r.t. $\theta = \{\phi, \xi\}$ and $\boldsymbol{K}$.

Initialize uniform samples from state space, $\{\boldsymbol{s}_t^{(m)}\}_{m=1,\dots,M}$
repeat
for each $\boldsymbol{s}_t^{(m)}$ **do**
draw one sample from $\omega(\boldsymbol{a}_t|\boldsymbol{s}_t^{(m)})$:
$\boldsymbol{a}_t \sim \omega(\boldsymbol{a}_t|\boldsymbol{s}_t^{(m)})$
transit to the final state $\boldsymbol{s}_{t+n}$
end for
J = $\frac{1}{M}\sum_m \log q(\boldsymbol{a}_t|\boldsymbol{s}_t^{(m)},\boldsymbol{s}_{t+n}) - \log\omega(\boldsymbol{a}_t|\boldsymbol{s}_t^{(m)})$
$\theta \leftarrow \theta + \eta_\theta \nabla_\theta J$ for r epoches
$\boldsymbol{K} \leftarrow \boldsymbol{K} + \eta_{\boldsymbol{K}} \nabla_{\boldsymbol{K}} J$
until convergence

5. Results

We proceed to show the efficacy of the proposed method on three canonical examples: linear dynamical systems, the inverted pendulum on fixed pivot and the Wilson–Cowan neural mass model. We then discuss the emergent dynamics of these systems following empowerment maximization.

5.1. Efficacy of the proposed impulse response procedure

First, we demonstrate the efficacy of the proposed empowerment calculation procedure.

5.1.1. Linear system

For the 2-dimensional linear systems, Eq. (5) is:

$$\boldsymbol{s}_{t+n} = \boldsymbol{s}_{t+n-1} + (\boldsymbol{A}\boldsymbol{s}_{t+n-1} + \boldsymbol{a}_t)dt + \boldsymbol{w}_{t+n}, \tag{17}$$

where $\boldsymbol{A} \in \mathbb{R}^{2\times 2}$ determines the vector field of the system. In this example, $\boldsymbol{A}$ is chosen so that the origin is unstable. Fig. 2(b) and (c), depicts the computed empowerment over the state space for horizons of 2 and 5 steps, respectively. As it is seen, the states around the origin have largest empowerment, which is intuitive given the dynamics of the system (because the origin is unstable, inputs applied for these initial conditions have the potential to access the largest swaths of state space).

As means of comparison, we have also performed calculation of the empowerment using a full iteration of actions (i.e., as in Fig. 1A). Fig. 2(d) shows the obtained empowerment landscape for this case. As seen, the proposed impulse response-based approach compares favorably to the full solution. Full details regarding simulations are found in Appendix.

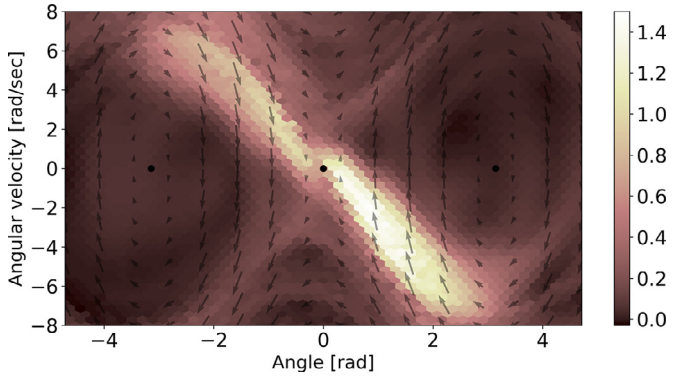

Fig. 3. The pendulum vector field and its corresponding empowerment landscape (in *nats*) obtained via empowerment calculation over 50 step states.

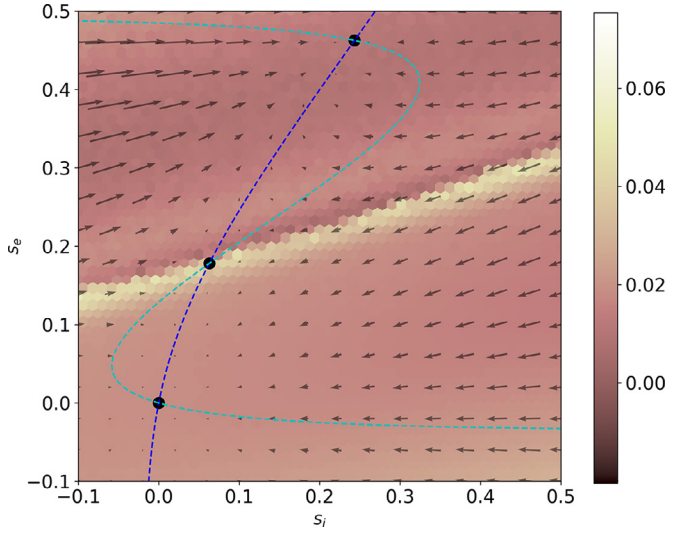

Fig. 4. The vector field, nullclines and empowerment landscape result from n step empowerment calculation for the Wilson-Cowan model (the empowerment values are in *nats*).

5.1.2. Pendulum

To show the performance of our method for nonlinear systems, we studied the simple pendulum with the following dynamics

$$\ddot{s} = -\frac{3g}{2l}\sin(s+\pi) + \frac{3}{ml^2}(u - 0.05\dot{s}) \tag{18}$$

For the sake of comparison with previous works [6,15], we considered the optimization over $n = 50$ steps. As seen in Fig. 3, the impulse response method produced an interpretable empowerment landscape, wherein the locus of high empowerment corresponds with the unstable manifold associated with the equilibrium at the origin (i.e., the pendulum in the inverted position). Like the unstable linear system, this is intuitive because on this manifold the input is able to drive the system to a larger swath of the state space. This is exactly equivalent to the idea of reachability in control systems analysis.

5.1.3. The wilson–Cowan model

We also demonstrate the performance of our method for the canonical neural mass model, the Wilson–Cowan model, which provides a course-grained representation of the overall activity of populations of neurons [24–26]. The 2-dimensional system describing the excitatory and inhibitory activity of these populations is:

$$\begin{aligned}\dot{s}_e &= -s_e + (1 - r_e s_e)\mathcal{F}_e(c_1 s_e - c_2 s_i + I_e + P)\\ \dot{s}_i &= -s_i + (1 - r_i s_i)\mathcal{F}_i(c_3 s_e - c_4 s_i + I_i + Q)\end{aligned} \tag{19}$$

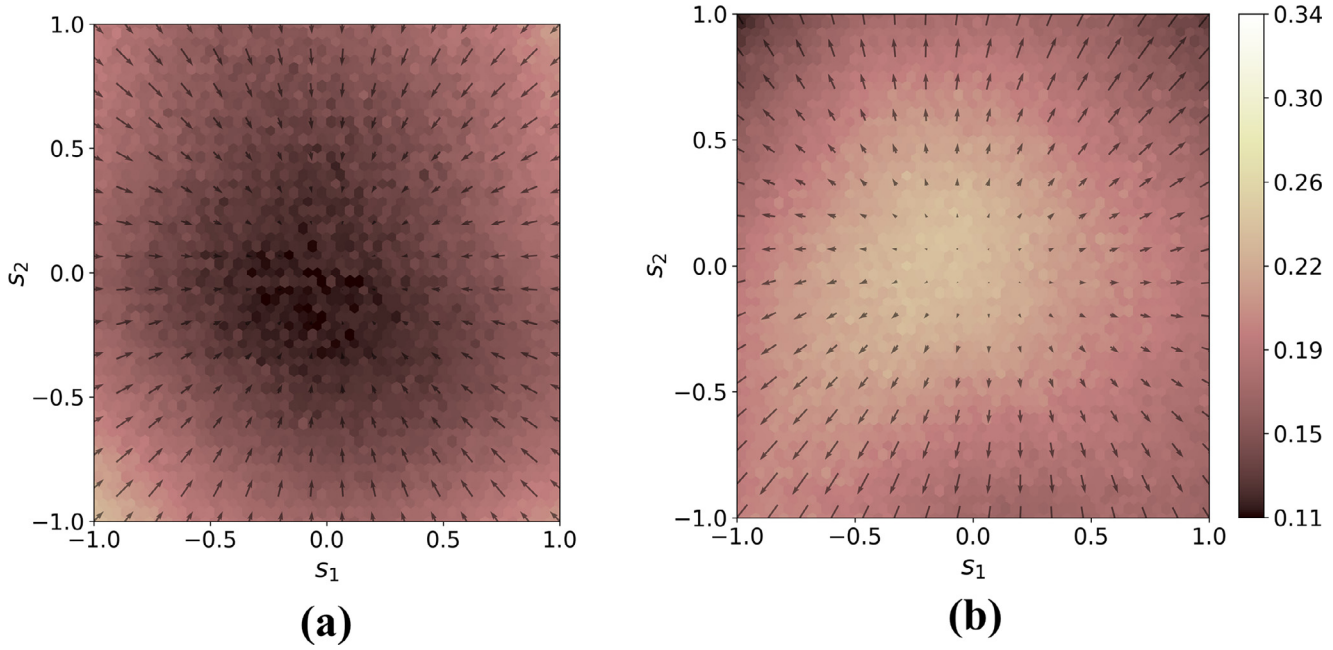

Fig. 5. Empowerment maximization of a linear dynamical system/environment with one stable equilibrium point (a). After optimization, the resultant environment (b) exhibits a dominant unstable equilibrium.

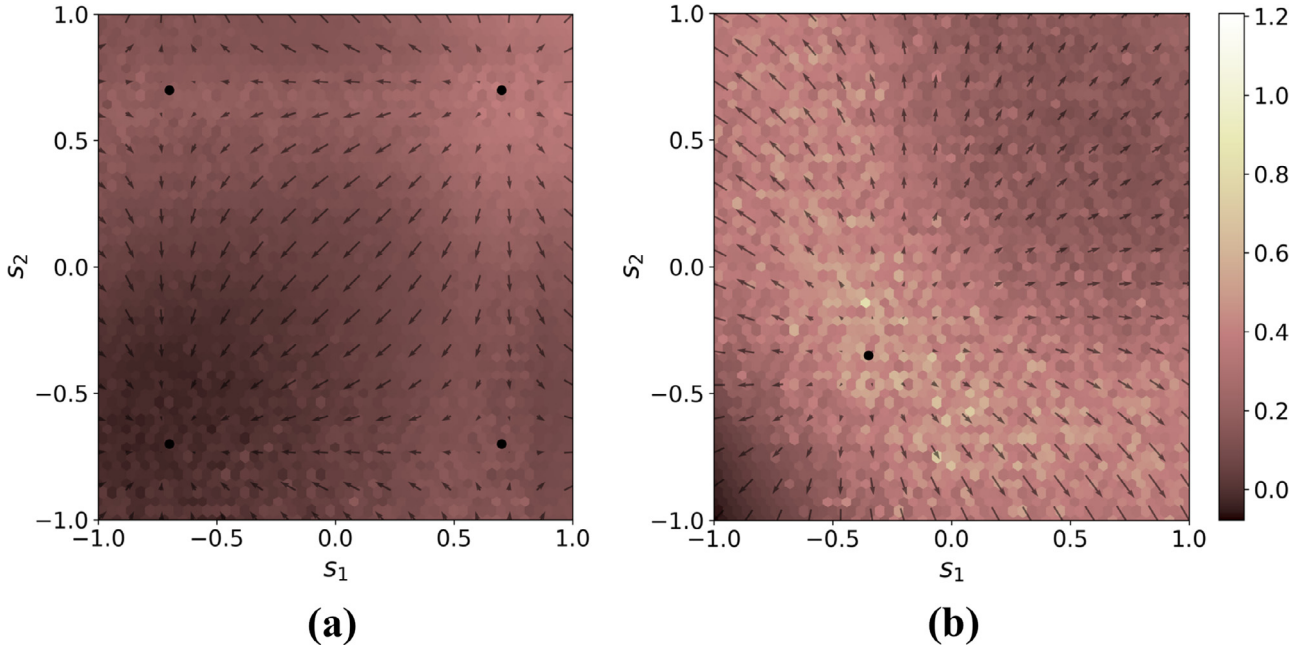

Fig. 6. Empowerment maximization of a nonlinear dynamical system/environment with four equilibrium points (a). After optimization, the number of equilibria is reduced and a single dominant unstable node emerges (b).

where s_e and s_i are the overall activity in the excitatory and inhibitory populations. Here, r_j, $j \in \{i, e\}$ represents a constant describing the refractory period. c_1, c_2, c_3 and c_4 present the strength of excitatory and inhibitory interactions. P and Q are the excitation level in the system and I_e and I_i are the control input currents that affect the respective populations. Also, $\mathcal{F}_j$ is a sigmoid function:

$$\mathcal{F}_j(s) = \frac{1}{1+\exp[-a_j(s-\theta_j)]} - \frac{1}{1+\exp(a_j\theta_j)} \tag{20}$$

where a and θ are free parameters representing the slope and the threshold, respectively. We applied our method to the Wilson-Cowan model over $n = 100$ steps. Particularly, we study the case wherein the system exhibits 3 multiple fixed points (2 stable fixed points and an unstable fixed point, see Appendix for details). Fig. 4, depicts the empowerment landscape obtained using the impulse response method. As seen, the locus of high empowerment corresponds to the unstable manifold associated with the unstable fixed point.

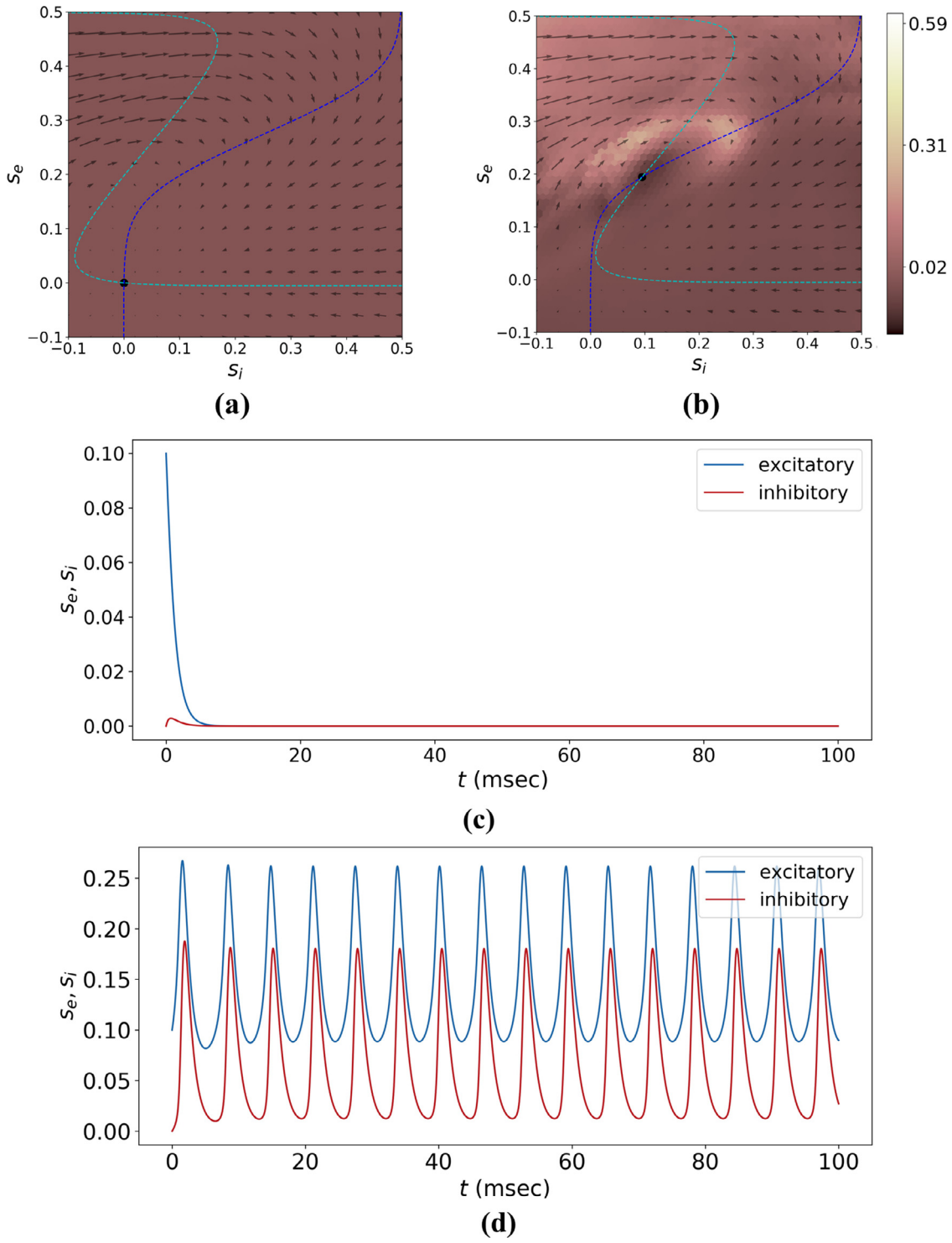

Fig. 7. Comparison of the empowerment landscapes and the environments dynamics for empowerment maximization in the Wilson–Cowan model. In (a), the system initially exhibits a stable fixed point and the empowerment landscape is flat ($P = 0$). In (b), after learning the optimal environment parameterization, i.e. P, the environment exhibits a limit cycle ($P = 1.177$). (c) and (d) present the comparison between the time response of the system before and after empowerment maximization.

5.2. Creating usable environments

We now move to the main problem considered in this paper: optimization of the empowerment with respect to the system/environment dynamics. We performed this study for the same systems considered above.

5.2.1. Linear and nonlinear systems

For simplicity, we first consider a horizon of two steps and a linear parameterization as follows:

$$\boldsymbol{s}_{t+2} = \boldsymbol{s}_t + f(f(\boldsymbol{s}_t) + \boldsymbol{K}\boldsymbol{s}_t + \boldsymbol{a}_t)dt + \boldsymbol{w}_{t+2} \tag{21}$$

Fig. 5, shows the vector field and the landscape of a stable linear system before and after optimization of the linear parameterization (again, see Appendix for details). Perhaps intuitively, the optimization has taken the original system (wherein the origin is asymptotically stable) and destabilized it. As a consequence, the environment is more 'usable' in the sense that a greater diversity of states is accessible (asymptotically) to an agent. The usable environment in Fig. 5(b) depicts higher empowerment around the center as opposed to Fig. 5(a).

We also consider the case of a nonlinear system with multiple equilibria (again, see Appendix for details). As shown, the optimization tends to accentuate the dominance of a single unstable equilibrium point. In Fig. 6(a), the basal system has 4 equilibria (2 saddle, one stable and one unstable equilibrium). The optimization leaves a dominant unstable equilibrium and destroys the rest of them, Fig. 6(b).

These results are intuitive insofar as the optimization appears to be simplifying the environment dynamics and making as much of the state space to be accessible or reachable as possible. Thus, an agent interacting with the optimized environments can 'do more' than one interacting with the basal system. Although we have considered only the simplified case of a linear parameterization, there are clear ways to generalize this concept including parameterizing the dynamics along a basis set of nonlinear functions.

5.2.2. The wilson–Cowan model

We carry out further simulations for the Wilson-Cowan model to look into the sorts of dynamics that emerge after learning the optimal parametrization of the environment. In Fig. 7, we have considered P in Eq. (19) as the environment parameterization and performed the empowerment optimization. Fig. 7(a) and (b) depicts the corresponding empowerment landscape of the initial and the optimal environment. As seen, the emergent dynamics from empowerment maximization for $n = 300$ create a limit cycle. This observation again is intuitive since the limit cycle permits different initial conditions to remain 'separated' asymptotically (versus converging to a single stable fixed point). This observation is also intriguing from a scientific perspective, since such limit cycle oscillations are thought to be pervasive in actual neuronal dynamics.

6. Discussion and conclusion

In our work, we have explored the use of empowerment as a way to shape an environment so as to render it more usable by an agent. From the perspective of dynamical systems and control theory, this problem amounts to altering the dynamics of the system at hand so that as much of the state space is reachable as possible. To do so, we suggest a computational simplification to aid in the calculation of empowerment (noting that the calculation of empowerment is itself an optimization problem over the space of agent actions/inputs). We then use a variational approach to facilitate the optimization of the environment as intended. Our simulation results show that the maximally usable environment created via this procedure leads to 'simple' dynamics with a single unstable fixed point and a large number of reachable states, which agrees with basic intuition.

Returning to the idea alluded to in the Introduction, one possible avenue for these results is to enable a study of biophysical neuronal dynamics. That is, we can fashion neurons as environments and study how their dynamics mediate information capacity. This may help us to derive general insights into the functionality of neural circuits and also reveal principles for network design that do not require task specificity. Finally, one can connect these results back to the general idea of agent policy design, by viewing the parameterization as a degree of freedom that can be chosen by the agent, so that it can both shape the environment and then exploit it according to a secondary objective. It is also worth mentioning that in our simulations, we constrained the degree of freedom as a linear parametrization of the environment; a potential future direction is augmenting the environment according to a more general parameterization, for example through the use of basis functions that allow for nonlinear manipulation of the vector field.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Elham Ghazizadeh: Conceptualization, Methodology, Software, Validation, Writing - original draft, Writing - review & editing, Visualization. **ShiNung Ching:** Conceptualization, Methodology, Writing - original draft, Writing - review & editing, Supervision, Funding acquisition.

Acknowledgments

ShiNung Ching holds a Career Award at the Scientific Interface from the Burroughs-Wellcome Fund. This work was partially supported by AFOSR 15RT0189, NSF ECCS 1509342 and NSF CMMI 1653589, from the US Air Force Office of Scientific Research and the US National Science Foundation, respectively.

Appendix

In this section, we provide the details of parameters used in our simulations. Particularly, the dynamical model parameters for pendulum and the Wilson-Cowan model are obtained from the corresponding references [6,24]. For the remaining, i.e. linear and nonlinear systems, the model parameter are chosen to highlight the key idea of empowerment maximization with respect to the system dynamics.

A.1. Dynamical models parameters:

In Fig. 2

$$A = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$$

In Fig. 3, the parameter set-up for pendulum are the same as the model provided in [6].

In Fig. 4, $c_1 = 12, c_2 = 4, c_3 = 13, c_4 = 11, a_e = 1.2, a_i = 1, \theta_e = 2.8, \theta_i = 4$ and $P = Q = 0$. We have set $[I_e, I_i] = [u_1, u_2]$.

We obtained excitatory and inhibitory nullclines by setting $\dot{s}_e = 0$ and $\dot{s}_i = 0$, respectively. The fixed points of the system are located where the nullclines intersect.

In Fig. 5(a)

$$A = \begin{bmatrix} -0.5 & 0 \\ 0 & -0.5 \end{bmatrix}$$

In Fig. 5(b)

$$K = \begin{bmatrix} 0.626 & 0.032 \\ 0.022 & 0.674 \end{bmatrix}$$

In Fig. 6(a)

$$f(\boldsymbol{s}) = \begin{cases} s_x^2 - 0.5 \\ s_y^2 - 0.5 \end{cases}$$

In Fig. 6(b)

$$K = \begin{bmatrix} 1.533 & -0.414 \\ -0.430 & 1.566 \end{bmatrix}$$

In Fig. 7, $c_1 = 16, c_2 = 12, c_3 = 15, c_4 = 4, a_e = 1.3, a_i = 2, \theta_e = 4\theta_i = 3.7$. We have set $[I_e, I_i] = [u_1, u_2]$. In Fig. 7(b), $k = P$.

A.2. Simulation parameters:

We have used TensorFlow [27,28] framework for our implementations. The NNs structure for our simulations are as follows:

Linear system: $d = 2$,

$\omega(\boldsymbol{a}_t|\boldsymbol{s}_t)$: 16 ELU + 16 ELU + {d idenity, d exp}

$q(\boldsymbol{a}_t|\boldsymbol{s}_t, \boldsymbol{s}_{t+n})$: 16 ELU + 16 ELU + {d idenity, d exp}

Pendulum: $d = 2$,

$\omega(\boldsymbol{a}_t|\boldsymbol{s}_t)$: 128 tanh + 128 tanh + 128 tanh + 128 tanh + {1 idenity, 1 exp}

$q(\boldsymbol{a}_t|\boldsymbol{s}_t, \boldsymbol{s}_{t+n})$: 128 tanh + 128 tanh + 128 tanh + 128 tanh + {1 idenity, 1 exp}

Nonlinear system: We have used the same structure as the linear one.

The Wilson-Cowan model: $d = 2$,

$\omega(\boldsymbol{a}_t|\boldsymbol{s}_t)$: 16 ELU + 16 ELU + 16 ELU + {d idenity, d exp}

$q(\boldsymbol{a}_t|\boldsymbol{s}_t, \boldsymbol{s}_{t+n})$: 16 ELU + 16 ELU + 16 ELU + {d idenity, d exp}

References

[1] R.S. Sutton, A.G. Barto, Reinforcement Learning: An Introduction, MIT press, 2018.

[2] J. Garcıa, F. Fernández, A comprehensive survey on safe reinforcement learning, J. Mach. Learn. Res. 16 (1) (2015) 1437–1480.

[3] A.S. Klyubin, D. Polani, C.L. Nehaniv, Keep your options open: an information-based driving principle for sensorimotor systems, PLoS One 3 (12) (2008) e4018.

[4] A.S. Klyubin, D. Polani, C.L. Nehaniv, Empowerment: a universal agent-centric measure of control, in: The 2005 IEEE Congress on Evolutionary Computation, 2005., 1, IEEE, 2005, pp. 128–135.

[5] H. Touchette, S. Lloyd, Information-theoretic approach to the study of control systems, Physica A 331 (1–2) (2004) 140–172.

[6] M. Karl, M. Soelch, P. Becker-Ehmck, D. Benbouzid, P. van der Smagt, J. Bayer, Unsupervised real-time control through variational empowerment, arXiv preprint arXiv:1710.05101 (2017).

[7] A.H. Qureshi, M.C. Yip, Adversarial imitation via variational inverse reinforcement learning, arXiv preprint arXiv:1809.06404 (2018).

[8] K. Gregor, D.J. Rezende, D. Wierstra, Variational intrinsic control, arXiv preprint arXiv:1611.07507 (2016).

[9] R. Blahut, Computation of channel capacity and rate-distortion functions, IEEE Trans. Inf. Theory 18 (4) (1972) 460–473.

[10] S. Mohamed, D.J. Rezende, Variational information maximisation for intrinsically motivated reinforcement learning, in: Advances in Neural Information Processing Systems, 2015, pp. 2125–2133.

[11] D.B.F. Agakov, The im algorithm: a variational approach to information maximization, Adv. Neural Inf. Process. Syst. 16 (2004) 201.

[12] T. Hayakawa, T. Kaneko, T. Aoyagi, A biologically plausible learning rule for the infomax on recurrent neural networks, Front. Comput. Neurosci. 8 (2014) 143.

[13] T. Toyoizumi, J.-P. Pfister, K. Aihara, W. Gerstner, Generalized bienenstock–cooper–munro rule for spiking neurons that maximizes information transmission, Proc. Natl. Acad. Sci. 102 (14) (2005) 5239–5244.

[14] C. Salge, C. Glackin, D. Polani, Empowerment–an Introduction, in: Guided Self-Organization: Inception, Springer, 2014, pp. 67–114.

[15] T. Jung, D. Polani, P. Stone, Empowerment for continuous agent?environment systems, Adapt. Behav. 19 (1) (2011) 16–39.

[16] C. Salge, C. Glackin, D. Polani, Approximation of empowerment in the continuous domain, Adv. Complex Syst. 16 (02n03) (2013) 1250079.

[17] I. Belghazi, S. Rajeswar, A. Baratin, R.D. Hjelm, A. Courville, Mine: mutual information neural estimation, arXiv preprint arXiv:1801.04062 (2018).

[18] T. Raiko, M. Tornio, Variational bayesian learning of nonlinear hidden state-space models for model predictive control, Neurocomputing 72 (16–18) (2009) 3704–3712.

[19] R. Ranganath, S. Gerrish, D. Blei, Black box variational inference, in: Artificial Intelligence and Statistics, 2014, pp. 814–822.

[20] D. Ritchie, P. Horsfall, N.D. Goodman, Deep amortized inference for probabilistic programs, arXiv preprint arXiv:1610.05735 (2016).

[21] R. Shu, H.H. Bui, S. Zhao, M.J. Kochenderfer, S. Ermon, Amortized inference regularization, arXiv preprint arXiv:1805.08913 (2018).

[22] D.P. Kingma, M. Welling, Auto-encoding variational bayes, arXiv preprint arXiv:1312.6114 (2013).

[23] D.J. Rezende, S. Mohamed, D. Wierstra, Stochastic backpropagation and approximate inference in deep generative models, arXiv preprint arXiv:1401.4082 (2014).

[24] H.R. Wilson, J.D. Cowan, Excitatory and inhibitory interactions in localized populations of model neurons, Biophys. J. 12 (1) (1972) 1–24.

[25] H.R. Wilson, J.D. Cowan, A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue, Kybernetik 13 (2) (1973) 55–80.

[26] P. Dayan, L. Abbott, et al., Theoretical neuroscience: computational and mathematical modeling of neural systems, J. Cogn. Neurosci. 15 (1) (2003) 154–155.

[27] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, et al., Tensorflow: a system for large-scale machine learning., in: OSDI, 16, 2016, pp. 265–283.

[28] J.V. Dillon, I. Langmore, D. Tran, E. Brevdo, S. Vasudevan, D. Moore, B. Patton, A. Alemi, M. Hoffman, R.A. Saurous, Tensorflow distributions, 2017. arXiv preprint arXiv: 1711.10604

Elham Ghazizadeh joined the Ph.D. program of Electrical and Systems Engineering at Washington University in St. Louis in 2016. She is currently a full-time research assistant working with Dr. ShiNung Ching. She completed her M.Sc. and B.Sc. degrees in Electrical and Computer Engineering from Shahid Bahonar University of Kerman, Iran in 2016 and 2014, respectively. Her main research interests are at the intersection of computational neuroscience and machine learning.

ShiNung Ching is an Associate Professor in the Department of Electrical and Systems Engineering at Washington University in St. Louis (St. Louis, USA). Dr. Ching completed his B.Eng (Hons.) and M.A.Sc degrees in Electrical and Computer Engineering from McGill University, Canada and the University of Toronto, Canada. He earned his Ph.D. in Electrical Engineering from the University of Michigan in 2009. His research interests are at the intersection of control theory and systems neuroscience, particularly in using systems and control theoretic concepts to study the link between dynamics and function in neuronal networks. Dr. Ching has received the CAREER Award from the US National Science Foundation, the Young Investigator Program award from the US AFOSR, a Career Award at the Scientific Interface from the Burroughs-Welcome Fund.