Hidden Markov Models and Gibbs formalism

E. Verbitskiy

Noise constantly affects all sorts of information: from corruption of data transmitted over communication channels to random mutations in DNA sequences. Effects of noise on information sources need to be properly understood. When can the information be recovered? When is it irretrievably lost? When is the transmission of partial information still of use? These fundamental questions are of great relevance in applied fields like Information Theory and Bioinformatics. In the past decades, these fields have seen enormous progress, which stems from the application of probabilistic modeling to information sources. Probabilistic model often used is to model the effects of noise in various applications such as speech recognition, language processing, bioinformatics, image analysis, is the so-called Hidden Markov Model.

Hidden Markov Models: Let $\{X_k\}$ be a stationary Markov chain with a finite state space S. The process $\{Y_k\}$ is called a Hidden Markov process if

- given $\{X_k\}$, the random variables $\{Y_k\}$ are conditionally independent;
- the distribution of Y_n depends on $\{X_k\}$ only through X_n.

It is easy to show that these conditions imply that there exists a function h such that

$$Y_n = f(X_n, \epsilon_n),$$

where $\{\epsilon_n\}$ is a sequence of independent identically distributed random variables.

Key object for treatment of HMM's are the so-called smoothing probabilities [1–3]:

$$\mathbb{P}(X_k = s | Y_m = y_m \dots Y_n = y_n), \quad m \le k \le n.$$

For example, these probabilities are basis for optimal **denoising** of HMM's, i.e., estimating $\hat{X}_k$ – the most probable input symbol, given the realization of HMM $\{Y_t\}_{t=1}^N$.

Under various conditions, the smoothing probabilities have very nice properties, namely, the **exponential forgetting**:

$$\Big| \mathbb{P}(X_k = s | Y_m = y_m \dots Y_n = y_n) - \mathbb{P}(X_k = s | Y_{m-v} = y_{m-v} \dots Y_{n+u} = y_{n+u}) \Big| \le C(\rho^v + \rho^u),$$

for $m \le k \le n$, $u, v \ge 0$, and some $\rho \in (0, 1)$. It turns out that this property is closely related to the fact that the Hidden Markov process $\{Y_n\}$ is **Gibbs**.

First objective of this Bachelor project is to review the known sufficient conditions for exponential forgetting, and put these results in the thermodynamic perspective. Furthermore, we will investigate the exponential forgetting for more general hidden processes.

Prospective student should be familiar with basic probability theory, statistics. There is a possibility to perform some numerical simulations as well, hence affinity with Matlab will be beneficial. Following the Master course of A. Opuku on Gibbs measures will be highly beneficial as well.

[1] Olivier Cappé, Eric Moulines, and Tobias Rydén, *Inference in hidden Markov models*, Springer, 2005.

[2] Jüri Lember, *On approximation of smoothing probabilities for hidden Markov models*, Statist. Probab. Lett. **81** (2011), no. 2, 310–316.

[3] T. Weissman, E. Ordentlich, G. Seroussi, S. Verdú, and M.J. Weinberger, *Universal discrete denoising: Known channel*, IEEE Transactions on Information Theory **51** (2005), no. 1, 5–28.