Recursive Method of Moments Identification of Hidden Markov Models using Convex Optimization

Robert Mattila[1], Vikram Krishnamurthy[2], Bo Wahlberg[1]

Introduction

- **Problem:** Estimate the parameters of an HMM from observations, assuming either known sensor or system dynamics. Expectation-maximization is only locally convergent and potentially slow.
- **Algorithm:** Recursively estimate the second order moments and simultaneously solve a convex/convexly relaxed optimization problem representing the decomposition of the moments into the HMM parameters.
- **Results:** Works well for identifying (potentially) time-varying system dynamics with known sensor. Works sometimes when identifying the sensor.

Problem Statement

Consider an HMM with transition matrix T, observation matrix O and stationary distribution π. Given a sequence of observations $\{y_k\}$, we consider estimation of

- T given O;
- O given T.

Method of Moments

The matrix $S_{ij} = \Pr[y_k = i, y_{k-1} = j]$ can be written

$$S = OT\operatorname{diag}(\pi)O^T \triangleq OAO^T$$

and can empirically be estimated from data using (recursive) stochastic approximation.

Identifying both T and O

In general non-convex, see e.g. [1]:

$$\begin{aligned} \min_{A,O} \quad & \|S - OAO^T\|_F^2 \\ s.t. \quad & \begin{cases} A \geq 0,\ \mathbb{1}^T A \mathbb{1} = 1, \\ O \geq 0,\ \mathbb{1}^T O = \mathbb{1}^T. \end{cases} \end{aligned}$$

The transition matrix can easily be recovered from A.

Known Sensor Dynamics

Reduces to a convex problem:

$$\begin{aligned} \min_{A} \quad & \|S - OAO^T\|_F^2 \\ s.t. \quad & A \geq 0,\ \mathbb{1}^T A \mathbb{1} = 1. \end{aligned}$$

The constraints can be removed by using $\sin^2 x + \cos^2 x \equiv 1$ to parametrize A, see [2]:

$$\min_{\alpha} \|S - OA(\alpha)O^T\|_F^2.$$

Known System Dynamics

Assuming A known (via T and π) the optimization problem is in general non-convex.

Convex Relaxation

Expand the cost function

$$\begin{aligned} \|S - OAO^T\|_F^2 &= \sum_{i,j}(S_{ij} - \sum_{l,m} A_{lm} \underbrace{O_{il}O_{mj}^T}_{=\lambda_{ilmj}})^2 \\ &= \sum_{i,j}(S_{ij} - \sum_{l,m} A_{lm}\lambda_{ilmj})^2. \end{aligned}$$

Formulated in λ, the problem is again convex:

$$\begin{aligned} \min_{\lambda} \quad & \sum_{i,j}(S_{ij} - \sum_{l,m} A_{lm}\lambda_{ilmj})^2 \\ s.t. \quad & \lambda_{ijkl} \geq 0, \\ & \lambda_{ijkl} = \lambda_{lkji}, \\ & \sum_{i,k} \lambda_{ijkl} = Y, \\ & W = \operatorname{vec}[O(\lambda)]\operatorname{vec}[O(\lambda)]^T \succeq 0, \\ & \operatorname{rank}(W) = 1, \\ & \sum_{j,k,l} \pi_j \lambda_{ijkl} = \sum_j S_{ij}, \end{aligned}$$

where Y is the number of observations and the rank-condition can be approximated using a nuclear norm minimization.

Problem: We find solutions in λ that are not solutions of the original problem (in O).

Solution (?): Alternate between optimizing in λ and O.

Conclusion

We simultaneously estimate S and solve the optimization problem recursively (using e.g. a gradient descent or primal-dual scheme). Works *well* for estimating the system dynamics, works *sometimes* for estimating the sensor dynamics.

References

[1] B. Vanluyten, J. C. Willems, and B. De Moor, “Structured nonnegative matrix factorization with applications to hidden Markov realization and clustering.” In *Linear Algebra and its Applications*, vol. 429, no. 7, pp. 1409-1424, 2008.

[2] R. Mattila, V. Krishnamurthy and B. Wahlberg. “Recursive Identification of Chain Dynamics in Hidden Markov Models Using Non-Negative Matrix Factorization”. In *54th IEEE Conference on Decision and Control*, Osaka, Japan, 2015.

[1] Department of Automatic Control and ACCESS Linnaeus Centre, School of Electrical Engineering, KTH Royal Institute of Technology, Stockholm, Sweden.

[2] Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, Canada.