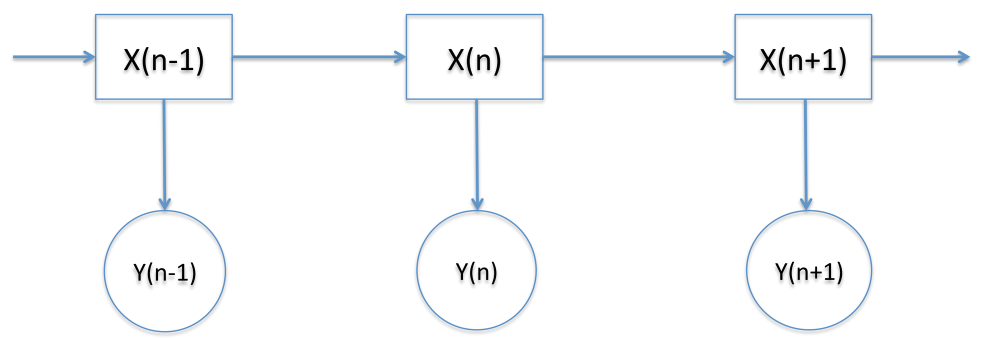
Estimate Exponential Forgetting in HMM and its Applications

Felix X.-F. Ye, Yi-An Ma and Hong Qian   University of Washington Applied Mathematics

Introduction

The forward filtering in HMM possesses the property of exponential loss of memory with some mild assumptions. We will estimate this asymptotic rate numerically and apply it to SG-MCMC algorithm.

- Latent state sequence $\{X_n\}$ in the finite set S with N states follows an irreducible and aperiodic Markov Chain M and initial probability distribution $\mathbf{p}_0$.
- The pdf for observation sequences $\{Y_n\}$ in $\mathbb{R}^d$ is $P(Y_n \in dy | X_n = i, \phi_i) = b_i(y)\mu(dy)$. Generally, it will be a Gaussian.
- The parameter of an HMM is $\theta = (M, \phi)$.



The forward algorithm gives

$$\mathbf{p}_n = P(X_n, Y_{1:n}|\theta) = \mathbf{p}_0 M D_1 \dots M D_n \quad (1)$$

where $D_j = \text{diag}(b_i(Y_j))$. One could think D_j are random matrices sampled in i.i.d manner. The conditional probability $\boldsymbol{\rho}_n = P(X_n|Y_{1:n},\theta) = \mathbf{p}_n/(\mathbf{p}_n \cdot \mathbb{1})$.
For two sequences $\boldsymbol{\rho}_n$ and $\boldsymbol{\rho}'_n$, starting with two different initial conditions, $\mathbf{p}_0$ and $\mathbf{p}'_0$ will synchronize eventually,

$$\|\boldsymbol{\rho}_n - \boldsymbol{\rho}'_n\| \to_{n\to+\infty} 0$$

Atar *et al* prove the asymptotic rate of exponential loss of memory is bounded above by the gap of the first two Lyapunov exponents of eq (1). However, explicit estimate of the gap is not practically useful.

$$\limsup_{n\to+\infty} \frac{1}{n} \log \|\boldsymbol{\rho}_n - \boldsymbol{\rho}'_n\| \le \lambda_2 - \lambda_1$$

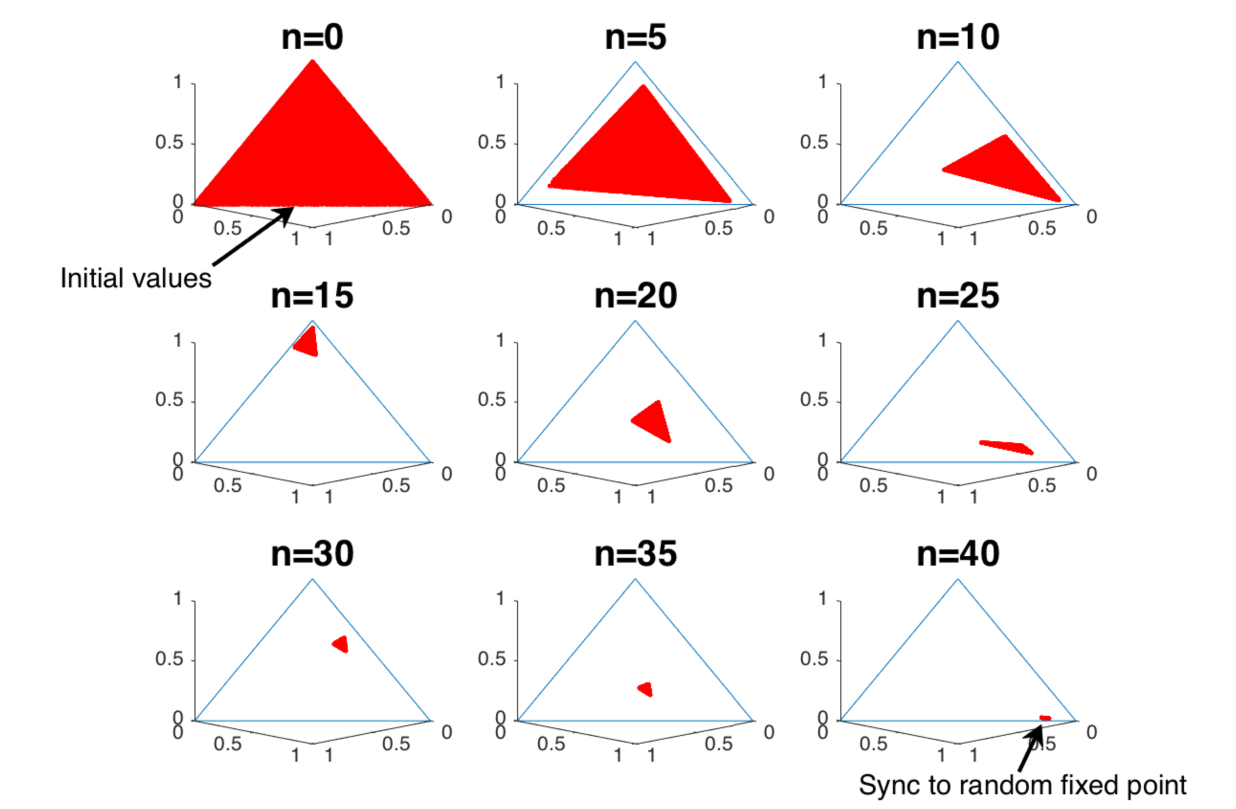
References

[1] R. Atar and O. Zeitouni, Lyapunov exponents for finite state nonlinear filtering, *SIAM J. Control Optim.* Vol. 35, No. 1, pp. 36-55, Jan 1997.

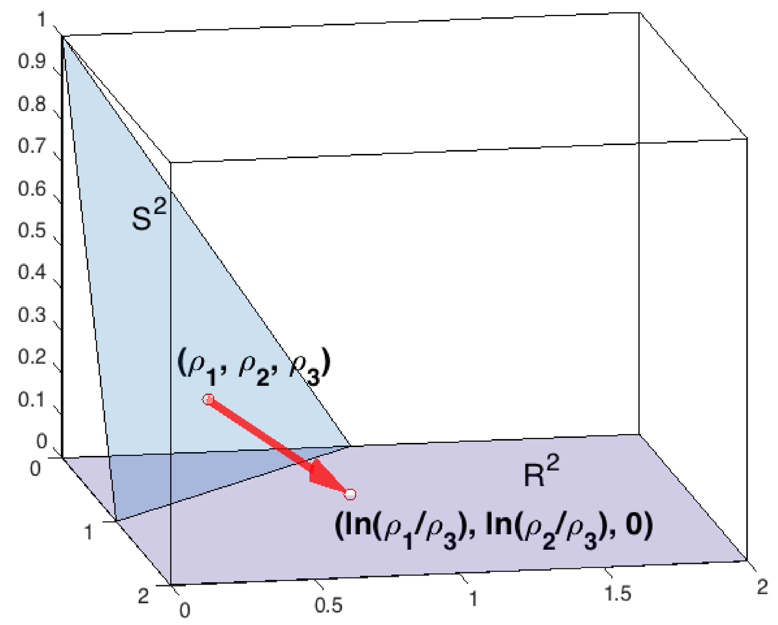
[2] Y.-A. Ma, N. Foti and E. B. Fox, *Stochastic gradient MCMC methods for hidden Markov models*, arXiv:1706.04632

[3] F. X.-F., Ye, Y.-A., Ma and H. Qian, *A numerical algorithm for calculating the rate of exponential forgetting in HMM*, manuscript in preparation.

Synchronization



RDS



Project $\boldsymbol{\rho}$ to $\mathbf{r}$:

$$[\rho_1 \dots \rho_N] \downarrow [\underbrace{\ln\frac{\rho_1}{\rho_N}}_{r_1} \dots \underbrace{\ln\frac{\rho_{N-1}}{\rho_N}}_{r_{N-1}}, 0]$$

It is a 1-to-1 mapping for interior points. The derivative of the map and its inverse exist.

The dynamics for $\boldsymbol{\rho}$ in S^{N-1} induces a random map for $\mathbf{r}$ in $\mathbb{R}^{N-1}$.

$$\mathbf{r}^{(n)} = \underbrace{\mathbf{d}^{(n)}}_{\text{random translation}} + \underbrace{F(\mathbf{r}^{(n-1)})}_{\text{deterministic map}} \quad (2)$$

$$\mathbf{d}^{(n)} = [\ln\frac{b_1(Y_n)}{b_N(Y_n)} \dots \ln\frac{b_{N-1}(Y_n)}{b_N(Y_n)}, 0] \quad (3)$$

$$F_i(\mathbf{r}) = \ln\left(\frac{\sum_{j=1}^N \exp(r_j) M_{ji}}{\sum_{j=1}^N \exp(r_j) M_{jN}}\right) \quad (4)$$

It naturally defines an i.i.d induced random dynamical system (RDS) in $\mathbb{R}^{N-1}$. The Jacobian $J(\mathbf{r}) = \nabla F(\mathbf{r})$ doesn't dependent on $\mathbf{d}$. The maximum Lyapunov exponent of induced RDS is

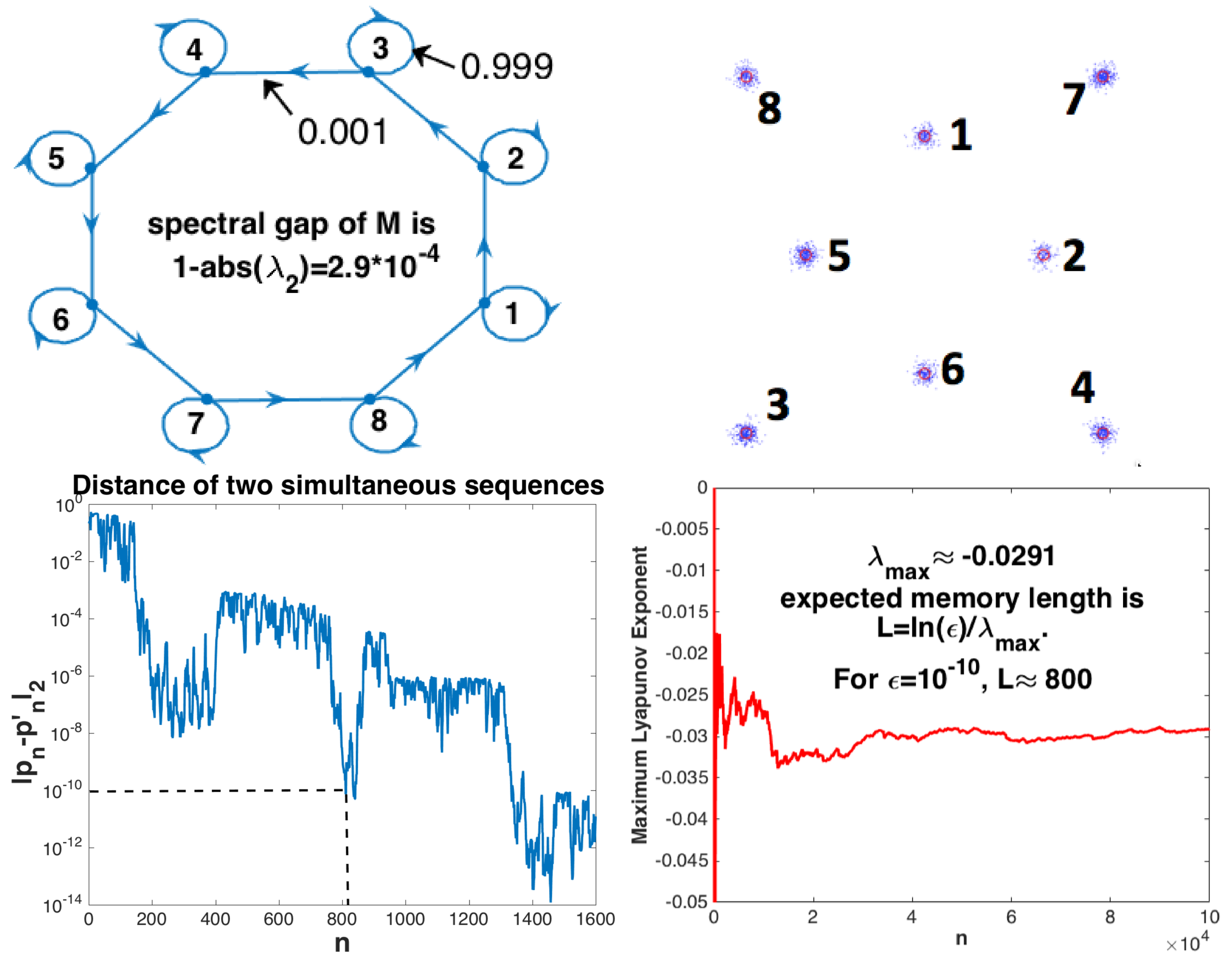
$$\lambda_{\max} = \lambda_2 - \lambda_1 = \limsup_{n\to+\infty} \frac{1}{n} \log \|J(\mathbf{r}^{n-1}) \dots J(\mathbf{r}^0)\| \quad (5)$$

The distance of two sequences with different initial conditions has the following behavior,

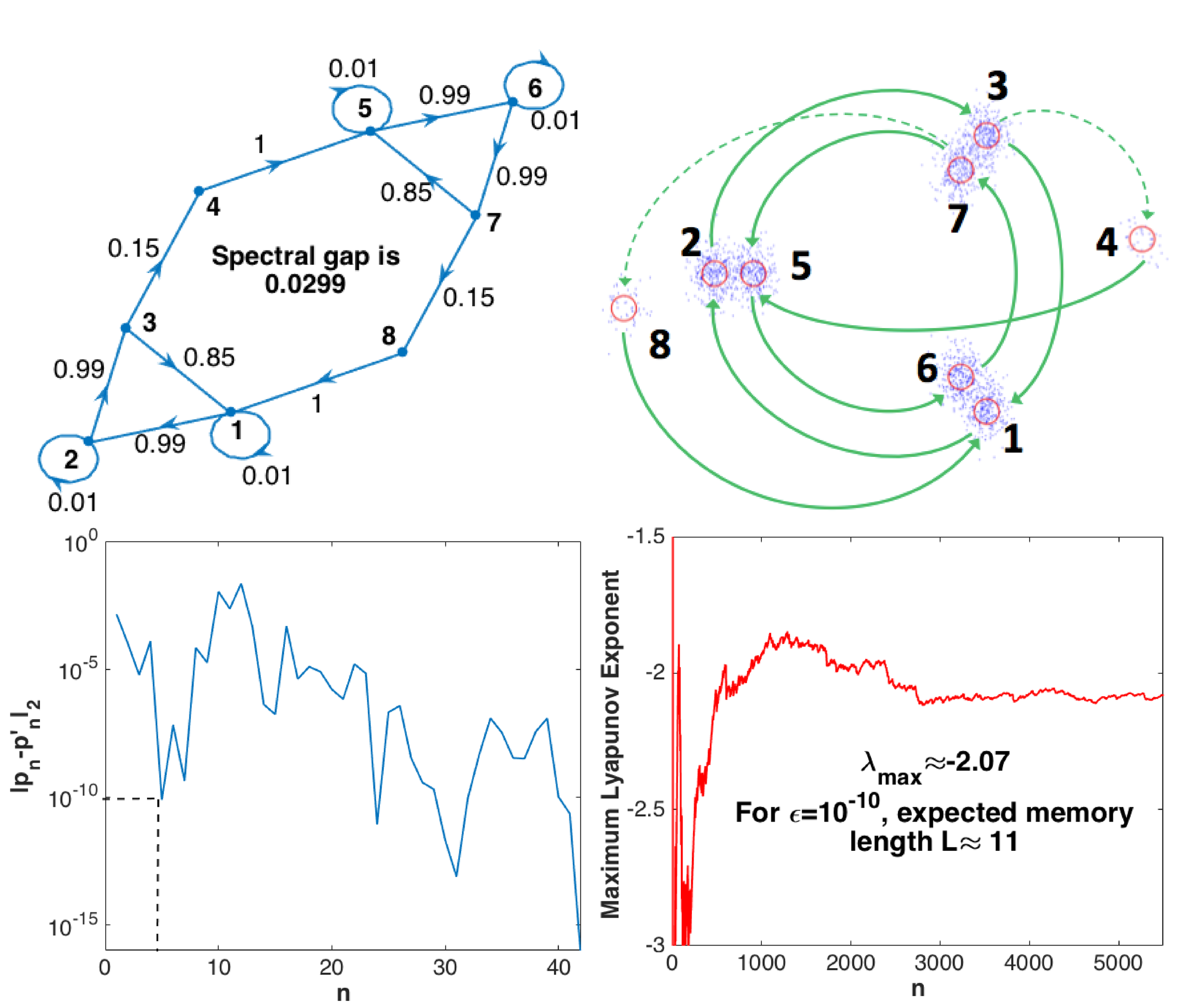
$$\|\boldsymbol{\rho}_n - \boldsymbol{\rho}'_n\| \le C\exp(\lambda_{\max} n)\|\boldsymbol{\rho}_0 - \boldsymbol{\rho}'_0\| \quad (6)$$

Synthetic Example

Diagonally Dominant

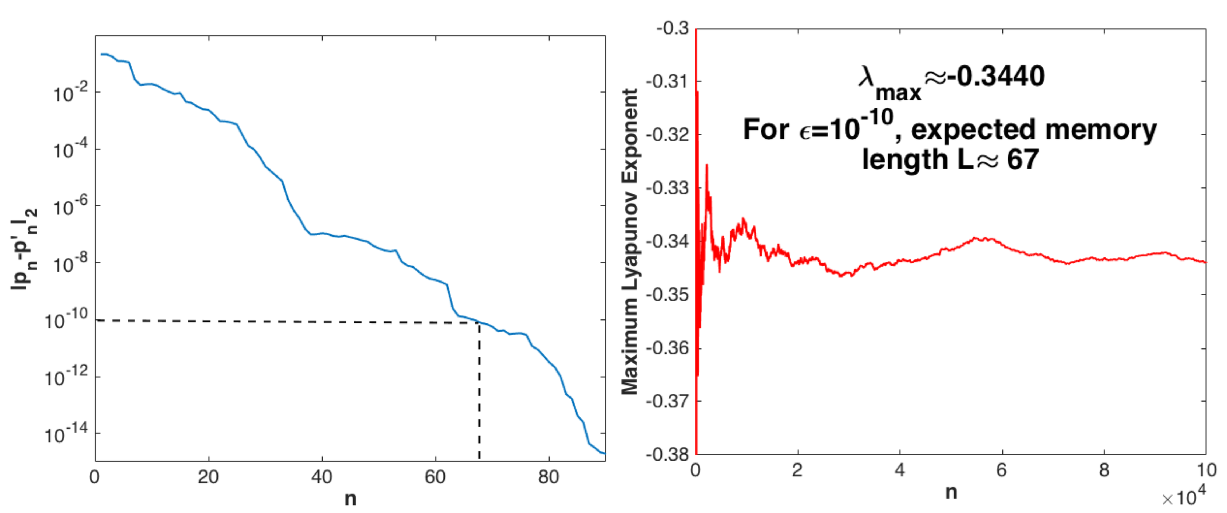


Reversed Cycles



Ion Channel Example

Inference of ion channel data. For a given parameter set θ (not shown),



Ion Channel Example

The potential function is defined

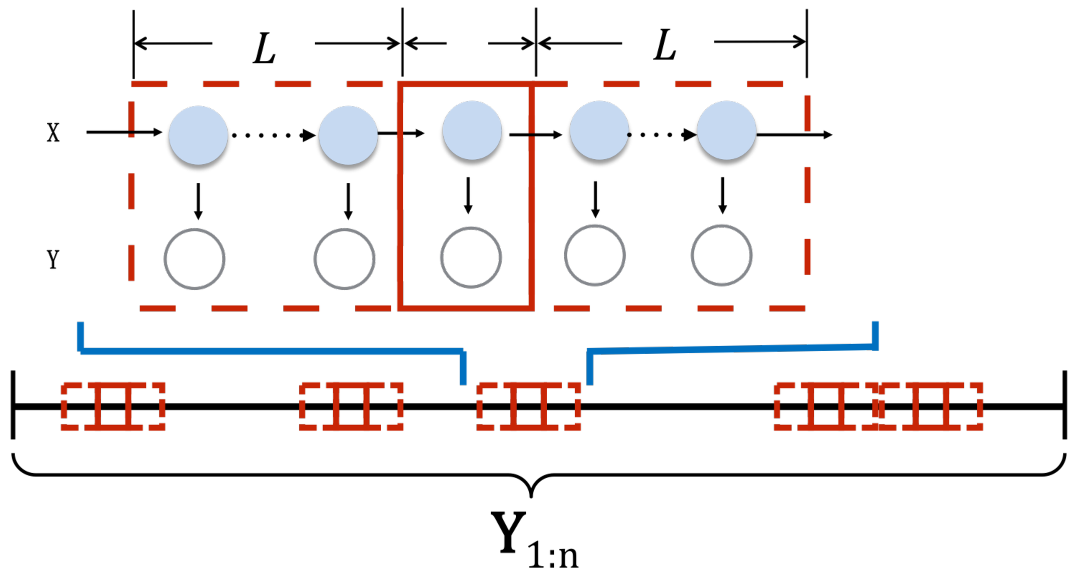
$$U(\theta) \propto -\ln P(\theta|Y_{1:n}) \quad (7)$$

$$\text{(Bayes Rule, marginalize X)} \propto -\ln(\mathbf{p}_n \cdot \mathbb{1}) - \ln P(\theta)$$

The gradient of potential function is

$$\frac{\partial U(\theta)}{\partial \theta_i} \propto -\sum_{j=1}^{n} \frac{\mathbf{p}_0 M D_1 \dots \frac{\partial (M D_j)}{\partial \theta_i} \dots M D_n \mathbb{1}}{\mathbf{p}_n \cdot \mathbb{1}} - \frac{\partial \ln P(\theta)}{\partial \theta_i} \quad (8)$$

When n is massive, eq (8) is not feasible. But one can harness the memory decay property,



Approximate eq (8) by using the memory length $L(\theta)$

$$-\sum_{j=1}^{n} \frac{\mathbf{p}_0 P(y_{LM}) \frac{\partial (M D_j)}{\partial \theta_i} P(y_{RM}) \mathbb{1}}{\mathbf{p}_0 P(y_{LM}) M D_j P(y_{RM}) \mathbb{1}} \quad (9)$$

$$P(y_{LM}) = M D_{j-L} \cdots M D_{j-1}$$

$$P(y_{RM}) = M D_{j+1} \cdots M D_{j+L}$$