Pretty darn good control: when are approximate solutions better than approximate models*

Abstract

The text of your abstract. 150 – 250 words.

Introduction

Figures brainstorm

Figure 1: 1-D and 3-D model conceptual figure. something about the objective / decision Figure 2: Stability / multistability in the 1D and 3D models. state space + time views

Figure 3: The 1-D optimal management solution. 'constant escapement' intuition etc

Results figures: - timeseries example of management under the RL policy. probably compare to managing under 1-D solution / rule-of-thumb methods - state-space view of management doughnut - visualization / encapsulation of the policy heatmaps, slices, policy vs position along ellipse - reward plot over time (comparing methods)

Fisheries: approximate models and approximate solutions

Reinforcement learning

Reinforcement learning (RL) is a way of approaching *control problems* through machine learning. An RL application can be conceptually separated into two parts: an *agent*, and an *environment*. The *environment* is commonly a computer simulation, although it sometimes can be a real world system. The *agent*, on the other hand, is a computer program which interacts with the environment.

*This material is based upon work supported by the National Science Foundation under Grant No. DBI-1942280.

That is, the agent may act on the environment and change its state. This, moreover, gives the agent a *reward* which encodes the control goal. The main part of an RL algorithm is then to progressively improve the agent's *policy*, in order to maximize the cumulative reward received. This is done by aggregating experience and learning from it. See Fig. **??**.

In our case, the environment will be a computational model of the population dynamics of a fishery, with the environment state being a vector of all the fish populations, $S = (V_1, V_2, H)$. The system evolves naturally according to (**??**). At each time step, the agent harvests one of the populations, say A. Specifically, it chooses a quota q: the fraction of A that will be fished. This changes the state as

$$(V_1,\ V_2,\ H) \mapsto ((1-q)V_1,\ V_2,\ H)$$

and secures a reward of qV_1 to the agent. Afterwards, the environment undergoes its natural dynamics.

Mathematical framework for RL

Mathematically, RL may be formulated using a discrete time *partially observable Markov decision process (POMDP)*. This formalization is rather flexible and allows one, e.g., to account for situations where the agent may not fully observe the environment state, or where the only observations available to the agent are certain functions of the underlying state. For the sake of clarity, we will present here only the class of POMDPs which are relevant to our work: fully observed POMDPs with trivial emission function (FMDPs for short). An FMDP may be defined by the following data:

- $\mathcal{S}$: *state space*, the set of states of the environment,
- $\mathcal{A}$: *action space*, the set of actions which the agent may choose from,
- $T(s_{t+1}|s_t, a_t)$: *transition operator*, a conditional distribution which describes the dynamics of the system,
- $r(s_t, a_t)$: *reward function*, the reward obtained after performing action $a_t \in \mathcal{A}$ in state s_t,
- $d(s_0)$: *initial state distribution*, the initial state of the environment is sampled from this distribution,
- $\gamma \in [0, 1]$: *discount factor*.

Acknowledgements

The title of this piece references a mathematical biology workshop at NIMBioS organized by Paul Armsworth, Alan Hastings, Megan Donahue, and Carl Towes in 2011 that first posed the question addressed here.

References