Inverse Filtering in Hidden Markov Models

Robert Mattila, Cristian R. Rojas, Vikram Krishnamurthy[†] and Bo Wahlberg

KTH Royal Institute of Technology, Sweden and [†]Cornell Tech, Cornell University, USA

Summary

Problem:

Given posteriors $\{\pi_k\}_{k=1}^N$ from a *hidden Markov model* (HMM) filter with known transition matrix P.

Is it possible to reconstruct:

- the sequence of observations $\{y_k\}_{k=1}^N$?
- the observation matrix B?
- both?

Results:

Yes! It is possible to recover B and $\{y_k\}_{k=1}^N$ exactly in absence of noise. In presence of noise, estimates can be obtained via clustering.

Introduction

The stochastic filtering problem (given observations, compute the state posterior) is of paramount importance in many applications. In this work, we consider the corresponding *inverse problem*.

Motivation The underlying idea of inverse filtering problems ("*inform me about your state estimate and I will know your sensor characteristics, including your measurements*") has potential applications in:

- autonomous calibration of sensors, electronic warfare, cyberphysical security
  - *How can one determine how accurate an adversary's sensors are?*
- fault detection
  - *If multiple data batches are available, then change detection can be performed on the sequence of reconstructed observation likelihoods.*
- modeling of experts
  - *If the posterior distribution is estimated by querying a number of experts, then they can be bypassed in the future.*
- ...

Preliminaries

- *State* at time k: $x_k \in \{1, \dots, X\}$
- *Observation* at time k: $y_k \in \{1, \dots, Y\}$
- *Transition matrix*: $[P]_{ij} = \Pr[x_{k+1} = j | x_k = i]$
- *Observation matrix*: $[B]_{ij} = \Pr[y_k = j | x_k = i]$
- *Posterior distribution*:

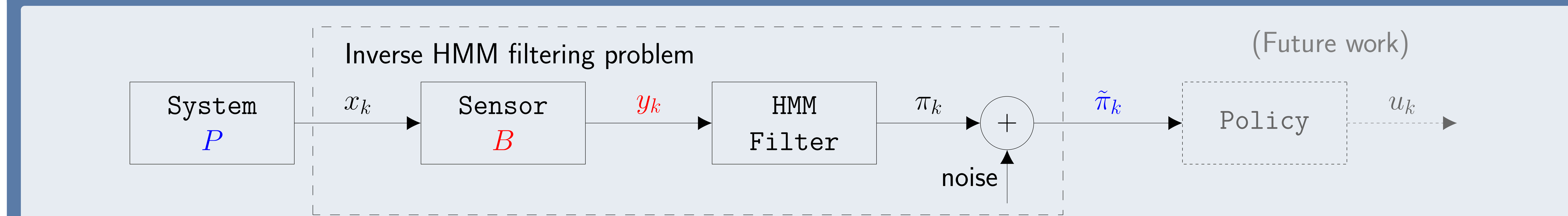
$$[\pi_k]_i = \Pr[x_k = i | y_1, \dots, y_k].$$

The HMM filter computes the posterior of the latent state, given observations from the system via the recursive updates:

$$\pi_k = \frac{\operatorname{diag}(b_{y_k}) P^T \pi_{k-1}}{\mathbb{1}^T \operatorname{diag}(b_{y_k}) P^T \pi_{k-1}}, \qquad (1)$$

where $B = [b_1 \dots b_Y]$.

Inverse HMM Filtering Problem



Problem Formulations

Noise-free: Consider the given data $\mathcal{D} = \{P, \{\pi_k\}_{k=0}^N\}$, where the posteriors have been generated by an HMM-filter sensor. Reconstruct the observations $\{y_k\}_{k=1}^N$ and the observation likelihood matrix B.

Noisy: Consider the given data $\mathcal{D} = \{P, \{\tilde{\pi}_k\}_{k=0}^N\}$, where $\tilde{\pi}_k$ is a noise-corrupted measurement of π_k (due to, e.g., quantization, measurement or modelling errors). Estimate the observations $\{y_k\}_{k=1}^N$ and the observation likelihood matrix B.

Naive Solution

Rewrite:

$$(1) \iff b_{y_k}^T P^T \pi_{k-1} \pi_k = \operatorname{diag}(b_{y_k}) P^T \pi_{k-1}. \qquad (2)$$

Formulate as a feasibility problem:

$$\begin{aligned} \min_{\{y_k\}_{k=1}^N, \{b_i\}_{i=1}^Y} \quad & \sum_{k=1}^N \| b_{y_k}^T P^T \pi_{k-1} \pi_k - \operatorname{diag}(b_{y_k}) P^T \pi_{k-1} \|_\infty \\ \text{s.t.} \quad & y_k \in \{1, \dots, Y\}, \\ & b_i \geq 0, \ [b_1 \dots b_Y]\mathbb{1} = \mathbb{1}. \end{aligned}$$

Note: This is a **computationally expensive** *mixed-integer linear program* (MILP).

Efficient Solution

Step 1: *Two useful lemmas*

Lemma 1: The HMM-filter update equation (2) can be equivalently written

$$\left(\pi_k (P^T \pi_{k-1})^T - \operatorname{diag}(P^T \pi_{k-1})\right) b_{y_k} = 0. \qquad (3)$$

Lemma 2: If $P, B > 0$, then the nullspace of the matrix

$$\pi_k (P^T \pi_{k-1})^T - \operatorname{diag}(P^T \pi_{k-1}), \qquad k > 1,$$

is of dimension one.

Step 2: *Recover the observation matrix B*

i) For every time k, compute a basis for the nullspace of (3) $\implies$ the direction of one column of B.

ii) There is a finite number of columns in B: stop when you have all Y unique directions.

- If $B \geq \beta > 0$, then the expected number of samples is less than $\beta^{-1}(1 + \frac{1}{2} + \dots + \frac{1}{Y})$.

iii) Normalize using the sum-to-one property.

Step 3: *Recover the observations y_k*

For every k, check which column of the recovered B-matrix that the nullspace of (3) is parallel to.

Noisy Case

When the posteriors are corrupted by noise (due to, e.g., quantization, measurement or modeling uncertainties) reformulate **Step 2**.ii) and **3** as a clustering problem.

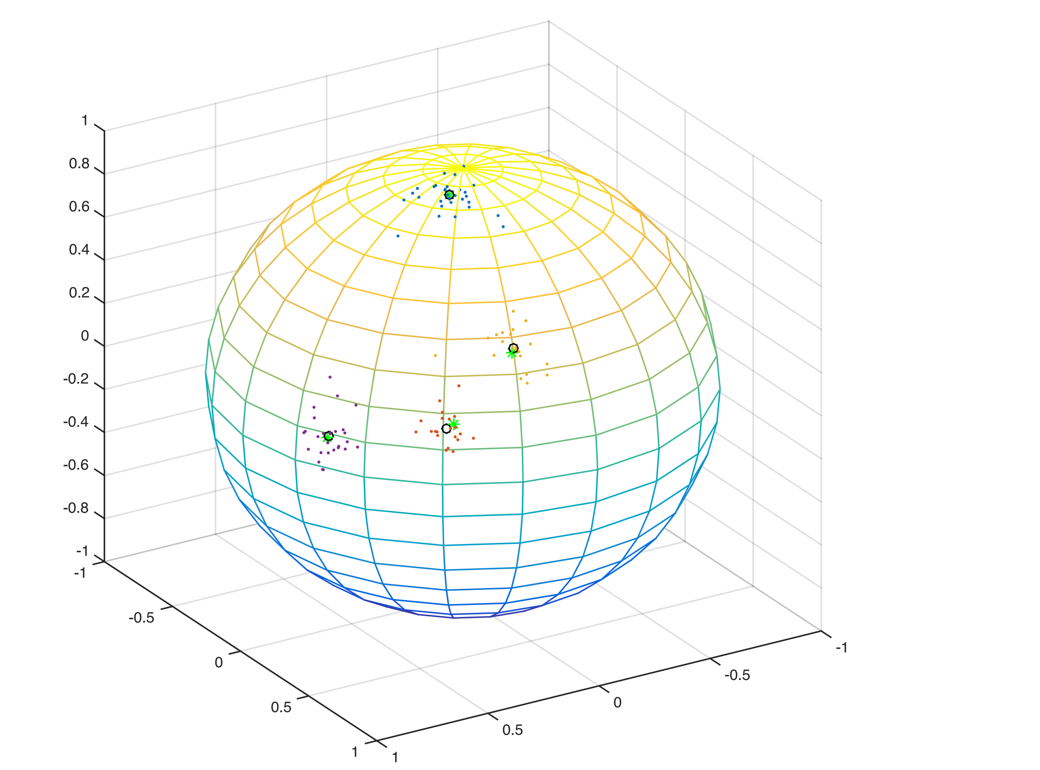
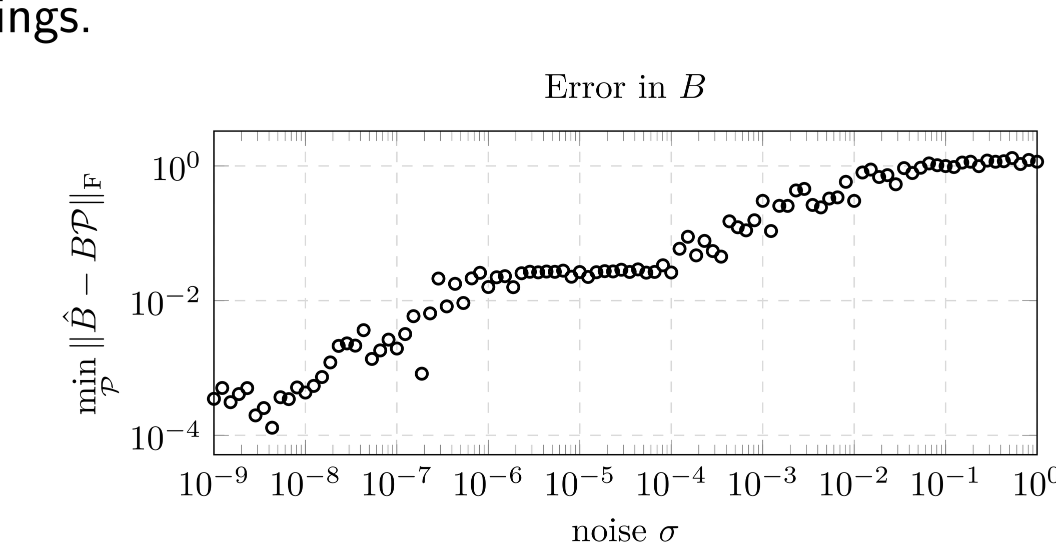
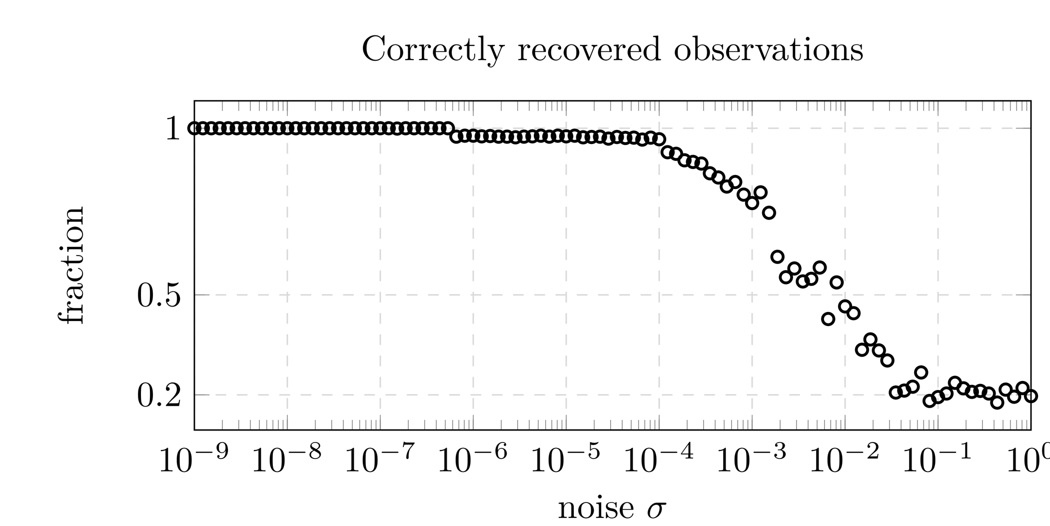


Figure 1: Every nullspace is a noisy estimate of one column of B. *Legend:* $\cdot$ – (perturbed) nullspace, $*$ – true column of B, $\circ$ – centroid (resulting estimate of a column of B).

Sleep Staging

Evaluated on real-world data from a system used for automatic sleep segmentation based on EEG readings.





Open Problems

- Uncertain, or even unknown, system dynamics P
- Only actions based on the filtered distribution can be observed (POMDP)

Contact Information

✉ : rmattila@kth.se   www: rmattila.github.io