Stochastic optimal control and sensori-motor integration

Hilbert J. Kappen
b.kappen@science.ru.nl

Vicenç Gómez
v.gomez@science.ru.nl

Donders Institute for Brain Cognition and Behaviour
Radboud University Nijmegen
6525 EZ Nijmegen, The Netherlands

Abstract

The paper discusses the problem of modelling intelligent behaviour using stochastic optimal control theory. The stochastic control solution requires state feed-back which requires vast computational resources both in terms of memory and computation. We argue that an efficient approach to this problem requires an integration of sensory and motor computation. We propose the path integral control framework as a natural theory for sensori-motor integration using a Bayesian framework.

1 Introduction

The over-aching goal of both artificial intelligence and cognitive (neuro-)science is to build models that provide an integrated understanding of how sensory stimuli are processed to yield motor actions. So far, we have a relatively good understanding of peripheral neural processing, both on the sensory and on the motor side, but the integration of these two modalities is much less understood. The progress is made difficult because the computational principles that underly sensori-motor integration are obscured by the neural hardware that implements them. As a result, bottom-up approaches that reveal 'what neurons are doing' need to be complemented by top-down, functional, approaches that guide our understanding on what principles may be used and how they are implemented.

On the sensory side, much progress has been made using ideas from statistics. Features in natural images that are 'statistically optimal' are in close agreement with the features that are computed by neurons in the brain of animals. On the motor side it is much less understood what principles guide our actions. For peripheral motor tasks, such as arm movement, or locomotion, or eye movement, there exists a basic understanding in terms of deterministic control theory and it is assumed that noise has a limited effect on the optimal solution.

Control theory is a natural candidate to also describe the complete behaviour of an animal or intelligent system, from now on referred to as the agent. Such a theory would be able to compute the behaviour of an agent in a complex, partially unknown, environment which contains other agents of various types with their own objectives. However, there are fundamental problems that have prevented the successful development of a theory of this type sofar:

- Due to limited sensing capability, the agent is to a large extend ignorant about the world around him, now and in the future. This setting is often referred to as *partial observability*. Partial observability may have a profound effect on what the agent should do. When the agent has full knowledge of a food location and the course of action of its predators, it can compute an optimal trajectory that leads to the food and avoids the predators. But this

course of actions may be disastrous when knowledge is uncertain. Partial observability can be encoded in terms of beliefs, which are distributions over the quantity of interest. However, the computation of controls in terms of beliefs is very costly.

- Furthermore, due to the stochastic nature of the problem, selecting the optimal action requires current state information. This *policy* should either be computed in real-time for any state that is visited or should be stored, requiring either massive computation or massive memory.

Thus, there is both a *representation problem* and a *computation problem*: representing the environment, the beliefs, and entire policies requires vast memory; computing the control for any given instance is intractable in terms of number of operations. These problems are absent for deterministic control problems and are tractable for Gaussian models and explain why stochastic optimal control theory has failed to provide a successful computational framework to model intelligent behaviour.

In our research, we aim to address these problems using the path integral and KL control formalism which connects the fields stochastic optimal control theory with statistical inference and statistical physics[1, 2, 3]. The optimal control can be computed using the efficient approximate inference methods that have been developed in the machine learning community. The path integral control methods have been applied with great success in robotics by the group of Stefan Schaal (see for instance [4]) showing their superiority to RL and adaptive control methods and in biological systems [5, 6]. In this paper we review path integral control theory. We then give an example for coordination of agents. Finally we discuss how this approach can be used for sensori-motor integration.

2 Path integral control theory

Here we summarise a simplified setting of path integral control theory. Consider the stochastic control problem

$$dx = f(x,t)dt + g(x,t)(udt + d\xi) \qquad C = \left\langle \phi(x_T) + \int dt V(x_t,t) + \frac{1}{2}u^T R u \right\rangle \quad (1)$$

with x the state and u the control and where I have suppressed all component notation. If one assumes that there exists a constant λ such that the matrices R and the noise covariance matrix ν satisfy $\lambda \mathbf{1} = R\nu$, it can be shown that the optimal cost-to-go is given by the Feyman-Kac formula

$$J(x,t) = -\lambda \log \int d\tau q(\tau|x,t) \exp\left(-S(\tau)/\lambda\right) \qquad S = \phi(x_T) + \int_t^T dt V(x_t,t) \quad (2)$$

where τ denotes a trajectory starting at x,t and $q(\tau|x,t)$ the distribution over trajectories under the *uncontrolled stochastic dynamics* Eq. 1 with $u = 0$ and S the *action*. There is a corresponding Gibbs distribution over optimally controlled trajectories

$$p(\tau|x,t) = q(\tau|x,t) \exp(J(x,t)/\lambda - S(\tau|x,t)/\lambda) \quad (3)$$

The optimal control can be expressed as an expectation value with respect to $p(\tau|x,t)$:

$$u_j(x,t)dt = \langle d\xi_j \rangle \quad (4)$$

The path integral control theory can be obtained as a particular instance of a larger class of problems that minimizes the Kullback-Leibler (KL) divergence between distributions over trajectories. This is illustrated in the simplest case of a discrete state space and discrete time.

Let q denote a Markov process on this state space. Consider a control problem to find a Markov process p that minimizes $C = KL(p||q) + \langle V \rangle_p$, with $KL(p||q)$ the KL divergence and $p(\tau), q(\tau)$ distributions over trajectories τ according to the Markov processes p and q, respectively and τ a trajectory of length T starting at a given initial state x_0. The optimal control solution is given by the discrete time version of Eq. 3, with optimal cost $C(x_0) = -\log \psi(x_0)$. The solution p is a Markov process with

$$p_t(x_t|x_{t-1}) = q(x_t|x_{t-1}) \exp(-V(x_t)) \frac{\beta_t(x_t)}{\beta_{t-1}(x_{t-1})} \quad (5)$$

where the functions $\beta_t(x)$ are found by message passing.

It can be shown that when q and R do not explicitly depend on time in the limit of $T \to \infty$ the KL control problem becomes an extremal eigenvalue problem [2].

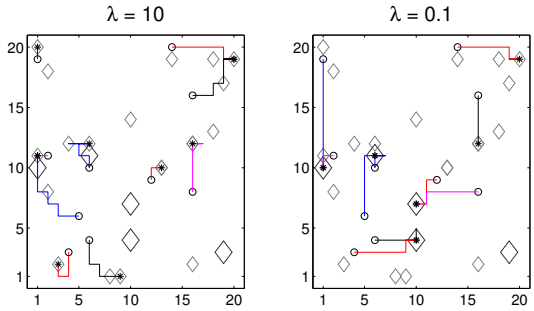

Figure 1: Approximate inference KL-stag-hunt using BP for $M = 10$ hunters in a large grid. (**Left**) Risk dominant control is obtained for $\lambda = 10$, where all hunters go for a hare. (**Right**) Payoff dominant control is obtained for $\lambda = 0.1$. In this case, all hunters cooperate to capture the stags except the ones on the upper-right corner, who are too far away from the stag to reach it in $T = 10$ steps. Their optimal choice is to go for a hare. $N = 400$, $R_s = -10$, $H = 2M$ and $R_h = -2$.

3 Multi Agent cooperative game (KL-stag-hunt)

In this section we consider a variant of the stag hunt game, a prototype game of social conflict between personal risk and mutual benefit [7]. The original two-player stag hunt game consists of two hunters that can either hunt a hare by themselves giving a small reward, or cooperate to hunt a stag and getting a bigger reward, see table 1.

Both stag hunting (*payoff* equilibrium, top-left) and hare hunting (*risk-dominant* equilibrium, bottom-right) are *Nash equilibria*,

	Stag	Hare
Stag	$\mathbf{3}, \mathbf{3}$	$0, 1$
Hare	$1, 0$	$\mathbf{1}, \mathbf{1}$

Table 1: Payoff matrix for the stag-hung game: if both go for the stag, they both get a reward of 3. If one hunter goes for the stag and the other for the hare, they get a reward of 0 and 1 respectively.

We define the KL-stag-hunt game as a multi-agent version of the original stag hunt game where M agents live in a 2d grid of N locations and can move to adjacent locations on the grid. The grid also contains hares and stags at certain fixed locations. The game is played for a finite time T and at each time-step all the agents move.

We formulate the problem as a KL control problem. The uncontrolled dynamics factorizes among the agents. It allows an agent to stay on the current position or move to an adjacent position (if possible) with equal probability, thus performing a random walk on the grid. The state dependent cost $V(x)/\lambda$ defines the profit when two agents are at the same time at the location of of a stag, or individual agents are at a hare location.

Computing the exact solution using this procedure becomes infeasible even for small number of agents, since the joint state space scales as N^M. The belief propagation (BP) algorithm is an alternative approximate algorithm that we can run on an extended factor graph and has polynomial time and space complexity [3].

The result is illustrated in Figure 1, where an example for $\lambda = 10$ and $\lambda = 0.1$ are shown. For high λ (left plot), each hunter catches one of the hares. In this case, the cost function is dominated by KL term. For small enough values of λ (right plot), the V/λ term dominates and both hunters cooperate to catch the stag. Thus λ can be seen as a parameter that controls whether the optimal strategy is risk dominant or payoff dominant.

This example shows how KL control can be used to model a multi-agent cooperative game. It explains the emergence of cooperation in terms of an effective temperature parameter λ. Approximate inference methods like BP provide an efficient and good approximation for large systems where exact inference is not feasible.

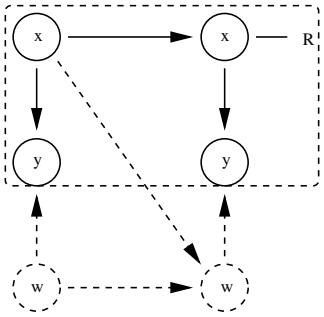

Figure 2: Sensori-motor model of world and brain coupled through observations. x denotes the activity of neural network/hidden graphical model at a given time. y are sensory neurons (retina, cochlea) and w represents the state of the outside world. The brain affects the world ($x \to w$) and knows about the world through observations y only.

Discussion: Sensor-motor integration

Using the path integral control theory, the control computation and the representation problem can be integrated as a single inference computation in a Bayesian network providing an integrated approach to sensori-motor control for both natural and artificial systems as schematically depicted in fig. 2:

- (Hidden) graphical models ($x \to x$) that yield efficient 'coarse grained' representation for the stochastic control problem in terms of suitable features of the problem ($x \to y$).
- Efficient inference algorithms to compute the optimal control in these models (as for instance illustrated above for the stag hunt game).
- For biological systems, the rewards are functions of the brain state (pleasure, pain) and not of the outside world, as is usually assumed for artificial systems: eating food sends a pleasure signal to the brain, activating a particular brain state. The control problem is to get the brain into the food state, which requires actions in the outside world.

The optimal hidden dynamics has the combined role to represent the environment dynamically (sensory problem) as well as to implement the optimal control (motor problem, see Eq. 5) by adapting the synaptic weights/hidden representation.

References

[1] H.J. Kappen. A linear theory for control of non-linear stochastic systems. *Physical Review Letters*, 95:200201, 2005.

[2] E. Todorov. Linearly-solvable markov decision problems. In B. Schölkopf, J. Platt, and T. Hoffman, editors, *Advances in Neural Information Processing Systems 19*, pages 1369–1376. MIT Press, Cambridge, MA, 2007.

[3] H.J. Kappen, V. Gómez, and M. Opper. Optimal control as a graphical model inference problem. *Machine Learning Journal*, 2012. DOI 10.1007/s10994-012-5278-7.

[4] E. Theodorou, J. Buchli, and S. Schaal. Reinforcement learning of motor skills in high dimensions: a path integral approach. In *IEEE International Conference on Robotics and automation (ICRA)*, pages 2397–2403, 2010.

[5] E. Todorov. Efficient computation of optimal actions. *Proceedings of the National Academy of Sciences*, 106:11478–11483, 2009.

[6] W. Yoshida, R. J. Dolan, and K.J. Friston. Game theory of mind. *PLOS Computational Biology*, 4:e1000254, 2008.

[7] Brian Skyrms, editor. *The Stag Hunt and Evolution of Social Structure*. Cambridge University Press, Cambridge, MA, USA, 2004.