Searching for the unknown dynamics in continuous partially observable MDPs

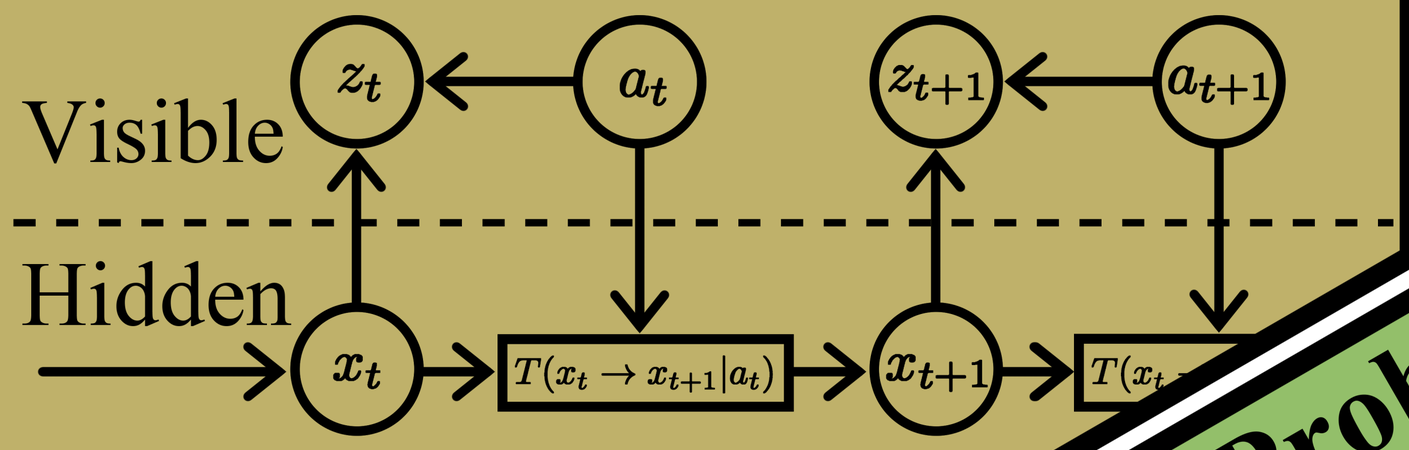
Erik Karlsson Nordling, Chalmers University of Technology
Mathematical Sciences
Supervisor: Prof. Marina Axelson-Fisk (Chalmers)

CHALMERS UNIVERSITY OF TECHNOLOGY

Problem

Find MLE

The examined problem is: Find the transition function that maximizes the posterior observation (z_t) probability.



Simplified problem

$$\min_\rho \ f(\rho)$$
$$\text{s.t.} \quad 1 = \int_{x\in\mathcal{S}} \rho(x|y)dx \ \forall \ y \in \mathcal{S}$$
$$\rho(x|y) \geq 0 \ \forall \ x,y \in \mathcal{S}$$

Finite-dimensional Master Problem

Solve for the **linear basis approximation:**

$$\hat{\rho}(x|y) = \boldsymbol{v}^T\boldsymbol{\phi}(x,y)$$

Include penalty term for equality constraint.

$$z^*(\boldsymbol{\phi},\eta) = \min_{\boldsymbol{v}\geq\mathbf{0}} \ f(\boldsymbol{v}^T\boldsymbol{\phi}) + \tfrac{\eta}{2}||\mathbf{1} - \boldsymbol{v}^T\textstyle\int_{x\in\mathcal{S}} \boldsymbol{\phi}(x,\cdot)dx||_{L_2}$$

Is solved with 0-bounded Newton's algorithm.

$$\nabla_{\boldsymbol{v}} f(\hat{\rho}) = \boldsymbol{\phi}\cdot\frac{\delta f}{\delta\rho}\Big|_{\rho=\hat{\rho}}$$
$$H_{\boldsymbol{v}}(f)(\hat{\rho}) = \boldsymbol{\phi}\cdot\frac{\delta^2 f}{\delta\rho^2}\Big|_{\rho=\hat{\rho}}\cdot\boldsymbol{\phi}$$

Subspace Search Subproblem

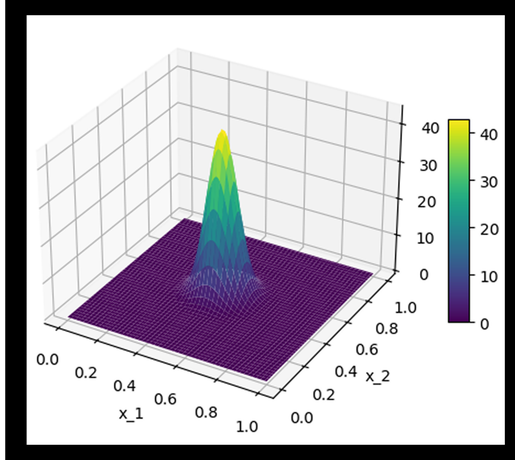
Assume that $\hat{\rho}^*$ solves the restricted master problem. We may evaluate an updated basis vector $\boldsymbol{\psi}$ by solving the functional 2nd-order approximation:

$$\min_{\boldsymbol{u}\geq\mathbf{0}} (\boldsymbol{u}^T\boldsymbol{\psi} - \hat{\rho}^*)\cdot\frac{\delta f}{\delta\rho}\Big|_{\rho=\hat{\rho}^*} + \frac{1}{2}(\boldsymbol{u}^T\boldsymbol{\psi} - \hat{\rho}^*)\cdot\frac{\delta^2 f}{\delta\rho^2}\Big|_{\rho=\hat{\rho}^*}(\boldsymbol{u}^T\boldsymbol{\psi} - \hat{\rho}^*)$$

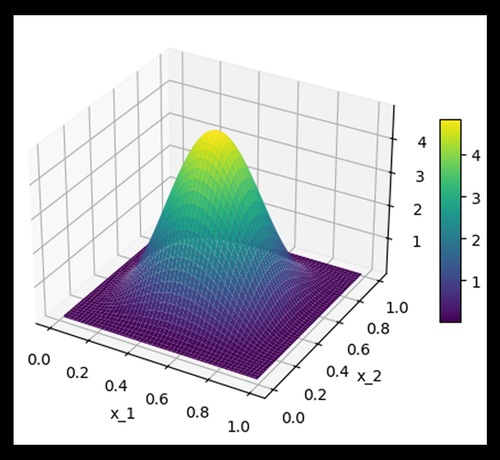
The problem is similar to the problem of approximating a target function with a nonnegative combination of the basis functions in $\boldsymbol{\psi}$. Different algorithms may be used to find a good $\boldsymbol{\psi}$. A **modified version of subspace pursuit** [1] was implemeted. Also let the penalty parameter increase each iteration.

Preliminary Results

Estimated state PDF



Subspace method



Baseline method

Algorithm has problems with dependency on prior state. Possible cause: penalty.
The representation found is sparse, and achieves higher objective value than the baseline method that uses more parameters.

[1] W. Dai and O. Milenkovic, "Subspace Pursuit for Compressive Sensing Signal Reconstruction," in IEEE Transactions on Information Theory, vol. 55, no. 5, pp. 2230-2249, May 2009, doi: 10.1109/TIT.2009.2016006

Method

Results