Embedding a State Space Model into a Markov Decision Process

Lars Relund Nielsen, Erik Jørgensen and Søren Højsgaard
Department of Genetics and Biotechnology, University of Aarhus, Denmark

Introduction

- Markov decision processes (MDPs) with discrete state spaces model sequential decision making over time.
- An application could be finding the optimal replacement decision for an animal. States in the MDP represent levels of traits of the animal and transition probabilities are based on biological models estimated from data.
- State space models (SSMs) usually deal with a "continuous state vector" and continuous responses. Therefore it is hard combining an MDP and an SSM.
- We have a method for embedding an SSM into an MDP such that predictions in the MDP are based on Bayesian updating, so hang on and see what happens!

Markov decision processes (MDPs)

- Discrete time-instances, states and decisions.
- A stage of an MDP is illustrated using a directed hypergraph in Fig. 1. Decisions *terminate* and *continue* are illustrated using hyperarcs.
- Use dynamic programming to find optimal decisions.

Linear normal state space models (SSMs)

- Models of phenomena evolving in time e.g. blood pressure and milk yield.
- Latent process evolves as a first order Markov process.

$$\theta_t = G\theta_{t-1} + \omega_t, \quad \omega_t \sim N(0, W)$$

- Y_t are observations which we model as a function depending on θ_t.

$$Y_t = F'\theta_t + \nu_t, \quad \nu_t \sim N(0, V)$$

Embedding the SSM

- Kalman filter/Bayesian updating implies the distribution of the posterior at time t can be calculated:

$$(\theta_t \mid Y_1, \dots, Y_t) \sim N(m_t, C_t)$$

  which can be found using the observation and the prior of the latent variable. Hence all previous observations do not have to be stored as state variables in the MDP, the observations can be represented by storing m_t and C_t instead.
- C_t does not depend on data and hence can be excluded from the states in the MDP.

Embedding the SSM (continued)

- Let $\tilde{m}_1, \dots, \tilde{m}_q$ denote the discretization of m_t then a stage can be illustrated as

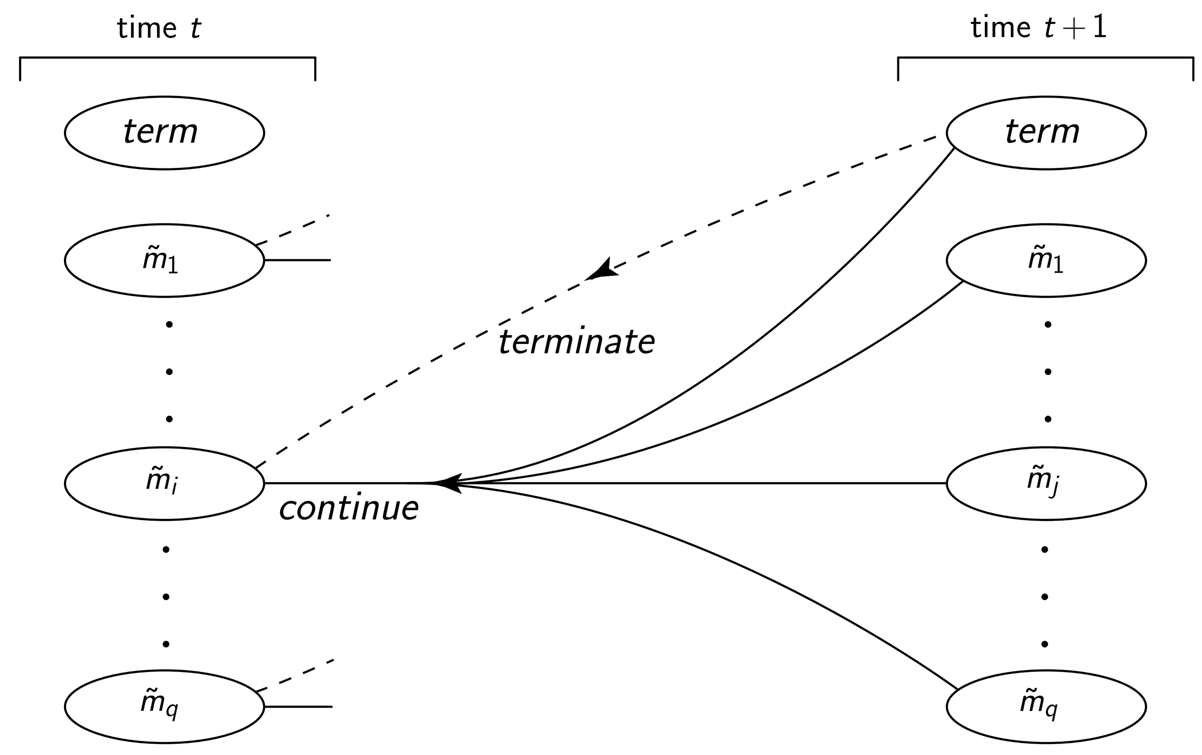


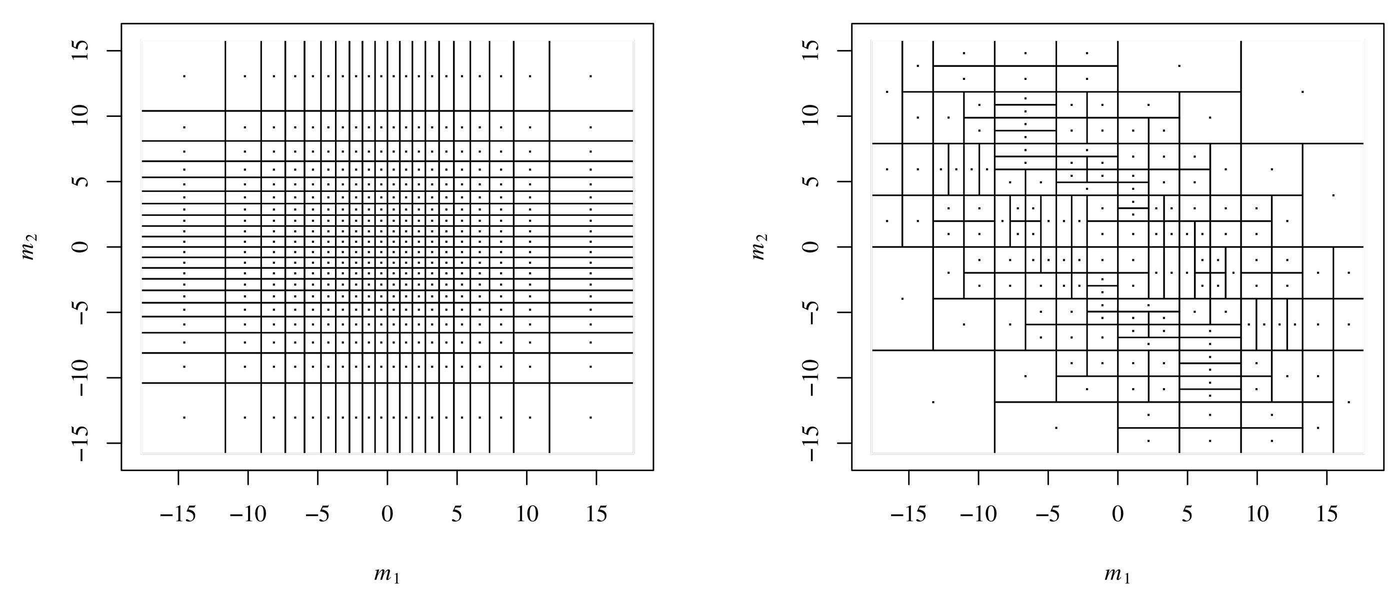
Fig. 1: MDP with the SSM embedded. The state *term* indicates that the data for the SSM has terminated either using a decision or involuntary.

- Transition probabilities $P(m_{t+1} \mid m_t)$ can be calculated on the basis of $(m_{t+1} \mid m_t) \sim N(\mu, \Sigma)$ with:

$$\mu = Gm_t, \quad \Sigma = GC_tG' + W - C_{t+1}$$

Discretizing the SSM

- Discretize m_t based on the relative entropy (RE) as criterion.
- Goal: Few stages but good approximation (low RE).
- Univariate: Discretize each variable in m_t.
- Multivariate: Discretize the vector m_t.
- Multivariate in general has fewer states for the same quality of the approximation (see Fig. 2).



(a) Univariate. (b) Multivariate.

Fig. 2: Partition of m_t (2-dim) using univariate and multivariate discretization. Multivariate discretization reduces the number of states by 58%.

Conclusions

The method can be applied to MDPs which use information from online biosensors, e.g. the method is currently used to calculate optimal replacement strategies for dairy cows based on daily yield measurements.