Multiagent Control as a Graphical Model Inference Problem

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

Manfred Opper OPPERM@CS.TU-BERLIN.DE
Department of Computer Science
D-10587 Berlin, TU Berlin, Germany

KL-control problems are a certain class of non-linear stochastic optimal control problems for which the optimal control cost C is a Kullback-Leibler divergence between the optimal control law p and the uncontrolled process q plus a state dependent expected cost of future states $\langle R\rangle_p$ (Kappen et al., 2012; Todorov, 2007).

$$C = \mathrm{KL}(p\|q) + \langle R\rangle_p.$$

In this work, we show that this class of problems corresponds to a probabilistic inference problem defined on a factor graph, where variable nodes denote the states of the system at different times and factor nodes encode either the uncontrolled process or the state costs. The optimal control is given by a marginal distribution that can be computed using standard methods such as the junction tree or belief propagation (BP).

We consider the following game defined on a grid where M agents (hunters) can move to adjacent locations for T time steps. The grid also contains hares and stags at fixed locations. Each hunter can choose between hunting a hare on his own, resulting in a small reward R_h, or hunting a stag, resulting in a larger reward $R_s \gg R_h$, but requiring cooperation of two hunters.

To define the factor graph associated to this problem, let x_i^t (variables) denote the position of hunter i at time t on the grid. Also, let s_j and h_k denote the positions of the jth stag and the kth hare respectively. The state dependent reward factor can be written as: $\psi_R(x^t) = \exp(-1/\lambda R(x^t))$,

$$R(x^t) = R_h \sum_{k=1}^{H}\sum_{i=1}^{M} \delta_{x_i^t,h_k} + R_s \sum_{j=1}^{S} \mathcal{I}\{(\sum_{i=1}^{M} x_i^t = s_j) > 1\}.$$

The uncontrolled dynamics factorizes among the agents $\psi_q(x^t|x^{t-1}) = \prod_i \psi_q(x_i^t|x_i^{t-1})$ and is defined as a random walk, allowing an agent to stay or to move to an adjacent position with equal probability.

We "clamp" x_i^0 to a given initial configuration and estimate the marginals (optimal controls) $p(x^{1:T}|x^0)$.

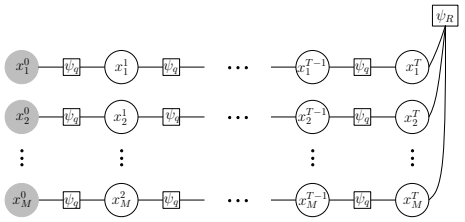

Figure 1. Factor graph representation for the KL-stag-hunt problem. Approximate optimal control can be obtained through the BP factor beliefs between two time slices.

Computing them exactly is intractable, since the state space scales as N^M. BP is an alternative approximate algorithm with polynomial complexity.

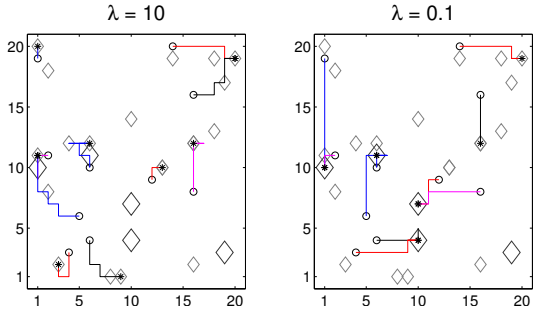

Figure 2. Examples of solutions using BP for 10 agents for different values of λ. (Left) **Risk** dominant optimal control: all hunters go for a hare. (Right) **Payoff** dominant optimal control: hunters cooperate to capture the stags. Small and big diamonds denote hares and stags respectively. Circles denote initial positions.

References

Kappen, H. J., Gómez, V., & Opper, M. (2012). Optimal control as a graphical model inference problem. *Machine Learning*, *87*, 159–182.

Todorov, E. (2007). Linearly-solvable Markov decision problems. In *NIPS 19*, 1369–1376. MIT Press.