

Convergence of the most probable paths in partially observed systems

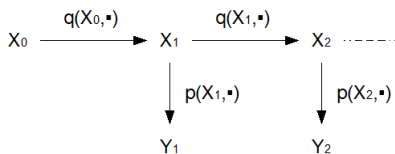
P. Chigansky, Y. Ritov

The Hebrew University, Israel

2011-03-21/ SAPS VIII, Le Mans

Hidden Markov model

HMM is a pair of processes $(X, Y) = (X_n, Y_n)_{n \in \mathbb{Z}_+}$, generated recursively:



where X is hidden and Y is observable.

HMM with finite state

- the hidden state process X is a finite state Markov chain with transition probabilities

$$q_{ij} = \mathbb{P}(X_n = j | X_{n-1} = i), \quad i, j \in \mathbb{S} = \{1, \dots, d\}$$

and $X_0 = x_0 \in \mathbb{S}$.

- the observations Y_1, Y_2, \dots are conditionally independent given $X_{1:\infty}$ and

$$\frac{d}{dy} \mathbb{P}(Y_n \leq y | X_{1:\infty}) = p(X_n, y), \quad \mathbb{P} - a.s.$$

- the transition probabilities q and the densities p are known

HMM with finite state

- the hidden state process X is a finite state Markov chain with transition probabilities

$$q_{ij} = \mathbb{P}(X_n = j | X_{n-1} = i), \quad i, j \in \mathbb{S} = \{1, \dots, d\}$$

and $X_0 = x_0 \in \mathbb{S}$.

- the observations Y_1, Y_2, \dots are conditionally independent given $X_{1:\infty}$ and

$$\frac{d}{dy} \mathbb{P}(Y_n \leq y | X_{1:\infty}) = p(X_n, y), \quad \mathbb{P} - a.s.$$

- the transition probabilities q and the densities p are known

HMM with finite state

- the hidden state process X is a finite state Markov chain with transition probabilities

$$q_{ij} = \mathbb{P}(X_n = j | X_{n-1} = i), \quad i, j \in \mathbb{S} = \{1, \dots, d\}$$

and $X_0 = x_0 \in \mathbb{S}$.

- the observations Y_1, Y_2, \dots are conditionally independent given $X_{1:\infty}$ and

$$\frac{d}{dy} \mathbb{P}(Y_n \leq y | X_{1:\infty}) = p(X_n, y), \quad \mathbb{P} - a.s.$$

- the transition probabilities q and the densities p are **known**

Path reconstruction

[Q] how to reconstruct $X_{1:n}$, given the observations $Y_{1:n}$?

[A] for a given loss function $\ell : \mathbb{S}^n \times \mathbb{S}^n \mapsto \mathbb{R}_+$, minimize the risk

$$\mathbb{E} \ell(\varphi(Y_{1:n}), X_{1:n}) \quad \text{over } \varphi : \mathbb{R}^n \mapsto \mathbb{S}^n$$

and use the minimizer $\varphi^*(Y_{1:n})$ to reconstruct $X_{1:n}$.

Path reconstruction

[Q] how to reconstruct $X_{1:n}$, given the observations $Y_{1:n}$?

[A] for a given loss function $\ell : \mathbb{S}^n \times \mathbb{S}^n \mapsto \mathbb{R}_+$, minimize the risk

$$\mathbb{E} \ell(\varphi(Y_{1:n}), X_{1:n}) \quad \text{over } \varphi : \mathbb{R}^n \mapsto \mathbb{S}^n$$

and use the minimizer $\varphi^*(Y_{1:n})$ to reconstruct $X_{1:n}$.

Local MAP path reconstruction

- “local” 0-1 loss:

$$\ell(x'_{1:n}, x_{1:n}) = \sum_{i=1}^n \mathbf{1}_{\{x'_i \neq x_i\}}$$

- the optimal estimator is the local MAP (smoothing) estimator:

$$\check{X}_m = \operatorname{argmax}_{j \in \mathcal{S}} \mathbb{P}(X_m = j | Y_{1:n}), \quad m = 1, \dots, n$$

which minimizes the sum of errors' probabilities:

$$\sum_{m=1}^n \mathbb{P}(\check{X}_m \neq X_m) \leq \sum_{m=1}^n \mathbb{P}(\varphi_m(Y_{1:n}) \neq X_m), \quad \forall \varphi : \mathbb{R}^n \mapsto \mathcal{S}^n$$

- the probabilities $\mathbb{P}(X_m = j | Y_{1:n})$ can be efficiently computed (Baum-Welch algorithm)

Local MAP path reconstruction

- “local” 0-1 loss:

$$\ell(x'_{1:n}, x_{1:n}) = \sum_{i=1}^n \mathbf{1}_{\{x'_i \neq x_i\}}$$

- the optimal estimator is the local MAP (smoothing) estimator:

$$\check{X}_m = \operatorname{argmax}_{j \in \mathcal{S}} \mathbb{P}(X_m = j | Y_{1:n}), \quad m = 1, \dots, n$$

which minimizes the sum of errors' probabilities:

$$\sum_{m=1}^n \mathbb{P}(\check{X}_m \neq X_m) \leq \sum_{m=1}^n \mathbb{P}(\varphi_m(Y_{1:n}) \neq X_m), \quad \forall \varphi : \mathbb{R}^n \mapsto \mathcal{S}^n$$

- the probabilities $\mathbb{P}(X_m = j | Y_{1:n})$ can be efficiently computed (Baum-Welch algorithm)

Local MAP path reconstruction

- “local” 0-1 loss:

$$\ell(x'_{1:n}, x_{1:n}) = \sum_{i=1}^n \mathbf{1}_{\{x'_i \neq x_i\}}$$

- the optimal estimator is the local MAP (smoothing) estimator:

$$\check{X}_m = \operatorname{argmax}_{j \in \mathcal{S}} \mathbb{P}(X_m = j | Y_{1:n}), \quad m = 1, \dots, n$$

which minimizes the sum of errors' probabilities:

$$\sum_{m=1}^n \mathbb{P}(\check{X}_m \neq X_m) \leq \sum_{m=1}^n \mathbb{P}(\varphi_m(Y_{1:n}) \neq X_m), \quad \forall \varphi : \mathbb{R}^n \mapsto \mathcal{S}^n$$

- the probabilities $\mathbb{P}(X_m = j | Y_{1:n})$ can be efficiently computed (Baum-Welch algorithm)

Global MAP path reconstruction

- “global” 0-1 loss:

$$\ell(x'_{1:n}, x_{1:n}) = \mathbf{1}_{\{x'_{1:n} \neq x_{1:n}\}}$$

- the optimal estimator is the global MAP estimator

$$\hat{X}_{1:n} = \operatorname{argmax}_{x_{1:n} \in \mathcal{S}^n} \mathbb{P}(X_{1:n} = x_{1:n} | Y_{1:n})$$

which minimizes the probability of the wrong guess:

$$\mathbb{P}(\hat{X}_{1:n} \neq X_{1:n}) \leq \mathbb{P}(\varphi(Y_{1:n}) \neq X_{1:n}), \quad \forall \varphi : \mathbb{R}^n \mapsto \mathcal{S}^n.$$

Global MAP path reconstruction

- “global” 0-1 loss:

$$\ell(x'_{1:n}, x_{1:n}) = \mathbf{1}_{\{x'_{1:n} \neq x_{1:n}\}}$$

- the optimal estimator is the global MAP estimator

$$\hat{X}_{1:n} = \operatorname{argmax}_{x_{1:n} \in \mathbb{S}^n} \mathbb{P}(X_{1:n} = x_{1:n} | Y_{1:n})$$

which minimizes the probability of the wrong guess:

$$\mathbb{P}(\hat{X}_{1:n} \neq X_{1:n}) \leq \mathbb{P}(\varphi(Y_{1:n}) \neq X_{1:n}), \quad \forall \varphi : \mathbb{R}^n \mapsto \mathbb{S}^n.$$

Global MAP path reconstruction

By the Bayes formula

$$\hat{X}_{1:n} = \operatorname{argmax}_{x_{1:n} \in \mathbb{S}^n} L_n(x_{1:n}; Y_{1:n}),$$

where L_n is the posterior “likelihood”

$$L_n(x_{1:n}; Y_{1:n}) = \prod_{i=1}^n q(x_{i-1}, x_i) p(x_i, Y_i)$$

- $\hat{X}_{1:n}$ can be computed efficiently by the Viterbi algorithm (DP)

Global MAP path reconstruction

By the Bayes formula

$$\hat{X}_{1:n} = \operatorname{argmax}_{x_{1:n} \in \mathbb{S}^n} L_n(x_{1:n}; Y_{1:n}),$$

where L_n is the posterior “likelihood”

$$L_n(x_{1:n}; Y_{1:n}) = \prod_{i=1}^n q(x_{i-1}, x_i) p(x_i, Y_i)$$

- $\hat{X}_{1:n}$ can be computed efficiently by the Viterbi algorithm (DP)

Convergence of the MAP path

Note: the optimal paths $\check{X}_{1:n}^n$ and $\hat{X}_{1:n}^n$ may change entirely when the new observation is added, e.g.:

$$\mathbb{P}\left(\hat{X}_{1:n}^n \neq \hat{X}_{1:n}^{n+1}\right) > 0$$

[Q]: Do the optimal paths $\hat{X}_{1:n}^n$ and $\check{X}_{1:n}^n$ stabilize ?
 I.e. do the limits exist:

$$\lim_{n \rightarrow \infty} \check{X}_{1:m}^n \quad \mathbb{P} - a.s. \quad \forall m \geq 1 \quad ?$$

and

$$\lim_{n \rightarrow \infty} \hat{X}_{1:m}^n \quad \mathbb{P} - a.s. \quad \forall m \geq 1 \quad ?$$

(In practical terms: is the memory bounded ?)

Convergence of the MAP path

Note: the optimal paths $\check{X}_{1:n}^n$ and $\hat{X}_{1:n}^n$ may change entirely when the new observation is added, e.g.:

$$\mathbb{P}\left(\hat{X}_{1:n}^n \neq \hat{X}_{1:n}^{n+1}\right) > 0$$

[Q]: Do the optimal paths $\hat{X}_{1:n}^n$ and $\check{X}_{1:n}^n$ stabilize ?
 i.e. do the limits exist:

$$\lim_{n \rightarrow \infty} \check{X}_{1:m}^n \quad \mathbb{P} - a.s. \quad \forall m \geq 1 \quad ?$$

and

$$\lim_{n \rightarrow \infty} \hat{X}_{1:m}^n \quad \mathbb{P} - a.s. \quad \forall m \geq 1 \quad ?$$

(In practical terms: is the memory bounded ?)

Convergence of the MAP path

Note: the optimal paths $\check{X}_{1:n}^n$ and $\hat{X}_{1:n}^n$ may change entirely when the new observation is added, e.g.:

$$\mathbb{P}\left(\hat{X}_{1:n}^n \neq \hat{X}_{1:n}^{n+1}\right) > 0$$

[Q]: Do the optimal paths $\hat{X}_{1:n}^n$ and $\check{X}_{1:n}^n$ stabilize ?
I.e. do the limits exist:

$$\lim_{n \rightarrow \infty} \check{X}_{1:m}^n \quad \mathbb{P} - a.s. \quad \forall m \geq 1 \quad ?$$

and

$$\lim_{n \rightarrow \infty} \hat{X}_{1:m}^n \quad \mathbb{P} - a.s. \quad \forall m \geq 1 \quad ?$$

(In practical terms: is the memory bounded ?)

Convergence of the MAP path

Note: the optimal paths $\check{X}_{1:n}^n$ and $\hat{X}_{1:n}^n$ may change entirely when the new observation is added, e.g.:

$$\mathbb{P}\left(\hat{X}_{1:n}^n \neq \hat{X}_{1:n}^{n+1}\right) > 0$$

[Q]: Do the optimal paths $\hat{X}_{1:n}^n$ and $\check{X}_{1:n}^n$ stabilize ?
I.e. do the limits exist:

$$\lim_{n \rightarrow \infty} \check{X}_{1:m}^n \quad \mathbb{P} - a.s. \quad \forall m \geq 1 \quad ?$$

and

$$\lim_{n \rightarrow \infty} \hat{X}_{1:m}^n \quad \mathbb{P} - a.s. \quad \forall m \geq 1 \quad ?$$

(In practical terms: is the memory bounded ?)

Convergence of the local MAP path

- by the martingale convergence the limit

$$\lim_n \mathbb{P}(X_m = j | Y_{1:n}) \text{ exists } \mathbb{P} - a.s. \quad \forall m \geq 1, \forall j \in \mathcal{S}$$

- if $\operatorname{argmax}_j \mathbb{P}(X_m = j | Y_{1:\infty})$ is unique \mathbb{P} -a.s., then local MAP path stabilize, i.e. the limit exists:

$$\lim_n \check{X}_m^n, \quad \mathbb{P} - a.s. \quad \forall m \geq 1.$$

- under certain conditions, the convergence is exponential (filtering stability, etc.)
- similar convergence for more general hidden state processes (and other local losses, e.g. MSE)

Convergence of the local MAP path

- by the martingale convergence the limit

$$\lim_n \mathbb{P}(X_m = j | Y_{1:n}) \text{ exists } \mathbb{P} - a.s. \quad \forall m \geq 1, \forall j \in \mathcal{S}$$

- if $\operatorname{argmax}_j \mathbb{P}(X_m = j | Y_{1:\infty})$ is unique \mathbb{P} -a.s., then local MAP path stabilize, i.e. the limit exists:

$$\lim_n \check{X}_m^n, \quad \mathbb{P} - a.s. \quad \forall m \geq 1.$$

- under certain conditions, the convergence is exponential (filtering stability, etc.)
- similar convergence for more general hidden state processes (and other local losses, e.g. MSE)

Convergence of the local MAP path

- by the martingale convergence the limit

$$\lim_n \mathbb{P}(X_m = j | Y_{1:n}) \text{ exists } \mathbb{P} - a.s. \quad \forall m \geq 1, \forall j \in \mathcal{S}$$

- if $\operatorname{argmax}_j \mathbb{P}(X_m = j | Y_{1:\infty})$ is unique \mathbb{P} -a.s., then local MAP path stabilize, i.e. the limit exists:

$$\lim_n \check{X}_m^n, \quad \mathbb{P} - a.s. \quad \forall m \geq 1.$$

- under certain conditions, the convergence is exponential (filtering stability, etc.)
- similar convergence for more general hidden state processes (and other local losses, e.g. MSE)

Convergence of the local MAP path

- by the martingale convergence the limit

$$\lim_n \mathbb{P}(X_m = j | Y_{1:n}) \text{ exists } \mathbb{P} - a.s. \quad \forall m \geq 1, \forall j \in \mathcal{S}$$

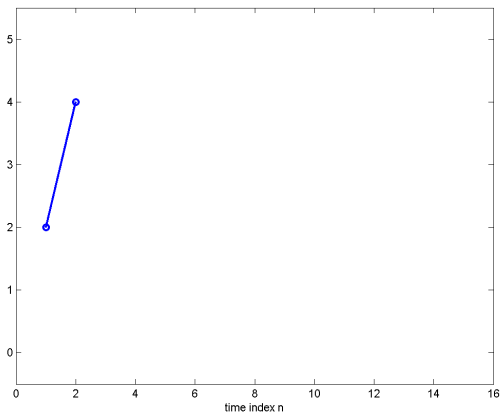
- if $\operatorname{argmax}_j \mathbb{P}(X_m = j | Y_{1:\infty})$ is unique \mathbb{P} -a.s., then local MAP path stabilize, i.e. the limit exists:

$$\lim_n \check{X}_m^n, \quad \mathbb{P} - a.s. \quad \forall m \geq 1.$$

- under certain conditions, the convergence is exponential (filtering stability, etc.)
- similar convergence for more general hidden state processes (and other local losses, e.g. MSE)

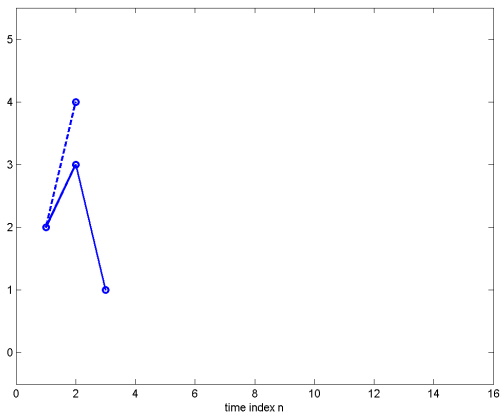
Typical *global* MAP path evolution

4-state Markov chain is hidden in additive $N(0, 1)$ noise



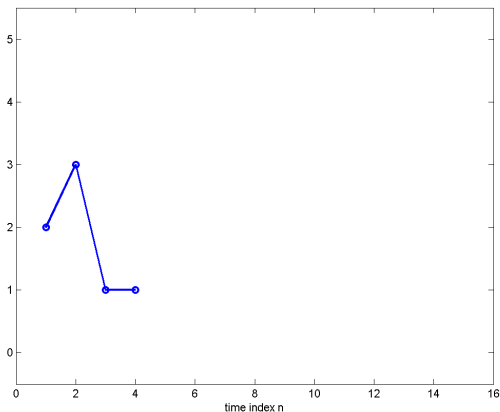
Typical *global* MAP path evolution

4-state Markov chain is hidden in additive $N(0, 1)$ noise



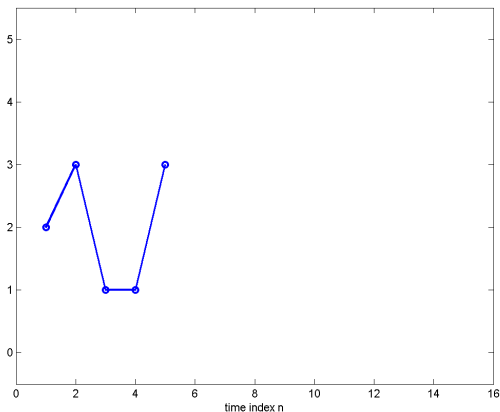
Typical *global* MAP path evolution

4-state Markov chain is hidden in additive $N(0, 1)$ noise



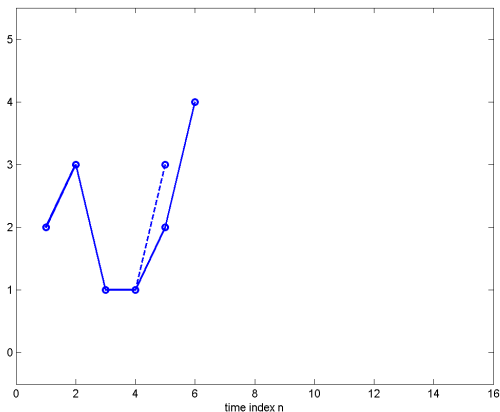
Typical *global* MAP path evolution

4-state Markov chain is hidden in additive $N(0, 1)$ noise



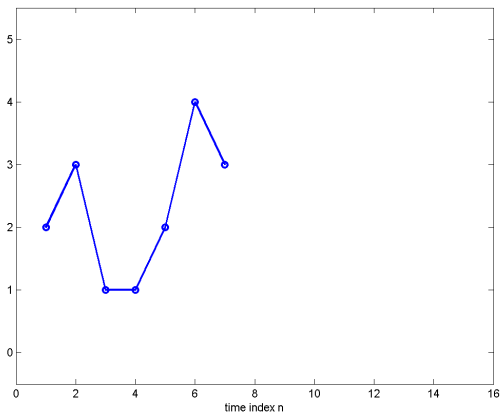
Typical *global* MAP path evolution

4-state Markov chain is hidden in additive $N(0, 1)$ noise



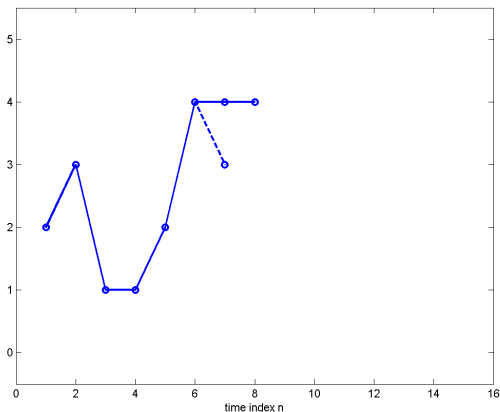
Typical *global* MAP path evolution

4-state Markov chain is hidden in additive $N(0, 1)$ noise



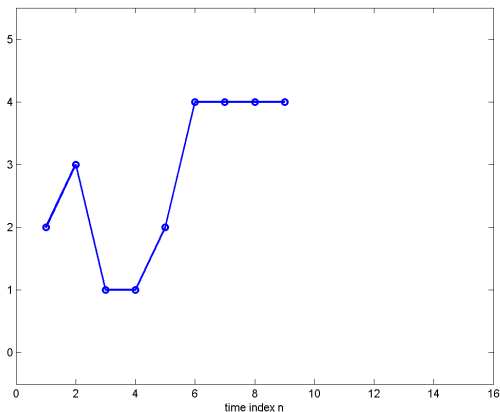
Typical *global* MAP path evolution

4-state Markov chain is hidden in additive $N(0, 1)$ noise



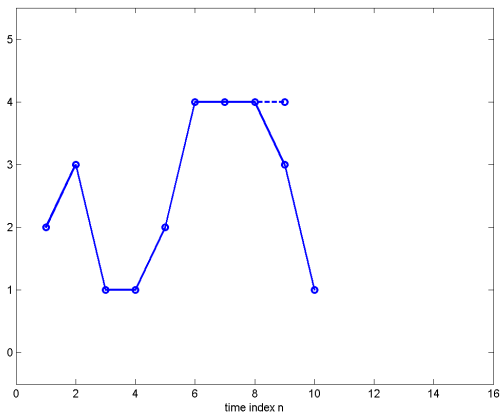
Typical *global* MAP path evolution

4-state Markov chain is hidden in additive $N(0, 1)$ noise



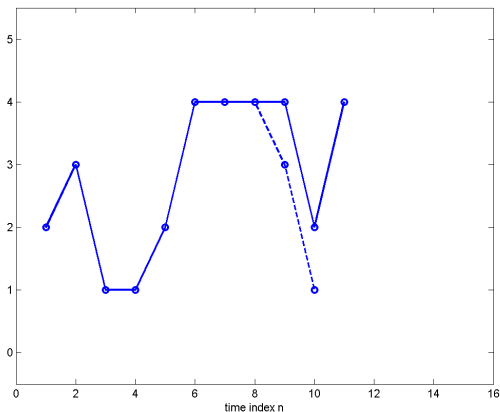
Typical *global* MAP path evolution

4-state Markov chain is hidden in additive $N(0, 1)$ noise



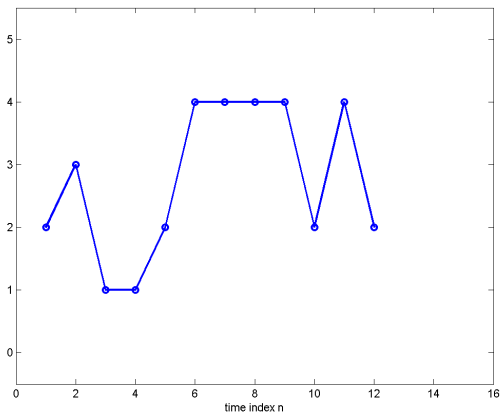
Typical *global* MAP path evolution

4-state Markov chain is hidden in additive $N(0, 1)$ noise



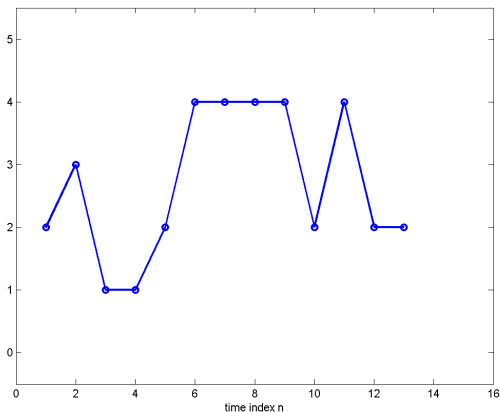
Typical *global* MAP path evolution

4-state Markov chain is hidden in additive $N(0, 1)$ noise



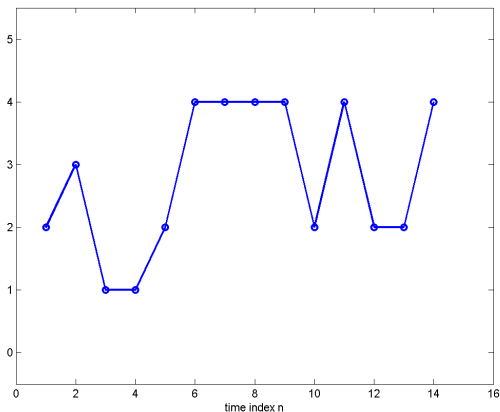
Typical *global* MAP path evolution

4-state Markov chain is hidden in additive $N(0, 1)$ noise



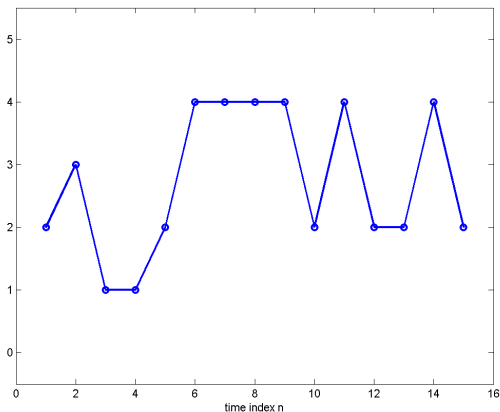
Typical *global* MAP path evolution

4-state Markov chain is hidden in additive $N(0, 1)$ noise



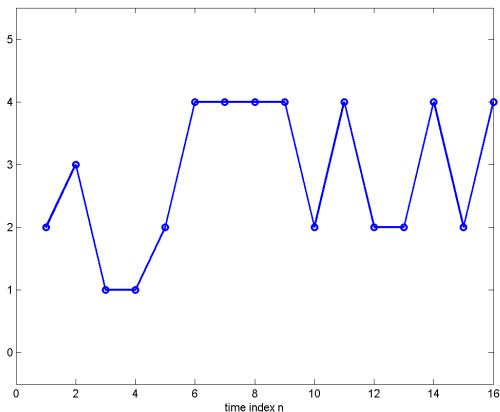
Typical *global* MAP path evolution

4-state Markov chain is hidden in additive $N(0, 1)$ noise



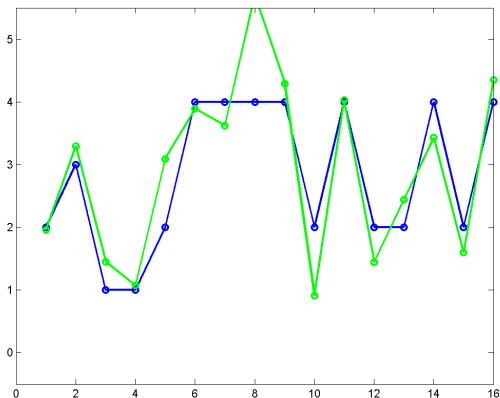
Typical *global* MAP path evolution

4-state Markov chain is hidden in additive $N(0, 1)$ noise



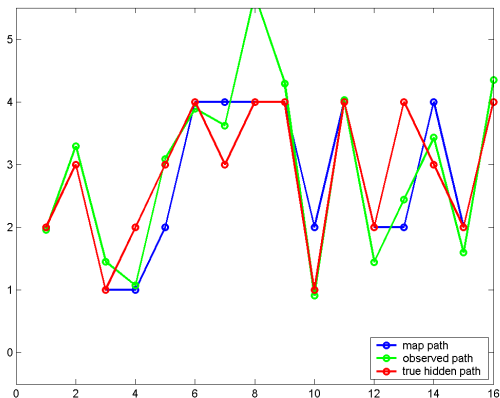
Typical *global* MAP path evolution

4-state Markov chain is hidden in additive $N(0, 1)$ noise



Typical *global* MAP path evolution

4-state Markov chain is hidden in additive $N(0, 1)$ noise



Does the global MAP path stabilize ?

- Recall that $\hat{X}_{1:n}^n$ is the maximizer of the likelihood:

$$L_n(x_{1:n}, Y_{1:n}) = \prod_{m=1}^n q(x_{m-1}, x_m) p(x_m, Y_m) = \dots q(x_{k-1}, x_k) p(x_k, Y_k) q(x_k, x_{k+1}) \dots$$

- If the “locking” event

$$A_k(j_0) = \left\{ j_0 = \operatorname{argmax}_{j \in \mathcal{S}} q(u, j) p(j, Y_k) q(j, v), \forall u, v \in \mathcal{S} \right\}$$

occurs, then

$$\hat{X}_k^n = j_0 \implies \hat{X}_{1:k}^n \text{ is determined by } Y_{1:k} \text{ for all } n \geq k$$

Does the global MAP path stabilize ?

- Recall that $\hat{X}_{1:n}^n$ is the maximizer of the likelihood:

$$L_n(x_{1:n}, Y_{1:n}) = \prod_{m=1}^n q(x_{m-1}, x_m) p(x_m, Y_m) = \\ \cdots q(x_{k-1}, x_k) p(x_k, Y_k) q(x_k, x_{k+1}) \cdots$$

- If the “locking” event

$$A_k(j_0) = \left\{ j_0 = \operatorname{argmax}_{j \in \mathbb{S}} q(u, j) p(j, Y_k) q(j, v), \forall u, v \in \mathbb{S} \right\}$$

occurs, then

$$\hat{X}_k^n = j_0 \implies \hat{X}_{1:k}^n \text{ is determined by } Y_{1:k} \text{ for all } n \geq k$$

Does the global MAP path stabilize ?

- Recall that $\hat{X}_{1:n}^n$ is the maximizer of the likelihood:

$$L_n(x_{1:n}, Y_{1:n}) = \prod_{m=1}^n q(x_{m-1}, x_m) p(x_m, Y_m) = \\ \cdots q(x_{k-1}, x_k) p(x_k, Y_k) q(x_k, x_{k+1}) \cdots$$

- If the “locking” event

$$A_k(j_0) = \left\{ j_0 = \operatorname{argmax}_{j \in \mathbb{S}} q(u, j) p(j, Y_k) q(j, v), \forall u, v \in \mathbb{S} \right\}$$

occurs, then

$$\hat{X}_k^n = j_0 \implies \hat{X}_{1:k}^n \text{ is determined by } Y_{1:k} \text{ for all } n \geq k$$

Does the global MAP path stabilize ?

- Recall that $\hat{X}_{1:n}^n$ is the maximizer of the likelihood:

$$L_n(x_{1:n}, Y_{1:n}) = \prod_{m=1}^n q(x_{m-1}, x_m) p(x_m, Y_m) = \\ \cdots q(x_{k-1}, x_k) p(x_k, Y_k) q(x_k, x_{k+1}) \cdots$$

- If the “locking” event

$$A_k(j_0) = \left\{ j_0 = \operatorname{argmax}_{j \in \mathbb{S}} q(u, j) p(j, Y_k) q(j, v), \forall u, v \in \mathbb{S} \right\}$$

occurs, then

$$\hat{X}_k^n = j_0 \implies \hat{X}_{1:k}^n \text{ is determined by } Y_{1:k} \text{ for all } n \geq k$$

Does the global MAP path stabilize ?

- Recall that $\hat{X}_{1:n}^n$ is the maximizer of the likelihood:

$$L_n(x_{1:n}, Y_{1:n}) = \prod_{m=1}^n q(x_{m-1}, x_m) p(x_m, Y_m) = \\ \cdots q(x_{k-1}, x_k) p(x_k, Y_k) q(x_k, x_{k+1}) \cdots$$

- If the “locking” event

$$A_k(j_0) = \left\{ j_0 = \operatorname{argmax}_{j \in \mathbb{S}} q(u, j) p(j, Y_k) q(j, v), \forall u, v \in \mathbb{S} \right\}$$

occurs, then

$$\hat{X}_k^n = j_0 \implies \hat{X}_{1:k}^n \text{ is determined by } Y_{1:k} \text{ for all } n \geq k$$

Does the global MAP path stabilize ?

Theorem (A.Caliebe & U.Roesler, '02 / A.Caliebe '06)

If X is an irreducible recurrent finite state Markov chain and

$$\mathbb{P}_{i_0}(A_1(j_0)) > 0, \quad \text{for some } j_0 \text{ and } i_0, \quad (\text{CR})$$

then the limit

$$\hat{X}_m^\infty := \lim_{n \rightarrow \infty} \hat{X}_m^n$$

exists and, moreover, the sequence $(\hat{X}_m^\infty)_{m \geq 1}$ is a renewal process.

- A.Koloydenko & J.Lember, '07-'08 develop the argument and give a weaker and more sophisticated condition than (CR)
- no counterexample is known (for finite state chains)

Does the global MAP path stabilize ?

Theorem (A.Caliebe & U.Roesler, '02 / A.Caliebe '06)

If X is an irreducible recurrent finite state Markov chain and

$$\mathbb{P}_{i_0}(A_1(j_0)) > 0, \quad \text{for some } j_0 \text{ and } i_0, \quad (\text{CR})$$

then the limit

$$\hat{X}_m^\infty := \lim_{n \rightarrow \infty} \hat{X}_m^n$$

exists and, moreover, the sequence $(\hat{X}_m^\infty)_{m \geq 1}$ is a renewal process.

- A.Koloydenko & J.Lember, '07-'08 develop the argument and give a weaker and more sophisticated condition than (CR)
- no counterexample is known (for finite state chains)

Does the global MAP path stabilize ?

Theorem (A.Caliebe & U.Roesler, '02 / A.Caliebe '06)

If X is an irreducible recurrent finite state Markov chain and

$$\mathbb{P}_{i_0}(A_1(j_0)) > 0, \quad \text{for some } j_0 \text{ and } i_0, \quad (\text{CR})$$

then the limit

$$\hat{X}_m^\infty := \lim_{n \rightarrow \infty} \hat{X}_m^n$$

exists and, moreover, the sequence $(\hat{X}_m^\infty)_{m \geq 1}$ is a renewal process.

- A.Koloydenko & J.Lember, '07-'08 develop the argument and give a weaker and more sophisticated condition than (CR)
- no counterexample is known (for finite state chains)

HMM: continuous hidden state

- the hidden process X is a Markov chain with transition **density**

$$q(u, v) = \frac{d}{dv} \mathbb{P}(X_n \leq v | X_{n-1} = u), \quad u, v \in \mathbb{R}$$

and $X_0 = x_0 \in \mathbb{R}$.

- the MAP estimate

$$\hat{X}_{1:n}^n = \operatorname{argmax}_{x_{1:n}} L_n(x_{1:n}, Y_{1:n}) = \operatorname{argmax}_{x_{1:n}} \prod_{j=1}^n q(x_{j-1}, x_j) p(x_j, Y_j)$$

is the most probable path in the *infinitesimal* sense:

$$\lim_{\varepsilon \rightarrow 0} \varepsilon^{-n} \mathbb{P} \left(\|X_{1:n} - \hat{X}_{1:n}\|_{\infty} \leq \varepsilon \right) \geq \lim_{\varepsilon \rightarrow 0} \varepsilon^{-n} \mathbb{P} \left(\|X_{1:n} - \varphi(Y_{1:n})\|_{\infty} \leq \varepsilon \right), \quad \forall \varphi$$

HMM: continuous hidden state

- the hidden process X is a Markov chain with transition density

$$q(u, v) = \frac{d}{dv} \mathbb{P}(X_n \leq v | X_{n-1} = u), \quad u, v \in \mathbb{R}$$

and $X_0 = x_0 \in \mathbb{R}$.

- the MAP estimate

$$\hat{X}_{1:n}^n = \operatorname{argmax}_{x_{1:n}} L_n(x_{1:n}, Y_{1:n}) = \operatorname{argmax}_{x_{1:n}} \prod_{j=1}^n q(x_{j-1}, x_j) p(x_j, Y_j)$$

is the most probable path in the *infinitesimal* sense:

$$\lim_{\varepsilon \rightarrow 0} \varepsilon^{-n} \mathbb{P} \left(\|X_{1:n} - \hat{X}_{1:n}\|_{\infty} \leq \varepsilon \right) \geq \lim_{\varepsilon \rightarrow 0} \varepsilon^{-n} \mathbb{P} \left(\|X_{1:n} - \varphi(Y_{1:n})\|_{\infty} \leq \varepsilon \right), \quad \forall \varphi$$

Convergence of the global MAP path

Study case I: linear auto-regression driven by double-sided exponential noise with Gaussian observations:

$$\begin{aligned} X_n &= X_{n-1} + \eta_n, & \xi_n &\sim N(0, 1), & \eta_n &\sim \frac{1}{4}e^{-|x|/2}. \\ Y_n &= X_n + \xi_n, \end{aligned}$$

- $\hat{X}_{1:n}^n$ is the minimizer of

$$\begin{aligned} -\log L_n(x_{1:n}, Y_{1:n}) &\propto \sum_{i=1}^n \left(|x_i - x_{i-1}| + (x_i - Y_i)^2 \right) = \\ &\cdots |x_k - x_{k-1}| + (x_k - Y_k)^2 + |x_{k+1} - x_k| \cdots \end{aligned}$$

- the function

$$x \mapsto |x - a| + (x - y)^2 + |b - x|$$

attains its minimum at $x^* \in [y - 1, y + 1]$ and if $a \leq y \leq b$ then $x^* = y$

Convergence of the global MAP path

Study case I: linear auto-regression driven by double-sided exponential noise with Gaussian observations:

$$\begin{aligned} X_n &= X_{n-1} + \eta_n, & \xi_n &\sim N(0, 1), & \eta_n &\sim \frac{1}{4}e^{-|x|/2}. \\ Y_n &= X_n + \xi_n, \end{aligned}$$

- $\hat{X}_{1:n}^n$ is the minimizer of

$$\begin{aligned} -\log L_n(x_{1:n}, Y_{1:n}) &\propto \sum_{i=1}^n \left(|x_i - x_{i-1}| + (x_i - Y_i)^2 \right) = \\ &\cdots |x_k - x_{k-1}| + (x_k - Y_k)^2 + |x_{k+1} - x_k| \cdots \end{aligned}$$

- the function

$$x \mapsto |x - a| + (x - y)^2 + |b - x|$$

attains its minimum at $x^* \in [y - 1, y + 1]$ and if $a \leq y \leq b$ then $x^* = y$

Convergence of the global MAP path

Study case I: linear auto-regression driven by double-sided exponential noise with Gaussian observations:

$$\begin{aligned} X_n &= X_{n-1} + \eta_n, & \xi_n &\sim N(0, 1), & \eta_n &\sim \frac{1}{4}e^{-|x|/2}. \\ Y_n &= X_n + \xi_n, \end{aligned}$$

- $\hat{X}_{1:n}^n$ is the minimizer of

$$\begin{aligned} -\log L_n(x_{1:n}, Y_{1:n}) &\propto \sum_{i=1}^n \left(|x_i - x_{i-1}| + (x_i - Y_i)^2 \right) = \\ &\cdots |x_k - x_{k-1}| + (x_k - Y_k)^2 + |x_{k+1} - x_k| \cdots \end{aligned}$$

- the function

$$x \mapsto |x - a| + (x - y)^2 + |b - x|$$

attains its minimum at $x^* \in [y - 1, y + 1]$ and if $a \leq y \leq b$ then $x^* = y$

Convergence of the global MAP path

- Thus

$$\hat{X}_{m-1}^n \leq Y_{m-1} + 1$$

$$\hat{X}_{m+1}^n \geq Y_{m+1} - 1$$

and so

$$A_m := \{Y_{m-1} + 1 \leq Y_m \leq Y_{m+1} - 1\} \subseteq \\ \{\hat{X}_{m-1}^n \leq Y_m \leq \hat{X}_{m+1}^n\} \subseteq \{\hat{X}_m^n \equiv Y_m\}$$

- the convergence follows from recurrence of A_m 's, as in the discrete case.

Convergence of the global MAP path

- Thus

$$\hat{X}_{m-1}^n \leq Y_{m-1} + 1$$

$$\hat{X}_{m+1}^n \geq Y_{m+1} - 1$$

and so

$$A_m := \{Y_{m-1} + 1 \leq Y_m \leq Y_{m+1} - 1\} \subseteq \\ \{\hat{X}_{m-1}^n \leq Y_m \leq \hat{X}_{m+1}^n\} \subseteq \{\hat{X}_m^n \equiv Y_m\}$$

- the convergence follows from recurrence of A_m 's, as in the discrete case.

Convergence of the global MAP path

- Thus

$$\hat{X}_{m-1}^n \leq Y_{m-1} + 1$$

$$\hat{X}_{m+1}^n \geq Y_{m+1} - 1$$

and so

$$A_m := \{Y_{m-1} + 1 \leq Y_m \leq Y_{m+1} - 1\} \subseteq \\ \{\hat{X}_{m-1}^n \leq Y_m \leq \hat{X}_{m+1}^n\} \subseteq \{\hat{X}_m^n \equiv Y_m\}$$

- the convergence follows from recurrence of A_m 's, as in the discrete case.

Convergence of the global MAP path

- Thus

$$\hat{X}_{m-1}^n \leq Y_{m-1} + 1$$

$$\hat{X}_{m+1}^n \geq Y_{m+1} - 1$$

and so

$$A_m := \{Y_{m-1} + 1 \leq Y_m \leq Y_{m+1} - 1\} \subseteq \\ \{\hat{X}_{m-1}^n \leq Y_m \leq \hat{X}_{m+1}^n\} \subseteq \{\hat{X}_m^n \equiv Y_m\}$$

- the convergence follows from recurrence of A_m 's, **as in the discrete case.**

Convergence of the global MAP path

Study case II: linear Gaussian auto-regression:

$$\begin{aligned} X_n &= X_{n-1} + \eta_n, & \xi_n &\sim N(0, 1), & \eta_n &\sim N(0, 1). \\ Y_n &= X_n + \xi_n, \end{aligned}$$

- The conditional law of $X_{1:n}$ given $Y_{1:n}$ is Gaussian, for which the mode coincides with the expectation, i.e.

$$\hat{X}_{1:n}^n = \mathbb{E}(X_{1:n} | Y_{1:n})$$

- The limit exists by the martingale convergence:

$$\lim_{n \rightarrow \infty} \hat{X}_m^n = \lim_{n \rightarrow \infty} \mathbb{E}(X_m | Y_{1:n})$$

- The convergence is essentially different from the discrete case:

$$\mathbb{P}(\hat{X}_m^n = \hat{X}_m^{n'} \text{ for some } m) = 0, \quad \text{for any } n \neq n'.$$

No “locking” events!

Convergence of the global MAP path

Study case II: linear Gaussian auto-regression:

$$\begin{aligned} X_n &= X_{n-1} + \eta_n, & \xi_n &\sim N(0, 1), & \eta_n &\sim N(0, 1). \\ Y_n &= X_n + \xi_n, \end{aligned}$$

- The conditional law of $X_{1:n}$ given $Y_{1:n}$ is Gaussian, for which the mode coincides with the expectation, i.e.

$$\hat{X}_{1:n}^n = \mathbb{E}(X_{1:n} | Y_{1:n})$$

- The limit exists by the martingale convergence:

$$\lim_{n \rightarrow \infty} \hat{X}_m^n = \lim_{n \rightarrow \infty} \mathbb{E}(X_m | Y_{1:n})$$

- The convergence is essentially different from the discrete case:

$$\mathbb{P}(\hat{X}_m^n = \hat{X}_m^{n'} \text{ for some } m) = 0, \quad \text{for any } n \neq n'.$$

No “locking” events!

Convergence of the global MAP path

Study case II: linear Gaussian auto-regression:

$$\begin{aligned} X_n &= X_{n-1} + \eta_n, & \xi_n &\sim N(0, 1), & \eta_n &\sim N(0, 1). \\ Y_n &= X_n + \xi_n, \end{aligned}$$

- The conditional law of $X_{1:n}$ given $Y_{1:n}$ is Gaussian, for which the mode coincides with the expectation, i.e.

$$\hat{X}_{1:n}^n = \mathbb{E}(X_{1:n} | Y_{1:n})$$

- The limit exists by the martingale convergence:

$$\lim_{n \rightarrow \infty} \hat{X}_m^n = \lim_{n \rightarrow \infty} \mathbb{E}(X_m | Y_{1:n})$$

- The convergence is essentially different from the discrete case:

$$\mathbb{P}(\hat{X}_m^n = \hat{X}_m^{n'} \text{ for some } m) = 0, \quad \text{for any } n \neq n'.$$

No “locking” events!

Convergence of the global MAP path

Study case II: linear Gaussian auto-regression:

$$\begin{aligned} X_n &= X_{n-1} + \eta_n, & \xi_n &\sim N(0, 1), & \eta_n &\sim N(0, 1). \\ Y_n &= X_n + \xi_n, \end{aligned}$$

- The conditional law of $X_{1:n}$ given $Y_{1:n}$ is Gaussian, for which the mode coincides with the expectation, i.e.

$$\hat{X}_{1:n}^n = \mathbb{E}(X_{1:n} | Y_{1:n})$$

- The limit exists by the martingale convergence:

$$\lim_{n \rightarrow \infty} \hat{X}_m^n = \lim_{n \rightarrow \infty} \mathbb{E}(X_m | Y_{1:n})$$

- The convergence is essentially different from the discrete case:

$$\mathbb{P}(\hat{X}_m^n = \hat{X}_m^{n'} \text{ for some } m) = 0, \quad \text{for any } n \neq n'.$$

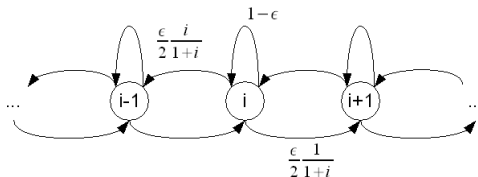
No “locking” events!

Convergence of the global MAP path

Study case III: **no convergence at all**

HMM with the bivariate state process $X_n = (U_n, V_n)$:

- U_n are (“auxiliary”) i.i.d. $U([0, 1])$ random variables
- V_n is a **positive recurrent** random walk on positive integers with “slow” ($\epsilon \ll 1$) transition probabilities:



and the initial (invariant) distribution: $\pi(j) \propto \frac{1}{j^2}$, $j \gg 1$

Convergence of the global MAP path

- The observations Y_n are sampled according to the rule:

$$U_n \in A_i, V_n < i \implies Y_n \sim U([0, 1])$$

$$U_n \in A_i, V_n \geq i \implies Y_n \sim U(A_i)$$

where

$$[0, 1] = \underbrace{[a_0, a_1)}_{A_1} \cup \underbrace{[a_1, a_2)}_{A_2} \cup \dots$$

and the lengths ℓ_i are exponentially decreasing:

$$\ell_i := |A_i| = a_i - a_{i-1} \propto \rho^i, \quad \rho < 1.$$

- The observation density is

$$p((u, v), y) = \mathbf{1}_{\{y \in [0, 1]\}} \mathbf{1}_{\{u \notin \cup_{i=1}^v A_i\}} + \sum_{i=1}^v \ell_i^{-1} \mathbf{1}_{\{(u, y) \in A_i \times A_i\}}.$$

Convergence of the global MAP path

- The observations Y_n are sampled according to the rule:

$$U_n \in A_i, V_n < i \implies Y_n \sim U([0, 1])$$

$$U_n \in A_i, V_n \geq i \implies Y_n \sim U(A_i)$$

where

$$[0, 1] = \underbrace{[a_0, a_1)}_{A_1} \cup \underbrace{[a_1, a_2)}_{A_2} \cup \dots$$

and the lengths ℓ_i are exponentially decreasing:

$$\ell_i := |A_i| = a_i - a_{i-1} \propto \rho^i, \quad \rho < 1.$$

- The observation density is

$$p((u, v), y) = \mathbf{1}_{\{y \in [0, 1]\}} \mathbf{1}_{\{u \notin \cup_{i=1}^v A_i\}} + \sum_{i=1}^v \ell_i^{-1} \mathbf{1}_{\{(u, y) \in A_i \times A_i\}}.$$

Convergence of the global MAP path

- i. If $\{Y_k \in A_j\}$ is observed, the choice

$$\hat{U}_k^n \in A_j \quad \text{and} \quad \hat{V}_k^n \geq j$$

yields a large gain $1/\ell_j$

- ii. Since the transitions are slow, constant trajectories $V_{1:n}$ are more probable and in view of (i), for some constant $c(n)$

$$\hat{V}_{1:n}^n \equiv c(n) \geq \{j : \max_{m \leq n} Y_m \in A_j\} =: j^*(n)$$

- iii. Since the initial distribution $\pi(j) \propto 1/j^2$, \hat{V}_1^n and thus the whole path $\hat{V}_{1:n}^n$, cannot be too high.

Convergence of the global MAP path

- i. If $\{Y_k \in A_j\}$ is observed, the choice

$$\hat{U}_k^n \in A_j \quad \text{and} \quad \hat{V}_k^n \geq j$$

yields a large gain $1/\ell_j$

- ii. Since the transitions are slow, constant trajectories $V_{1:n}$ are more probable and in view of (i), for some constant $c(n)$

$$\hat{V}_{1:n}^n \equiv c(n) \geq \{j : \max_{m \leq n} Y_m \in A_j\} =: j^*(n)$$

- iii. Since the initial distribution $\pi(j) \propto 1/j^2$, \hat{V}_1^n and thus the whole path $\hat{V}_{1:n}^n$, cannot be too high.

Convergence of the global MAP path

- i. If $\{Y_k \in A_j\}$ is observed, the choice

$$\hat{U}_k^n \in A_j \quad \text{and} \quad \hat{V}_k^n \geq j$$

yields a large gain $1/\ell_j$

- ii. Since the transitions are slow, constant trajectories $V_{1:n}$ are more probable and in view of (i), for some constant $c(n)$

$$\hat{V}_{1:n}^n \equiv c(n) \geq \{j : \max_{m \leq n} Y_m \in A_j\} =: j^*(n)$$

- iii. Since the initial distribution $\pi(j) \propto 1/j^2$, \hat{V}_1^n and thus the whole path $\hat{V}_{1:n}^n$, cannot be too high.

Convergence of the global MAP path

- it follows,

$$\hat{V}_m^n = \{j : \max_{m \leq n} Y_m \in A_j\} =: j^*(n), \quad n \geq 1$$

- (Y_n) visits any A_i eventually, hence

$$\hat{V}_m^n = j^*(n) \xrightarrow{n \rightarrow \infty} \infty, \quad \forall m \geq 1$$

- the MAP path **diverges** (and is **inadequate** estimate).

Convergence of the global MAP path

- it follows,

$$\hat{V}_m^n = \{j : \max_{m \leq n} Y_m \in A_j\} =: j^*(n), \quad n \geq 1$$

- (Y_n) visits any A_i eventually, hence

$$\hat{V}_m^n = j^*(n) \xrightarrow{n \rightarrow \infty} \infty, \quad \forall m \geq 1$$

- the MAP path **diverges** (and is **inadequate** estimate).

Convergence of the global MAP path

- it follows,

$$\hat{V}_m^n = \{j : \max_{m \leq n} Y_m \in A_j\} =: j^*(n), \quad n \geq 1$$

- (Y_n) visits any A_i eventually, hence

$$\hat{V}_m^n = j^*(n) \xrightarrow{n \rightarrow \infty} \infty, \quad \forall m \geq 1$$

- the MAP path **diverges** (and is **inadequate** estimate).

The strong log-concave densities

Theorem (P.Chigansky & Y. Ritov, '10)

The limit

$$\lim_{n \rightarrow \infty} \hat{X}_m^n$$

exists \mathbb{P} -a.s. for all $m \geq 1$, if

- a1. $q(u, v) \propto e^{-\alpha(u, v)}$, where $(u, v) \mapsto \alpha(u, v) \geq 0$ is twice cont. diff. and convex
- a2. there is a non-decreasing function $g : \mathbb{R}_+ \mapsto \mathbb{R}_+$ with polynomial growth, such that for all $M > 0$

$$\alpha(x, y) \leq M \quad \implies \quad \left| \frac{\partial^2}{\partial x \partial y} \alpha(x, y) \right| \leq g(M), \quad \forall x, y \in \mathbb{R}$$

The strong log-concave densities

Theorem (P.Chigansky & Y. Ritov, '10)

The limit

$$\lim_{n \rightarrow \infty} \hat{X}_m^n$$

exists \mathbb{P} -a.s. for all $m \geq 1$, if

- a1. $q(u, v) \propto e^{-\alpha(u, v)}$, where $(u, v) \mapsto \alpha(u, v) \geq 0$ is twice cont. diff. and convex
- a2. there is a non-decreasing function $g : \mathbb{R}_+ \mapsto \mathbb{R}_+$ with polynomial growth, such that for all $M > 0$

$$\alpha(x, y) \leq M \quad \implies \quad \left| \frac{\partial^2}{\partial x \partial y} \alpha(x, y) \right| \leq g(M), \quad \forall x, y \in \mathbb{R}$$

The strong log-concave densities

Theorem (P.Chigansky & Y. Ritov, '10)

The limit

$$\lim_{n \rightarrow \infty} \hat{X}_m^n$$

exists \mathbb{P} -a.s. for all $m \geq 1$, if

- a1. $q(u, v) \propto e^{-\alpha(u, v)}$, where $(u, v) \mapsto \alpha(u, v) \geq 0$ is twice cont. diff. and convex
- a2. there is a non-decreasing function $g : \mathbb{R}_+ \mapsto \mathbb{R}_+$ with polynomial growth, such that for all $M > 0$

$$\alpha(x, y) \leq M \quad \implies \quad \left| \frac{\partial^2}{\partial x \partial y} \alpha(x, y) \right| \leq g(M), \quad \forall x, y \in \mathbb{R}$$

The strong log-concave densities

- a3. $p(x, y) \propto e^{-\gamma(x, y)}$, where $x \mapsto \gamma(x, y) \geq 0$ is twice cont. diff. and strongly convex with

$$x^*(y) = \operatorname{argmin}_{x \in \mathbb{R}} \gamma(x, y) \in (-\infty, \infty)$$

and

$$\frac{\partial^2}{\partial x^2} \gamma(x, y) \geq \kappa > 0, \quad \forall x, y \in \mathbb{R}.$$

- a4. for some constant $C > 0$,

$$-\overline{\lim}_{n \rightarrow \infty} \frac{1}{n} \log L_n(X_{1:n}, Y_{1:n}) \leq C, \quad \mathbb{P} - a.s.$$

The strong log-concave densities

- a3. $p(x, y) \propto e^{-\gamma(x, y)}$, where $x \mapsto \gamma(x, y) \geq 0$ is twice cont. diff. and strongly convex with

$$x^*(y) = \operatorname{argmin}_{x \in \mathbb{R}} \gamma(x, y) \in (-\infty, \infty)$$

and

$$\frac{\partial^2}{\partial x^2} \gamma(x, y) \geq \kappa > 0, \quad \forall x, y \in \mathbb{R}.$$

- a4. for some constant $C > 0$,

$$- \overline{\lim}_{n \rightarrow \infty} \frac{1}{n} \log L_n(X_{1:n}, Y_{1:n}) \leq C, \quad \mathbb{P} - a.s.$$

The strong log-concave densities

Example: linear HMM:

$$X_n = aX_{n-1} + v_n,$$

$$Y_n = bX_n + w_n,$$

where

- stable case $|a| < 1$, $b \neq 0$
- (v_n) and (w_n) are independent and i.i.d.

$$v_1 \sim f_v(x) \propto e^{-|x|^{2+\delta}}$$

$$w_1 \sim f_w(x) \propto e^{-x^2(1+c|x|^{\delta'})}$$

with constants $c \geq 0$ and $\delta, \delta' \geq 0$

Mortensen's MAP path estimation in continuous time

- X is the solution of SDE

$$dX_t = \mu(X_t)dt + \sigma(X_t)dB_t, \quad X_0 = x \in \mathbb{R}$$

and Y is the “white noise” observation

$$dY_t = h(X_t)dt + W_t, \quad t \geq 0$$

- The global MAP path estimator is $\hat{X}_{[0, T]}^T$ is the maximizer (over φ) of the conditional Onsager-Machlup functional (Zeitouni-Dembo, 1987)

$$J(\varphi_{[0, T]}; Y_{[0, T]}) := \lim_{\varepsilon \rightarrow 0} \varepsilon^2 \log \mathbb{P} \left(\|X_{[0, T]} - \varphi_{[0, T]}\|_\infty \leq \varepsilon \mid Y_{[0, T]} \right)$$

- Does the $\lim_{T \rightarrow \infty} \hat{X}_t^T$ exist \mathbb{P} -a.s. for each fixed $t > 0$?

Mortensen's MAP path estimation in continuous time

- X is the solution of SDE

$$dX_t = \mu(X_t)dt + \sigma(X_t)dB_t, \quad X_0 = x \in \mathbb{R}$$

and Y is the “white noise” observation

$$dY_t = h(X_t)dt + W_t, \quad t \geq 0$$

- The global MAP path estimator is $\hat{X}_{[0,T]}^T$ is the maximizer (over φ) of the conditional Onsager-Machlup functional (Zeitouni-Dembo, 1987)

$$J(\varphi_{[0,T]}; Y_{[0,T]}) := \lim_{\varepsilon \rightarrow 0} \varepsilon^2 \log \mathbb{P} \left(\|X_{[0,T]} - \varphi_{[0,T]}\|_\infty \leq \varepsilon \mid Y_{[0,T]} \right)$$

- Does the $\lim_{T \rightarrow \infty} \hat{X}_t^T$ exist \mathbb{P} -a.s. for each fixed $t > 0$?

Mortensen's MAP path estimation in continuous time

- X is the solution of SDE

$$dX_t = \mu(X_t)dt + \sigma(X_t)dB_t, \quad X_0 = x \in \mathbb{R}$$

and Y is the “white noise” observation

$$dY_t = h(X_t)dt + W_t, \quad t \geq 0$$

- The global MAP path estimator is $\hat{X}_{[0, T]}^T$ is the maximizer (over φ) of the conditional Onsager-Machlup functional (Zeitouni-Dembo, 1987)

$$J(\varphi_{[0, T]}; Y_{[0, T]}) := \lim_{\varepsilon \rightarrow 0} \varepsilon^2 \log \mathbb{P} \left(\|X_{[0, T]} - \varphi_{[0, T]}\|_\infty \leq \varepsilon \mid Y_{[0, T]} \right)$$

- Does the $\lim_{T \rightarrow \infty} \hat{X}_t^T$ exist \mathbb{P} -a.s. for each fixed $t > 0$?

Open problems

- is there a finite state HMM for which MAP path doesn't converge ?
- weaker/no convexity in continuous state models ?
- connections to the optimal control (turnpike, etc.) ?
- connections to the nonlinear filter stability ?
- convergence of the Mortensen-Dembo-Zeitouni MAP path estimator in continuous time diffusion models ?

Open problems

- is there a finite state HMM for which MAP path doesn't converge ?
- weaker/no convexity in continuous state models ?
- connections to the optimal control (turnpike, etc.) ?
- connections to the nonlinear filter stability ?
- convergence of the Mortensen-Dembo-Zeitouni MAP path estimator in continuous time diffusion models ?

Open problems

- is there a finite state HMM for which MAP path doesn't converge ?
- weaker/no convexity in continuous state models ?
- connections to the optimal control (turnpike, etc.) ?
- connections to the nonlinear filter stability ?
- convergence of the Mortensen-Dembo-Zeitouni MAP path estimator in continuous time diffusion models ?

Open problems

- is there a finite state HMM for which MAP path doesn't converge ?
- weaker/no convexity in continuous state models ?
- connections to the optimal control (turnpike, etc.) ?
- connections to the nonlinear filter stability ?
- convergence of the Mortensen-Dembo-Zeitouni MAP path estimator in continuous time diffusion models ?

Open problems

- is there a finite state HMM for which MAP path doesn't converge ?
- weaker/no convexity in continuous state models ?
- connections to the optimal control (turnpike, etc.) ?
- connections to the nonlinear filter stability ?
- convergence of the Mortensen-Dembo-Zeitouni MAP path estimator in continuous time diffusion models ?