

State sequence prediction in imprecise hidden Markov models

Jasper De Bock & Gert de Cooman

27 July 2011



Jasper De Bock & Gert de Cooman



Jasper De Bock & Gert de Cooman



Research group
SYSTeMS



Jasper De Bock



Gert de Cooman



Research group
SYSTeMS



Jasper De Bock



Gert de Cooman



Filip Hermans



Erik Quaeghebeur



Keivan Shariatmadar



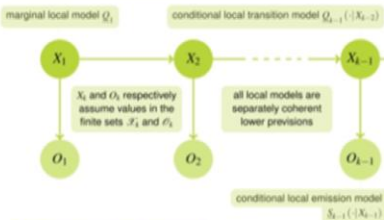
Arthur Van Camp

State sequence prediction in imprecise hidden Markov models

Interpretation of the graphical model

Interpretation of the graphical structure Our imprecise hidden Markov model (iHMM) represents the following irrelevance assessments: conditional on its mother variable, the non-parent non-descendants of any variable in the tree are epistemically irrelevant to this variable and its descendants.

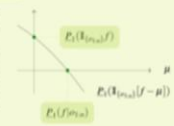
Epistemic irrelevance Y is irrelevant to X whenever the belief model (lower prevision E) does not change when we learn something about Y :
 $(\forall x \in \mathcal{X}(\mathcal{Y}))(\forall y \in \mathcal{Y}) E(x) = E(x|y)$.



Maximal state sequences

Conditioning the joint model

Since we assume that all local lower probabilities are strictly positive (in recent work, we dropped this assumption), $P_k(\{x_{k-1}\}) > 0$ and the Generalised Bayes Rule yields a uniquely coherent value of $P_k(f|y_{k-1})$, which has (this is very useful) the same sign as $P_k(\{x_{k-1}\})$.



Optimality criterion

We can express a strict preference \succ between state sequences $x_{1:n}$ and $x'_{1:n}$ as follows: $x_{1:n} \succ x'_{1:n} \Leftrightarrow E_k(\{x_{1:n}\} - \{x'_{1:n}\} | y_{1:n}) > 0$. This induces a strict partial order \succ on the set of state sequences $\mathcal{X}_{1:n}$, and we consider a sequence $x_{1:n}$ to be optimal when it is undominated, or maximal, in this strict partial order:

$$x_{1:n} \in \text{opt}(\mathcal{X}_{1:n} | y_{1:n}) \Leftrightarrow (\forall x_{1:n} \in \mathcal{X}_{1:n}) x_{1:n} \not\succeq x_{1:n}$$

$$\Leftrightarrow (\forall x_{1:n} \in \mathcal{X}_{1:n}) E_k(\{x_{1:n}\} - \{x_{1:n}\} | y_{1:n}) \leq 0$$

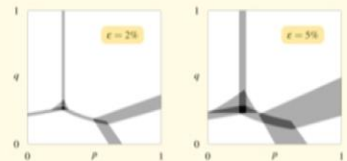
$$\Leftrightarrow (\forall x_{1:n} \in \mathcal{X}_{1:n}) E_k(\{x_{1:n}\} | y_{1:n}) \leq 0$$

In an analogous manner we define the optimal subsequences:

$$x_{k:n} \in \text{opt}(\mathcal{X}_{k:n} | y_{k:n}) \Leftrightarrow (\forall x_{k:n} \in \mathcal{X}_{k:n}) E_k(\{x_{k:n}\} - \{x_{k:n}\} | y_{k:n}) \leq 0$$

A first experiment

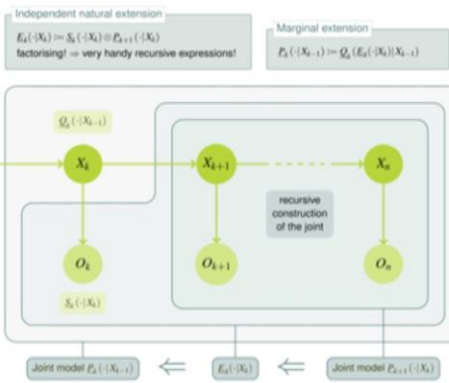
Motivation and description No algorithm, however cleverly designed, will be able to find all maximal sequences efficiently if there are too many. Because this number of maximal sequences is so important, we study its behaviour in more detail. We consider a binary, stationary iHMM with precise emission models. The imprecise marginal and transition models are generated by mixing precise models with a vacuous one, using a mixture coefficient ϵ . For a particular observation sequence of length three, we plot the number of maximal sequences as a function of the transition probabilities ρ and q . As this number grows from 1 to 4 the areas go from white to black.



Results We see that there are large regions of transition probability space where the number of maximal elements remains fairly small. The plots also display quite interesting behaviour. If we let the imprecision grow, by using higher ϵ , the areas with multiple maximal sequences become larger. They are expanded versions of the lines of indifference that occur in the precise case.

Jasper De Bock and Gert de Cooman

{Jasper.DeBock, Gert.deCooman}@UGent.be
 SYSTEms, Ghent University, Belgium



Principle of Optimality Using the recursive expressions for the joint model, we can derive an appropriate version of Bellman's Principle of Optimality:
 $x_{k:n} \in \text{opt}(\mathcal{X}_{k:n} | y_{k:n}) \Leftrightarrow x_{k+1:n} \in \text{opt}(\mathcal{X}_{k+1:n} | y_{k+1:n})$, which in turn implies that
 $\text{opt}(\mathcal{X}_{k:n} | y_{k:n}) \subseteq \bigcup_{x_{k+1:n} \in \text{opt}(\mathcal{X}_{k+1:n} | y_{k+1:n})} x_{k+1:n}$.

Backward-forward recursion We let k run backward from n to 1. For each k and every $x_{k+1:n} \in \mathcal{X}_{k+1:n}$, we determine $\text{opt}(\mathcal{X}_{k:n} | y_{k:n})$ with our alternative optimality criterion. We prove that we can do this efficiently by executing the forward running procedure demonstrated in the figure below. The sequences of μ -maximisers are the elements of $\text{opt}(\mathcal{X}_{k:n} | y_{k:n})$. In this binary example, $\mathcal{X} = \{0, 1\}$.

An alternative criterion If we limit ourselves to the possible sequences selected by the Principle of Optimality, we can reformulate the optimality criterion as follows:
 $x_{k:n} \in \text{opt}(\mathcal{X}_{k:n} | y_{k:n}) \Leftrightarrow \alpha_k^{\text{opt}}(x_{k:n}) \leq \alpha_k(x_{k:n})$.

repeat for l from k to n until **no**:
 $\alpha_k^{\text{opt}}(x_{k:n}) \geq \alpha_k^{\text{opt}}(x_{k:n-1})$?
 $\alpha_k^{\text{opt}}(x_{k:n}) = \max_{x_{k+1:n} \in \text{opt}(\mathcal{X}_{k+1:n} | y_{k+1:n})} \alpha_k(x_{k:n})$

Can be calculated efficiently by dynamical programming.

$\bar{x}_k(\{y_k\}) = \prod_{i=1}^k \bar{x}_i(\{y_i\}) \bar{Q}_i(\{k\} | y_{i-1})$

$\text{opt}(\mathcal{X}_{k+1:n} | y_{k+1:n})$

$\text{opt}(\mathcal{X}_{k:n} | y_{k:n})$

The EstiHMM algorithm

Computational complexity

Theoretical analysis We prove that the computational complexity is at worst quadratic in the length of the Markov chain, cubic in the number of states, and roughly speaking linear in the number of maximal sequences (each backward step in the backward-forward loop has a linear complexity in the number of maximal elements at that stage).

Empirical confirmation In order to demonstrate that our algorithm is indeed quite efficient, we let it determine the maximal sequences for a random output sequence of length 100. The iHMM we use to determine the maximal sequences is generated by mixing precise local models with a vacuous one, using a mixture coefficient ϵ . For $\epsilon = 2\%$, there are 5 maximal sequences and it takes 0.2 seconds to calculate them. If we let ϵ grow to 5%, we get 764 maximal sequences and these can be determined in 32 seconds. This demonstrates that the complexity is indeed linear in the number of solutions and that the algorithm can efficiently calculate the maximal sequences even for long output sequences.

State sequence prediction in imprecise hidden Markov models

State sequence prediction in imprecise hidden Markov models

Interpretation of the graphical model

Interpretation of the graphical structure Our imprecise hidden Markov model (iHMM) represents the following irrelevance assessments: conditional on its mother variable, the non-parent non-descendants of any variable in the tree are epistemically irrelevant to this variable and its descendants.

Epistemic irrelevance Y is irrelevant to X whenever the belief model (lower prevision E) does not change when we learn something about Y :
 $(\forall x \in \mathcal{X}(\mathcal{Y}))(\forall y \in \mathcal{Y}) E(x) = E(x|y)$.

Jasper De Bock and Gert de Cooman
 {Jasper.DeBock, Gert.deCooman}@UGent.be
 SYSTEms, Ghent University, Belgium



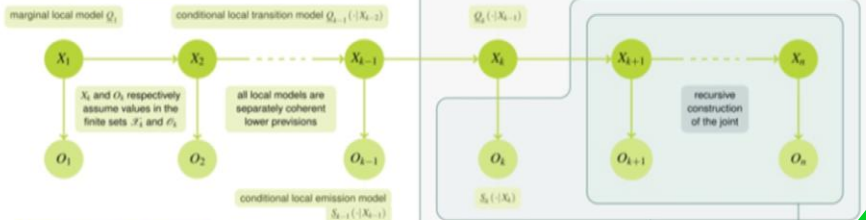
Independent natural extension

$$E_k(\cdot|X_k) = \bar{S}_k(\cdot|X_k) \oplus \underline{P}_{k+1}(\cdot|X_k)$$

factoring! \Rightarrow very handy recursive expressions!

Marginal extension

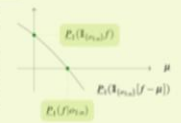
$$E_k(\cdot|X_{k-1}) = \underline{Q}_k(\bar{E}_k(\cdot|X_k)|X_{k-1})$$



Maximal state sequences

Conditioning the joint model

Since we assume that all local lower probabilities are strictly positive (in recent work, we dropped this assumption), $\underline{P}_k(\{x_{k+1}\}) > 0$ and the Generalised Bayes Rule yields a uniquely coherent value of $\underline{P}_k(f|f_{k+1})$, which has (this is very useful) the same sign as $\underline{P}_k(\{x_{k+1}\})$.



Optimality criterion

We can express a strict preference \succ between state sequences $x_{1:n}$ and $x_{1:n}$ as follows: $x_{1:n} \succ x_{1:n} \Leftrightarrow \underline{P}_k(\{x_{k+1}\} - \{x_{k+1}\})_{x_{1:n}} > 0$. This induces a strict partial order \succ on the set of state sequences $\mathcal{X}_{1:n}$, and we consider a sequence $x_{1:n}$ to be optimal when it is undominated, or maximal, in this strict partial order:

$$x_{1:n} \in \text{opt}(\mathcal{X}_{1:n}|o_{1:n}) \Leftrightarrow (\forall x_{1:n} \in \mathcal{X}_{1:n}) x_{1:n} \not\succeq x_{1:n}$$

$$\Leftrightarrow (\forall x_{1:n} \in \mathcal{X}_{1:n}) \underline{P}_k(\{x_{k+1}\} - \{x_{k+1}\})_{x_{1:n}} \leq 0$$

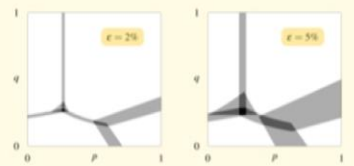
$$\Leftrightarrow (\forall x_{1:n} \in \mathcal{X}_{1:n}) \underline{P}_k(\{x_{k+1}\} - \{x_{k+1}\})_{x_{1:n}} \leq 0$$

In an analogous manner we define the optimal subsequences:

$$x_{k:n} \in \text{opt}(\mathcal{X}_{k:n}|x_{1:k-1}, o_{k:n}) \Leftrightarrow (\forall x_{k:n} \in \mathcal{X}_{k:n}) \underline{P}_k(\{x_{k+1}\} - \{x_{k+1}\})_{x_{k:n}} \leq 0$$

A first experiment

Motivation and description No algorithm, however cleverly designed, will be able to find all maximal sequences efficiently if there are too many. Because this number of maximal sequences is so important, we study its behaviour in more detail. We consider a binary, stationary iHMM with precise emission models. The imprecise marginal and transition models are generated by mixing precise models with a vacuous one, using a mixture coefficient ϵ . For a particular observation sequence of length three, we plot the number of maximal sequences as a function of the transition probabilities ρ and q . As this number grows from 1 to 4 the areas go from white to black.



Results We see that there are large regions of transition probability space where the number of maximal elements remains fairly small. The plots also display quite interesting behaviour. If we let the imprecision grow, by using higher ϵ , the areas with multiple maximal sequences become larger. They are expanded versions of the lines of indifference that occur in the precise case.



Principle of Optimality

Using the recursive expressions for the joint model, we can derive an appropriate version of Bellman's Principle of Optimality:

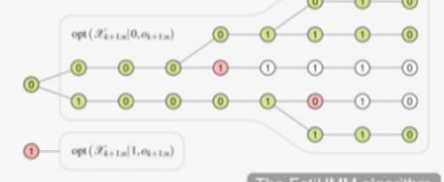
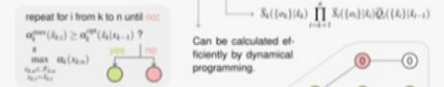
$$x_{k:n} \in \text{opt}(\mathcal{X}_{k:n}|x_{1:k-1}, o_{k:n}) \Leftrightarrow x_{k:n} \in \text{opt}(\mathcal{X}_{k+1:n}|x_{k+1}, o_{k+1:n}),$$

which in turn implies that

$$\text{opt}(\mathcal{X}_{k+1:n}|x_{k+1}, o_{k+1:n}) \subseteq \bigcup_{x_k \in \mathcal{X}_k} \text{opt}(\mathcal{X}_{k+1:n}|x_k, o_{k+1:n}).$$

Alternative criterion

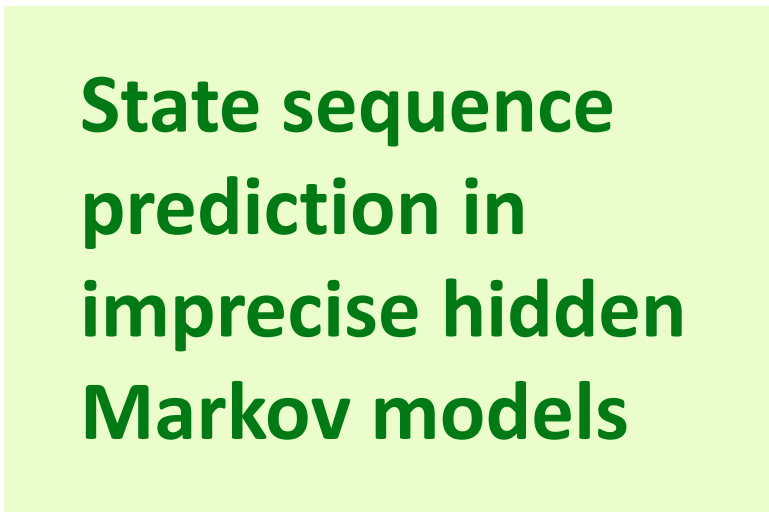
If we limit ourselves to the possible sequences selected by the Principle of Optimality, we can reformulate the optimality criterion as follows:

$$x_{k:n} \in \text{opt}(\mathcal{X}_{k:n}|x_{1:k-1}, o_{k:n}) \Leftrightarrow \alpha_k^{\text{opt}}(x_k|x_{1:k-1}) \leq \alpha_k(x_{k:n}).$$


Computational complexity

Theoretical analysis We prove that the computational complexity is at worst quadratic in the length of the Markov chain, cubic in the number of states, and roughly speaking linear in the number of maximal sequences (each backward step in the backward-forward loop has a linear complexity in the number of maximal elements at that stage).

Empirical confirmation In order to demonstrate that our algorithm is indeed quite efficient, we let it determine the maximal sequences for a random output sequence of length 100. The iHMM we use to determine the maximal sequences is generated by mixing precise local models with a vacuous one, using a mixture coefficient ϵ . For $\epsilon = 2\%$, there are 5 maximal sequences and it takes 0.2 seconds to calculate them. If we let ϵ grow to 5%, we get 764 maximal sequences and these can be determined in 32 seconds. This demonstrates that the complexity is indeed linear in the number of solutions and that the algorithm can efficiently calculate the maximal sequences even for long output sequences.



The imprecise hidden Markov model

Imprecise hidden Markov model

A sequence of hidden state variables

X_1

X_2

X_3

O_1

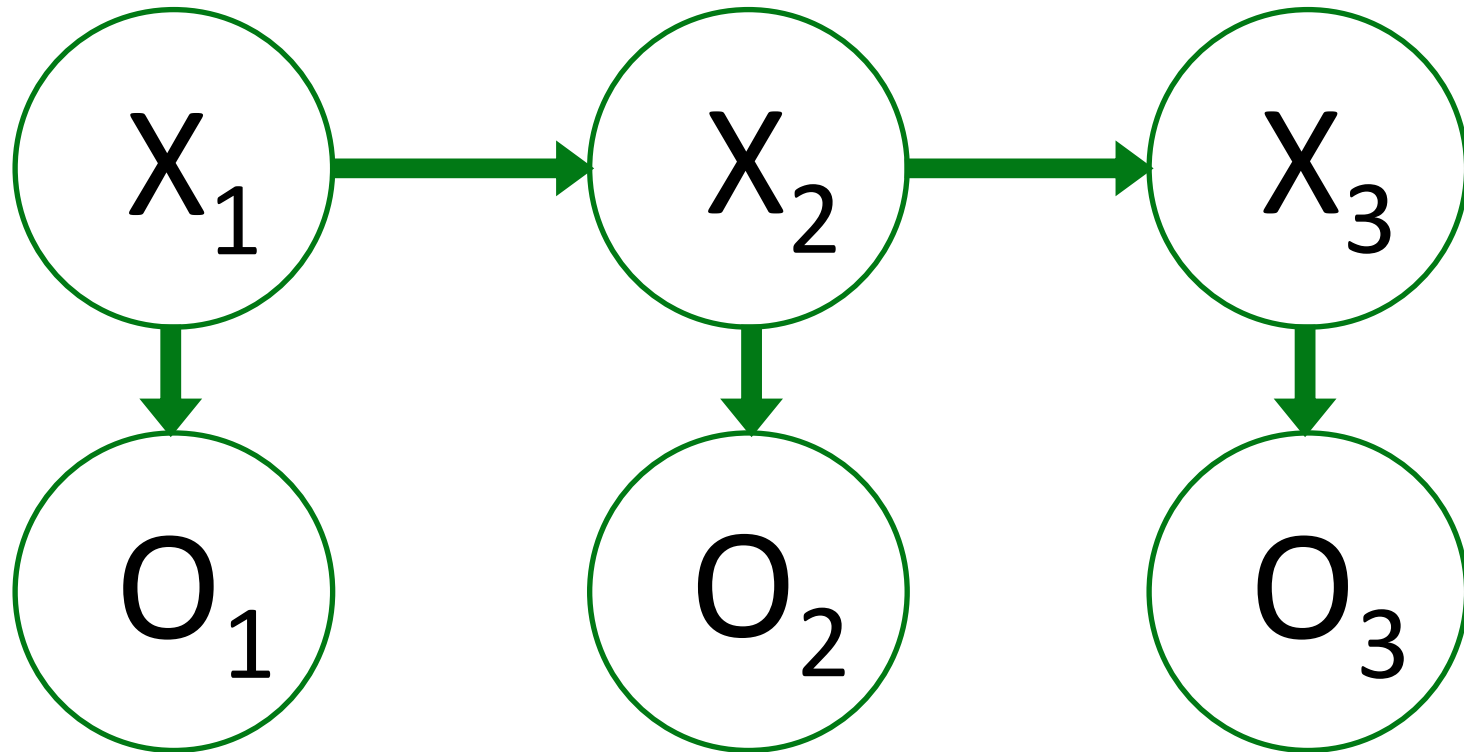
O_2

O_3

A sequence of observable variables

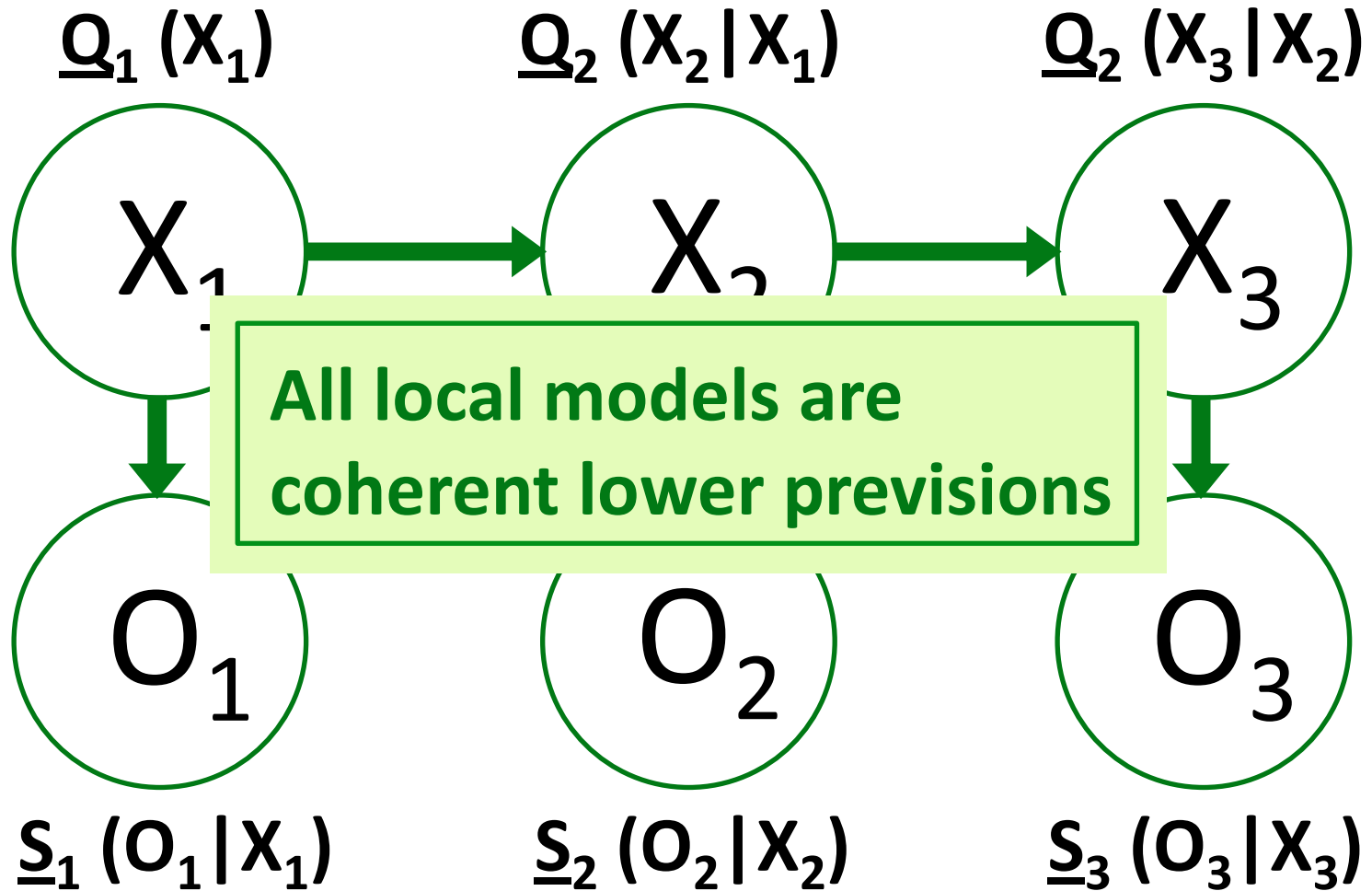
Imprecise hidden Markov model

A sequence of hidden state variables

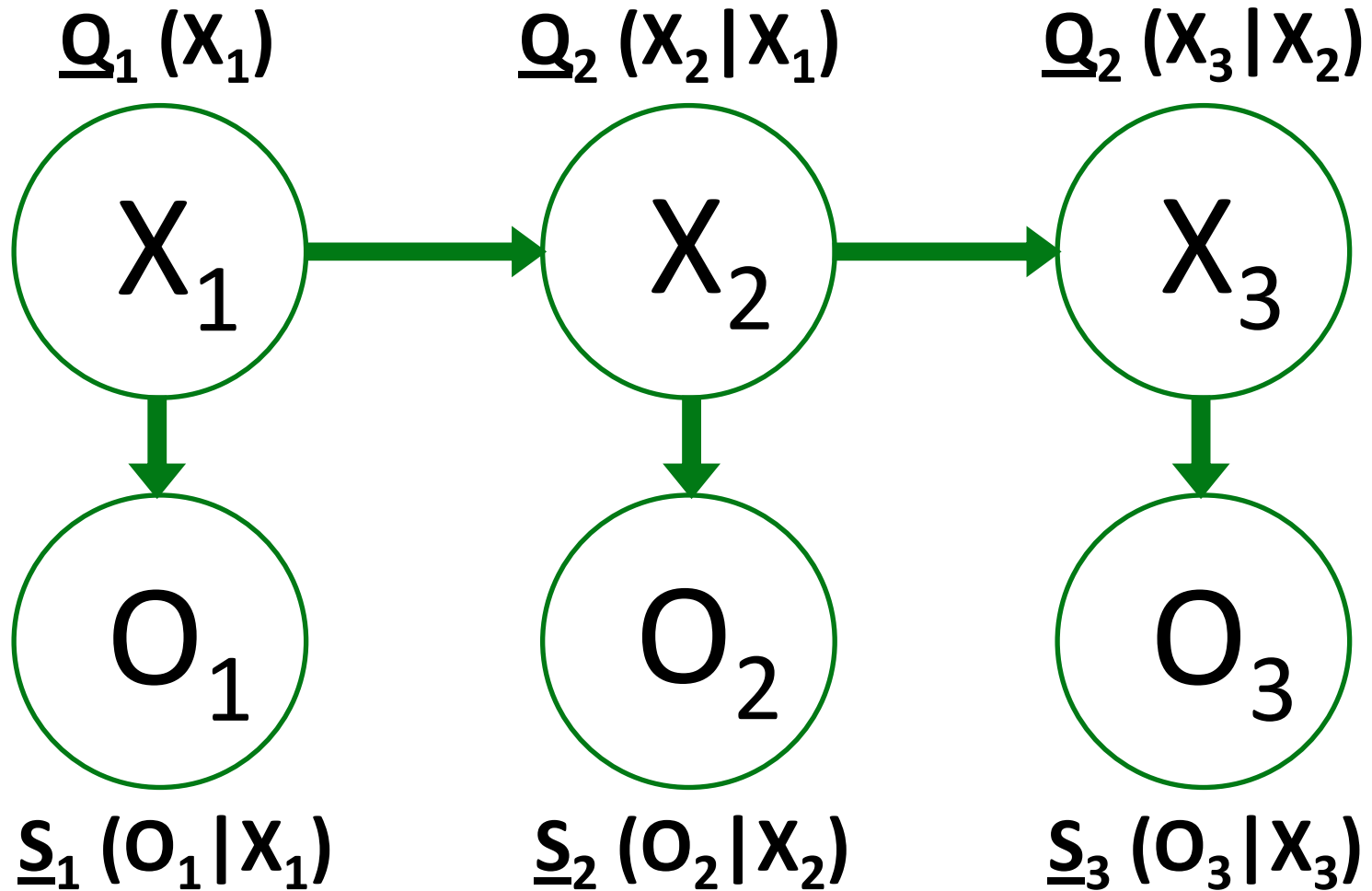


A sequence of observable variables

Imprecise hidden Markov model



Imprecise hidden Markov model

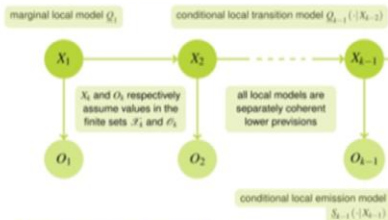


State sequence prediction in imprecise hidden Markov models

Interpretation of the graphical model

Interpretation of the graphical structure Our imprecise hidden Markov model (iHMM) represents the following irrelevance assessments: conditional on its mother variable, the non-parent non-descendants of any variable in the tree are epistemically irrelevant to this variable and its descendants.

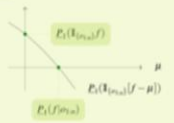
Epistemic irrelevance Y is irrelevant to X whenever the belief model (lower prevision E) about X does not change when we learn something about Y :
 $(\forall x \in \mathcal{X}(\mathcal{Y}))(\forall y \in \mathcal{Y}) E(x) = E(x|y)$.



Maximal state sequences

Conditioning the joint model

Since we assume that all local lower probabilities are strictly positive (in recent work, we dropped this assumption), $P_i(\{o_{i+1}\}) > 0$ and the Generalised Bayes Rule yields a uniquely coherent value of $P_i(f|o_{i+1})$, which has (this is very useful) the same sign as $P_i(\{i_{i+1}|f\})$.



Optimality criterion

We can express a strict preference \succ between state sequences $i_{1:n}$ and $i_{1:n}'$ as follows: $i_{1:n} \succ i_{1:n}' \Leftrightarrow P_i(\{i_{i+1} - i_{i+1}'\} | i_{1:n}) > 0$. This induces a strict partial order \succ on the set of state sequences $\mathcal{I}_{1:n}$, and we consider a sequence $i_{1:n}$ to be optimal when it is undominated, or maximal, in this strict partial order:

$$i_{1:n} \in \text{opt}(\mathcal{I}_{1:n} | o_{1:n}) \Leftrightarrow (\forall i_{1:n}' \in \mathcal{I}_{1:n}) i_{1:n} \not\succeq i_{1:n}'$$

$$\Leftrightarrow (\forall i_{1:n}' \in \mathcal{I}_{1:n}) P_i(\{i_{i+1} - i_{i+1}'\} | i_{1:n}) \leq 0$$

$$\Leftrightarrow (\forall i_{1:n}' \in \mathcal{I}_{1:n}) P_i(\{i_{i+1} - i_{i+1}'\} | i_{1:n}) \leq 0$$

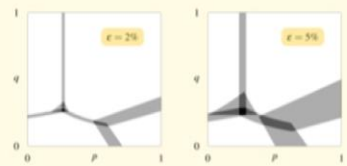
In an analogous manner we define the optimal subsequences:

$$i_{k:n} \in \text{opt}(\mathcal{I}_{k:n} | o_{k:n}) \Leftrightarrow (\forall i_{k:n}' \in \mathcal{I}_{k:n}) P_i(\{i_{i+1} - i_{i+1}'\} | i_{k:n}) \leq 0$$

A first experiment

Motivation and description

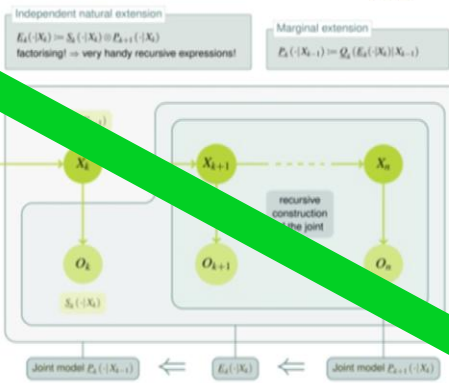
No algorithm, however cleverly designed, will be able to find all maximal sequences efficiently if there are too many. Because this number of maximal sequences is so important, we study its behaviour in more detail. We consider a binary, stationary iHMM with precise emission models. The imprecise marginal and transition models are generated by mixing precise models with a vacuous one, using a mixture coefficient ϵ . For a particular observation sequence of length three, we plot the number of maximal sequences as a function of the transition probabilities ρ and q . As this number grows from 1 to 4 the areas go from white to black.



Results We see that there are large regions of transition probability space where the number of maximal elements remains fairly small. The plots also display quite interesting behaviour. If we let the imprecision grow, by using higher ϵ , the areas with multiple maximal sequences become larger. They are expanded versions of the lines of indifference that occur in the precise case.

Jasper De Bock and Gert de Cooman

{Jasper.DeBock, Gert.deCooman}@UGent.be
 SYSTEms, Ghent University, Belgium



Principle of Optimality Using the recursive expressions for the joint model, we can derive an appropriate version of Bellman's Principle of Optimality:
 $i_{k:n} \in \text{opt}(\mathcal{I}_{k:n} | o_{k:n}) \Leftrightarrow i_{k+1:n} \in \text{opt}(\mathcal{I}_{k+1:n} | o_{k+1:n})$, which in turn implies that
 $\text{opt}(\mathcal{I}_{k:n} | o_{k:n}) \subseteq \bigcup_{o_{k+1}} \text{opt}(\mathcal{I}_{k+1:n} | o_{k+1:n})$.

Backward-forward recursion We let k run backward from n to 1. For each k and every $i_{k+1:n} \in \mathcal{I}_{k+1:n}$, we determine $\text{opt}(\mathcal{I}_{k:n} | o_{k:n})$ with our alternative optimality criterion. We prove that we can do this efficiently by executing the forward running procedure demonstrated in the figure below. The sequences of $\text{opt}(\mathcal{I}_{k:n} | o_{k:n})$ are the elements of $\text{opt}(\mathcal{I}_{1:n} | o_{1:n})$. In this binary example, $\mathcal{X} = \{0, 1\}$.

An alternative criterion If we limit ourselves to the possible sequences selected by the Principle of Optimality, we can reformulate the optimality criterion as follows:
 $i_{k:n} \in \text{opt}(\mathcal{I}_{k:n} | o_{k:n}) \Leftrightarrow \alpha_k^{\text{opt}}(i_k | i_{k+1:n}) \leq \alpha_k(i_k)$.

repeat for l from k to n until **exit**
 $\alpha_k^{\text{opt}}(i_k) \geq \alpha_k^{\text{opt}}(i_k | i_{k+1:n})$?
 $\max_{o_{k+1}} \alpha_k(o_{k+1})$?
 Can be calculated efficiently by dynamical programming.

The EstiHMM algorithm

Computational complexity

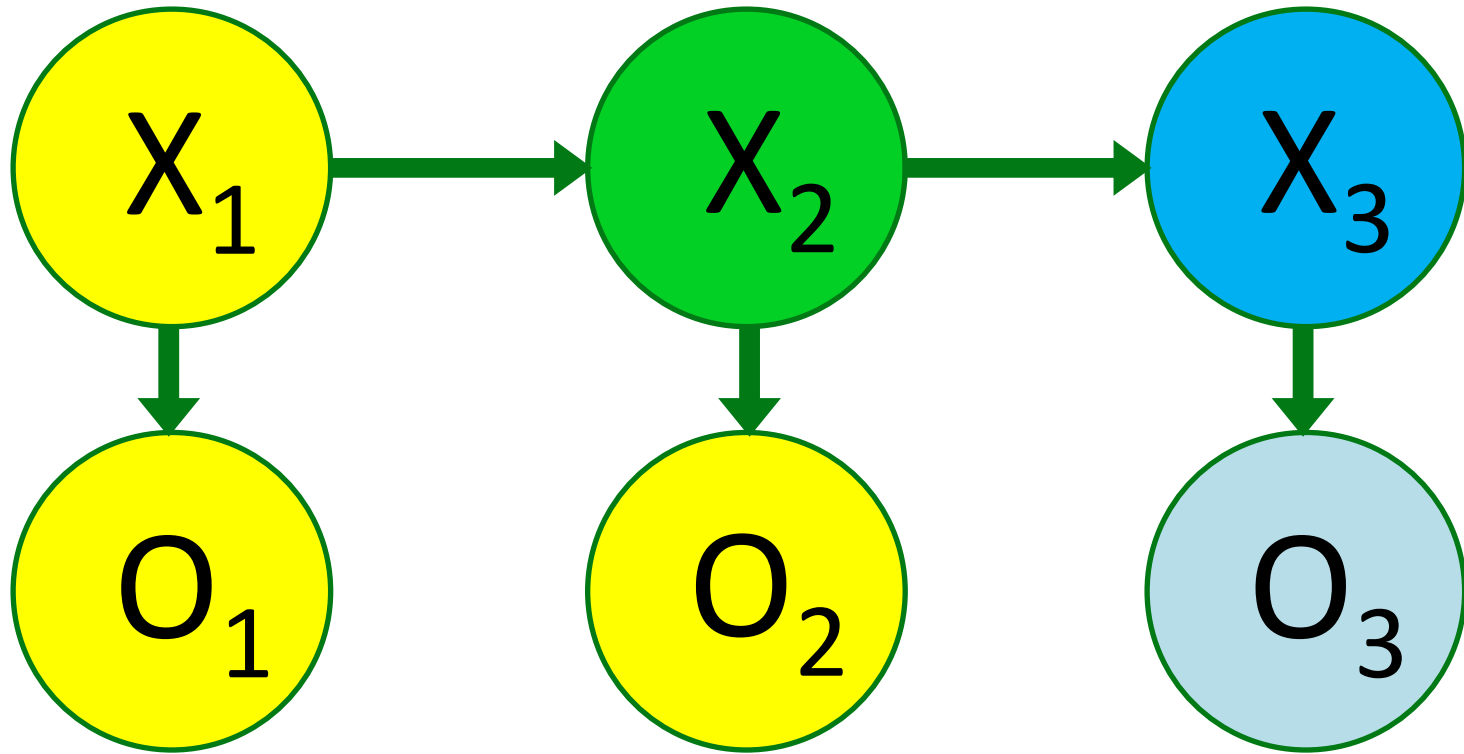
Theoretical analysis We prove that the computational complexity is at worst quadratic in the length of the Markov chain, cubic in the number of states, and roughly speaking linear in the number of maximal sequences (each backward step in the backward-forward loop has a linear complexity in the number of maximal elements at that stage).

Empirical confirmation In order to demonstrate that our algorithm is indeed quite efficient, we let it determine the maximal sequences for a random output sequence of length 100. The iHMM we use to determine the maximal sequences is generated by mixing precise local models with a vacuous one, using a mixture coefficient ϵ . For $\epsilon = 2\%$, there are 5 maximal sequences and it takes 0.2 seconds to calculate them. If we let ϵ grow to 5%, we get 764 maximal sequences and these can be determined in 32 seconds. This demonstrates that the complexity is indeed linear in the number of solutions and that the algorithm can efficiently calculate the maximal sequences even for long output sequences.

State sequence prediction in imprecise hidden Markov models

Epistemic Irrelevance

Epistemic irrelevance



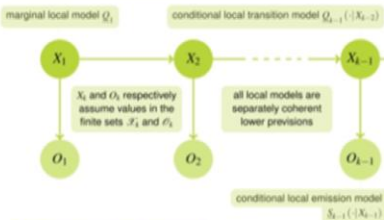
Conditional on its **mother variable**, the **non-parent non-descendants** of any **variable** in the tree are epistemically irrelevant to this **variable** and its **descendants**

State sequence prediction in imprecise hidden Markov models

Interpretation of the graphical model

Interpretation of the graphical structure Our imprecise hidden Markov model (iHMM) represents the following irrelevance assessments: conditional on its mother variable, the non-parent non-descendants of any variable in the tree are epistemically irrelevant to this variable and its descendants.

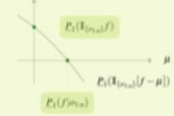
Epistemic irrelevance Y is irrelevant to X whenever the belief model (lower prevision E) about X does not change when we learn something about Y :
 $(\forall x \in \mathcal{X}(\mathcal{Y}))(\forall y \in \mathcal{Y}) E(x) = E(x|y)$.



Maximal state sequences

Conditioning the joint model

Since we assume that all local lower probabilities are strictly positive (in recent work, we dropped this assumption), $P_i(\{x_{i+1}\}) > 0$ and the Generalised Bayes Rule yields a uniquely coherent value of $P_i(f|y_{i+1})$, which has (this is very useful) the same sign as $P_i(\{x_{i+1}\})$.



Optimality criterion

We can express a strict preference \succ between state sequences $x_{1:n}$ and $x'_{1:n}$ as follows: $x_{1:n} \succ x'_{1:n} \Leftrightarrow E_i(\{x_{i+1}\} | x_{1:n}) > E_i(\{x_{i+1}\} | x'_{1:n})$. This induces a strict partial order \succ on the set of state sequences $\mathcal{X}_{1:n}$, and we consider a sequence $x_{1:n}$ to be optimal when it is undominated, or maximal, in this strict partial order:

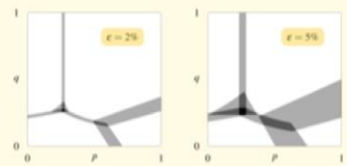
$$x_{1:n} \in \text{opt}(\mathcal{X}_{1:n}, v_{1:n}) \Leftrightarrow (\forall x'_{1:n} \in \mathcal{X}_{1:n} \setminus \{x_{1:n}\}) \\ (\forall v_{1:n} \in \mathcal{V}_{1:n}) E_i(\{x_{i+1}\} | x_{1:n}) > E_i(\{x_{i+1}\} | x'_{1:n}) \vee \\ (\forall v_{1:n} \in \mathcal{V}_{1:n}) E_i(\{x_{i+1}\} | x_{1:n}) = E_i(\{x_{i+1}\} | x'_{1:n}) \wedge v_{1:n}(x_{1:n}) > v_{1:n}(x'_{1:n})$$

In an analogous manner we define the optimal subsequences:

$$x_{k:n} \in \text{opt}(\mathcal{X}_{k:n}, v_{k:n}) \Leftrightarrow (\forall x'_{k:n} \in \mathcal{X}_{k:n}) E_i(\{x_{i+1}\} | x_{k:n}) > E_i(\{x_{i+1}\} | x'_{k:n}) \vee \\ (\forall x'_{k:n} \in \mathcal{X}_{k:n}) E_i(\{x_{i+1}\} | x_{k:n}) = E_i(\{x_{i+1}\} | x'_{k:n}) \wedge v_{k:n}(x_{k:n}) > v_{k:n}(x'_{k:n})$$

A first experiment

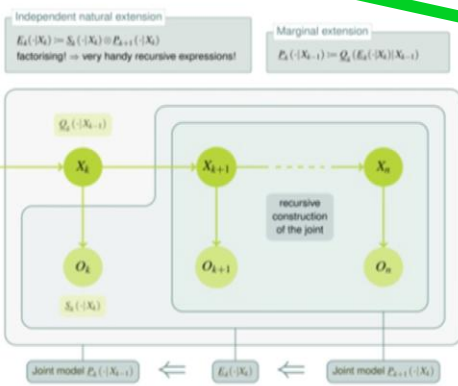
Motivation and description No algorithm, however cleverly designed, will be able to find all maximal sequences efficiently if there are too many. Because this number of maximal sequences is so important, we study its behaviour in more detail. We consider a binary, stationary iHMM with precise emission models. The imprecise marginal and transition models are generated by mixing precise models with a vacuous one, using a mixture coefficient ϵ . For a particular observation sequence of length three, we plot the number of maximal sequences as a function of the transition probabilities p and q . As this number grows from 1 to 4 the areas go from white to black.



Results We see that there are large regions of transition probability space where the number of maximal elements remains fairly small. The plots also display quite interesting behaviour. If we let the imprecision grow, by using higher ϵ , the areas with multiple maximal sequences become larger. They are expanded versions of the lines of indifference that occur in the precise case.

Jasper De Bock and Gert de Cooman

{Jasper.DeBock, Gert.deCooman}@UGent.be
 SYSTeMS, Ghent University, Belgium

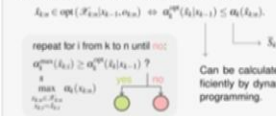


Principle of Optimality

Using the recursive expressions for the joint model, we can derive an appropriate version of Bellman's Principle of Optimality:
 $x_{k:n} \in \text{opt}(\mathcal{X}_{k:n}, v_{k:n}) \Leftrightarrow x_{k+1:n} \in \text{opt}(\mathcal{X}_{k+1:n}, v_{k+1:n})$, which in turn implies that

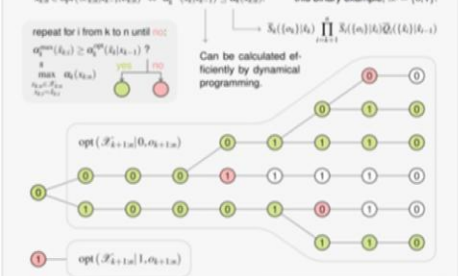
$$\text{opt}(\mathcal{X}_{k+1:n}, v_{k+1:n}) \subseteq \bigcup_{x_k \in \mathcal{X}_k} \text{opt}(\mathcal{X}_{k+1:n}, v_{k+1:n} | x_k)$$

An alternative criterion If we limit ourselves to the possible sequences selected by the Principle of Optimality, we can reformulate the optimality criterion as follows:
 $x_{k:n} \in \text{opt}(\mathcal{X}_{k:n}, v_{k:n}) \Leftrightarrow \alpha_k^{\text{opt}}(x_k | v_{k:n}) \leq \alpha_k(x_k | v_{k:n})$.



Backward-forward recursion

We let k run backward from n to 1 . For each k and every $x_{k+1:n} \in \mathcal{X}_{k+1:n}$, we determine $\text{opt}(\mathcal{X}_{k+1:n}, v_{k+1:n} | x_{k+1:n})$ with our alternative optimality criterion. We prove that we can do this efficiently by executing the forward running procedure demonstrated in the figure below. The sequences of $\text{opt}(\mathcal{X}_{k+1:n}, v_{k+1:n} | x_{k+1:n})$ are the elements of $\text{opt}(\mathcal{X}_{k+1:n}, v_{k+1:n})$. In this binary example, $\mathcal{X} = \{0, 1\}$.



The EstiHMM algorithm

Computational complexity

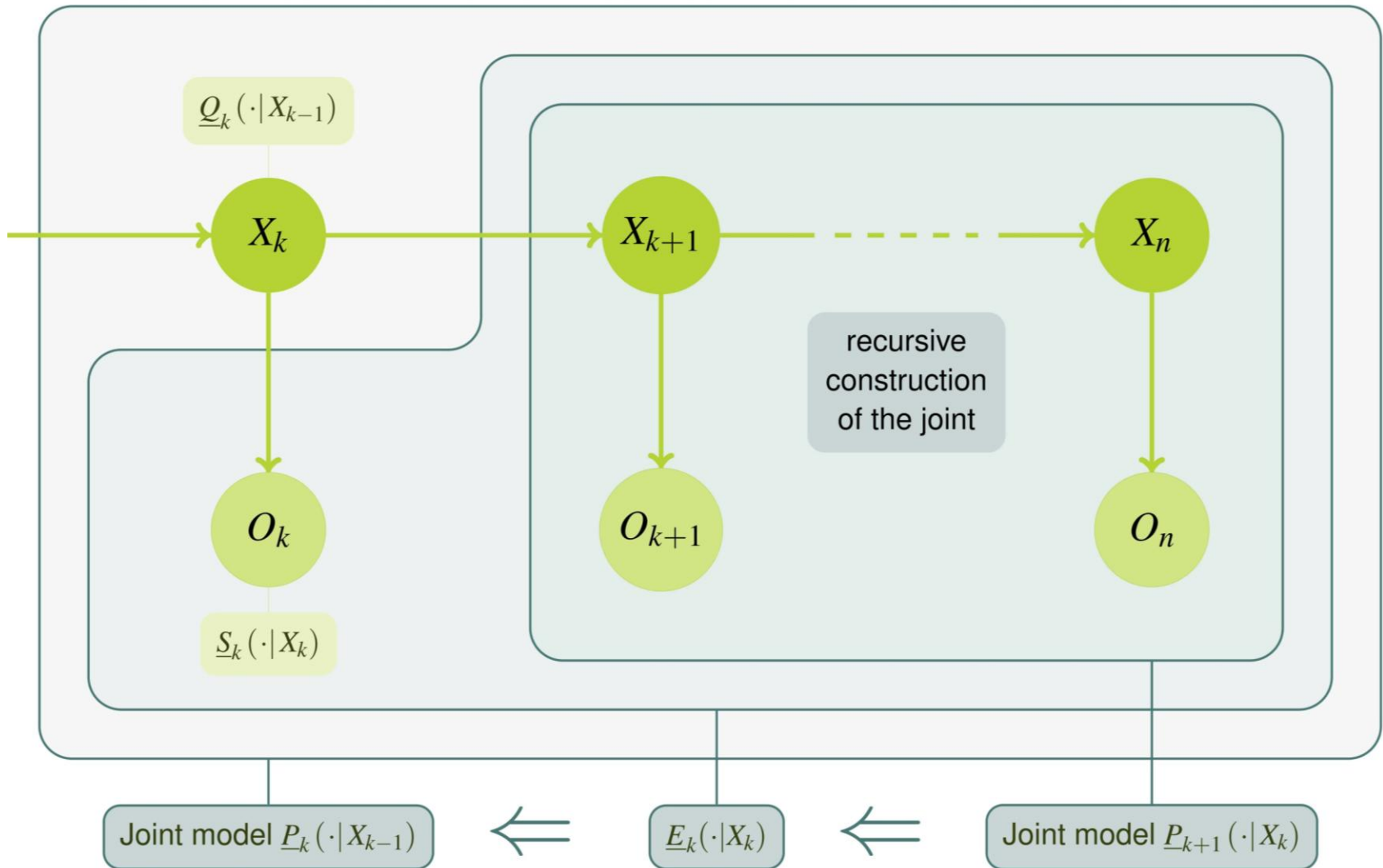
Theoretical analysis We prove that the computational complexity is at worst quadratic in the length of the Markov chain, cubic in the number of states, and roughly speaking linear in the number of maximal sequences (each backward step in the backward-forward loop has a linear complexity in the number of maximal elements at that stage).

Empirical confirmation In order to demonstrate that our algorithm is indeed quite efficient, we let it determine the maximal sequences for a random output sequence of length 100. The iHMM we use to determine the maximal sequences is generated by mixing precise local models with a vacuous one, using a mixture coefficient ϵ . For $\epsilon = 2\%$, there are 5 maximal sequences and it takes 0.2 seconds to calculate them. If we let ϵ grow to 5%, we get 764 maximal sequences and these can be determined in 32 seconds. This demonstrates that the complexity is indeed linear in the number of solutions and that the algorithm can efficiently calculate the maximal sequences even for long output sequences.

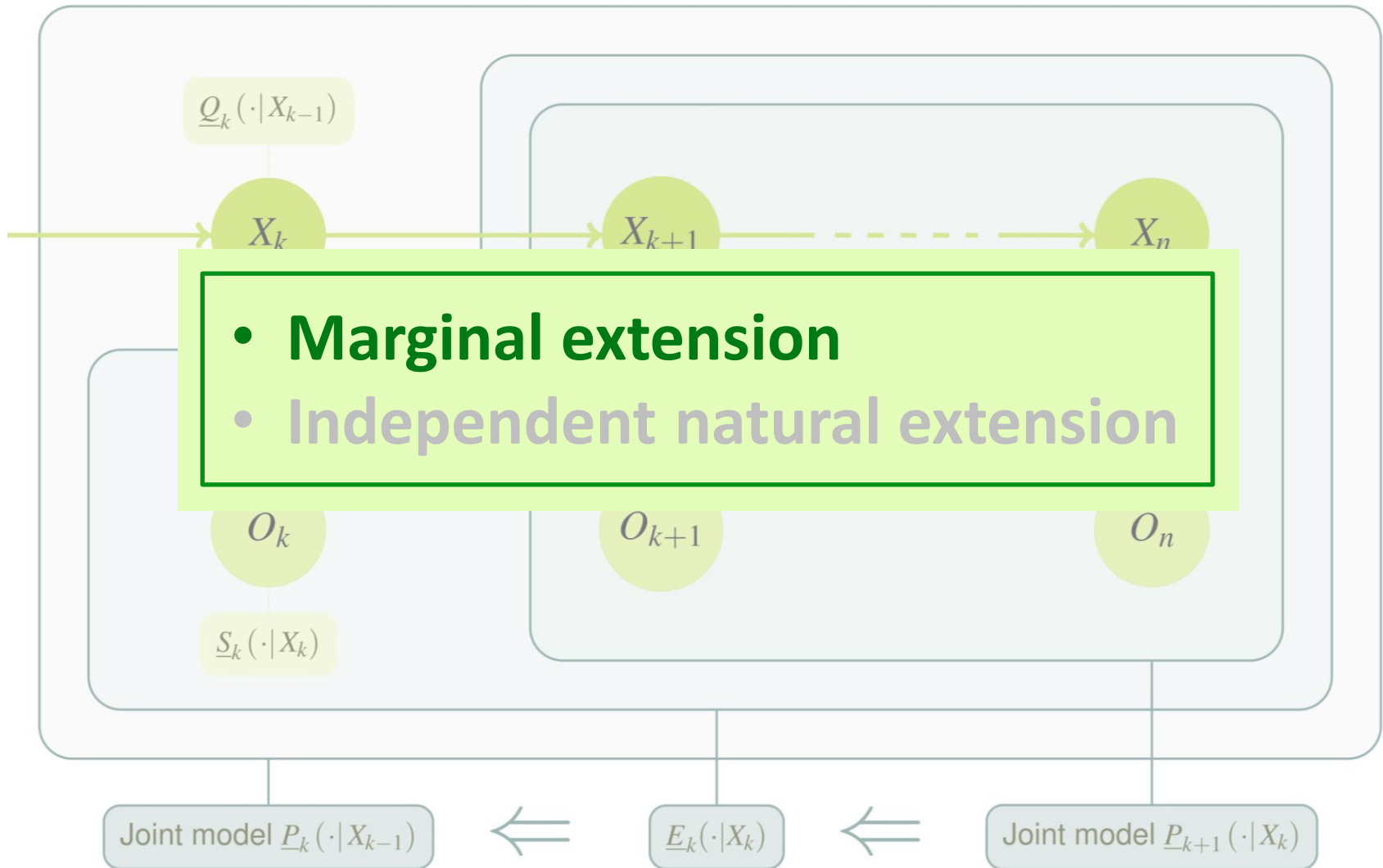
State sequence prediction in imprecise hidden Markov models

Recursive construction of a joint model for the imprecise hidden Markov model

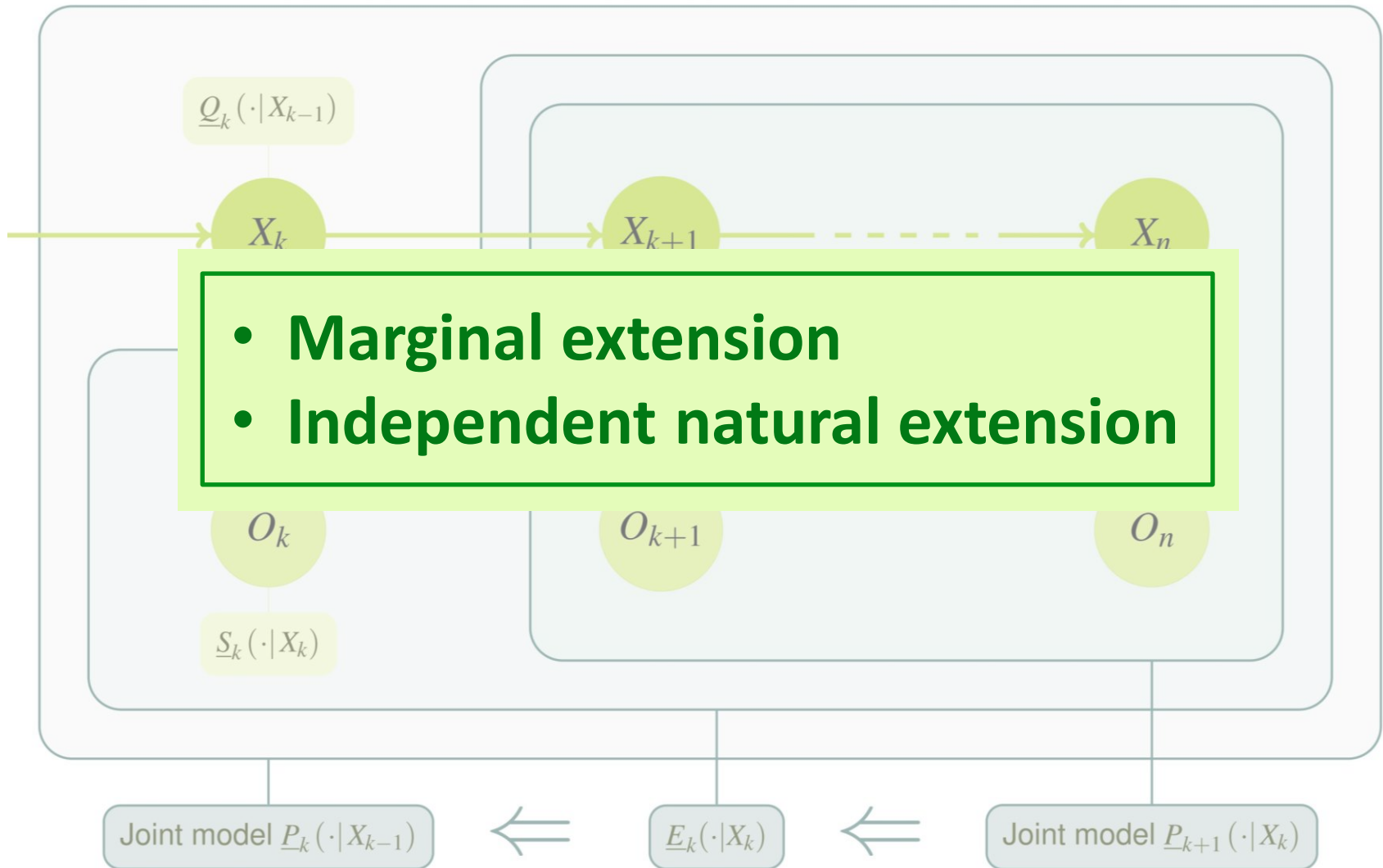
Recursive construction of a joint model



Recursive construction of a joint model



Recursive construction of a joint model

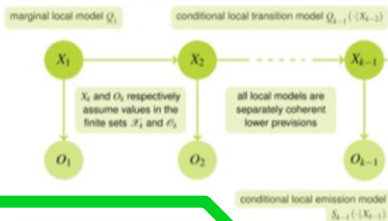


State sequence prediction in imprecise hidden Markov models

Interpretation of the graphical model

Interpretation of the graphical structure Our imprecise hidden Markov model (iHMM) represents the following irrelevance assessments: conditional on its mother variable, the non-parent non-descendants of any variable in the tree are epistemically irrelevant to this variable and its descendants.

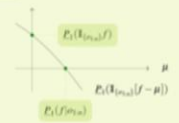
Epistemic irrelevance Y is irrelevant to X whenever the belief model (lower prevision P) about X does not change when we learn something about Y :
 $(\forall x \in \mathcal{X}(\mathcal{Y}))(\forall y \in \mathcal{Y}) P(x) = P(x|y)$.



Maximal state sequences

Conditioning the joint model

Since we assume that all local lower probabilities are strictly positive (in recent work, we dropped this assumption), $P_i(\{x_{i+1}\}) > 0$ and the Generalised Bayes Rule yields a uniquely coherent value of $P_i(f|f_{i+1})$, which has (this is very useful) the same sign as $P_i(\{x_{i+1}\})$.



Optimality criterion

We can express a strict preference \succ between state sequences $x_{1:n}$ and $x_{1:n}$ as follows: $x_{1:n} \succ x_{1:n} \Leftrightarrow P_i(\{x_{i+1}\} - \mathbb{1}_{\{x_{i+1}\}})_{x_{1:n}} > 0$. This induces a strict partial order \succ on the set of state sequences $\mathcal{X}_{1:n}$, and we consider a sequence $x_{1:n}$ to be optimal when it is undominated, or maximal, in this strict partial order:

$$x_{1:n} \in \text{opt}(\mathcal{X}_{1:n}|x_{1:n}) \Leftrightarrow (\forall x_{1:n} \in \mathcal{X}_{1:n}) x_{1:n} \not\succeq x_{1:n}$$

$$\Leftrightarrow (\forall x_{1:n} \in \mathcal{X}_{1:n}) P_i(\{x_{i+1}\} - \mathbb{1}_{\{x_{i+1}\}})_{x_{1:n}} \leq 0$$

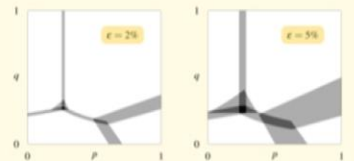
$$\Leftrightarrow (\forall x_{1:n} \in \mathcal{X}_{1:n}) P_i(\{x_{i+1}\} - \mathbb{1}_{\{x_{i+1}\}})_{x_{1:n}} \leq 0$$

In an analogous manner we define the optimal subsequences:

$$x_{k:n} \in \text{opt}(\mathcal{X}_{k:n}|x_{k:n}) \Leftrightarrow (\forall x_{k:n} \in \mathcal{X}_{k:n}) P_i(\{x_{i+1}\} - \mathbb{1}_{\{x_{i+1}\}})_{x_{k:n}} \leq 0$$

A first experiment

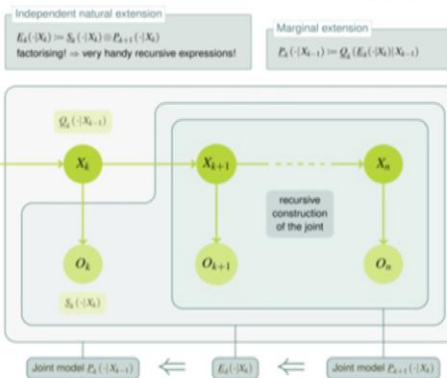
Motivation and description No algorithm, however cleverly designed, will be able to find all maximal sequences efficiently if there are too many. Because this number of maximal sequences is so important, we study its behaviour in more detail. We consider a binary, stationary iHMM with precise emission models. The imprecise marginal and transition models are generated by mixing precise models with a vacuous one, using a mixture coefficient ϵ . For a particular observation sequence of length three, we plot the number of maximal sequences as a function of the transition probabilities p and q . As this number grows from 1 to 4 the areas go from white to black.



Results We see that there are large regions of transition probability space where the number of maximal elements remains fairly small. The plots also display quite interesting behaviour. If we let the imprecision grow, by using higher ϵ , the areas with multiple maximal sequences become larger. They are expanded versions of the lines of indifference that occur in the precise case.

Jasper De Bock and Gert de Cooman

{Jasper.DeBock, Gert.deCooman}@UGent.be
 SYSTEms, Ghent University, Belgium



Principle of Optimality Using the recursive expressions for the joint model, we can derive an appropriate version of Bellman's Principle of Optimality.

Backward-forward recursion We let k run backward from n to 1. For each k and every

$\text{opt}(\mathcal{X}_{k+1:n}|x_{k+1:n}) \subseteq \bigcup_{x_k \in \mathcal{X}_k} \text{opt}(\mathcal{X}_{k+1:n}|x_k, x_{k+1:n})$

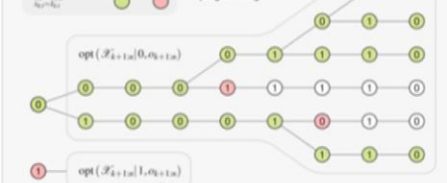
An alternative criterion If we limit ourselves to the possible sequences selected by the Principle of Optimality, we can reformulate the optimality criterion as follows:
 $x_{k:n} \in \text{opt}(\mathcal{X}_{k:n}|x_{k:n}) \Leftrightarrow \alpha_k^{\text{opt}}(k|x_{k:n}) \leq \alpha_k(x_{k:n})$

$$\text{repeat for } l \text{ from } k \text{ to } n \text{ until } \text{no}$$

$$\alpha_k^{\text{opt}}(k|x_{k:n}) \geq \alpha_k^{\text{opt}}(k|x_{k-1:n}) ?$$

$$\alpha_k^{\text{opt}}(k|x_{k:n}) = \max_{x_{k-1:n}} \alpha_k(x_{k-1:n})$$

Can be calculated efficiently by dynamical programming.



The EstiHMM algorithm

Computational complexity

Theoretical analysis We prove that the computational complexity is at worst quadratic in the length of the Markov chain, cubic in the number of states, and roughly speaking linear in the number of maximal sequences (each backward step in the backward-forward loop has a linear complexity in the number of maximal elements at that stage).

Empirical confirmation In order to demonstrate that our algorithm is indeed quite efficient, we let it determine the maximal sequences for a random output sequence of length 100. The iHMM we use to determine the maximal sequences is generated by mixing precise local models with a vacuous one, using a mixture coefficient ϵ . For $\epsilon = 2\%$, there are 5 maximal sequences and it takes 0.2 seconds to calculate them. If we let ϵ grow to 5%, we get 764 maximal sequences and these can be determined in 32 seconds. This demonstrates that the complexity is indeed linear in the number of solutions and that the algorithm can efficiently calculate the maximal sequences even for long output sequences.

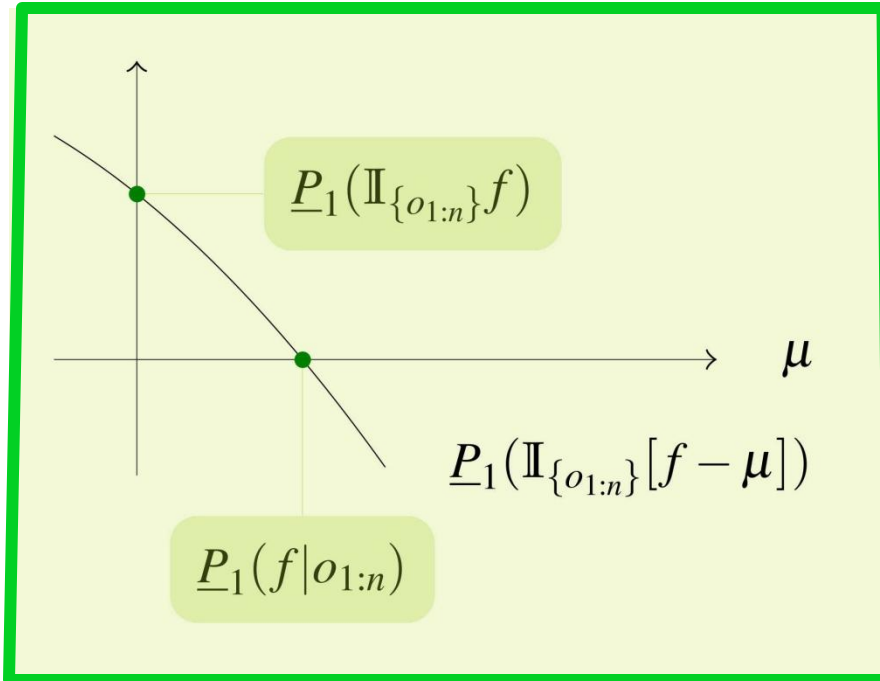


Conditioning the model on the observations

State sequence prediction in imprecise hidden Markov models

Conditioning the model on the observations

Generalised Bayes rule:
An extension of the Bayes rule to imprecise probabilities

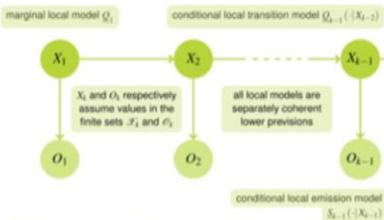


State sequence prediction in imprecise hidden Markov models

Interpretation of the graphical model

Interpretation of the graphical structure Our imprecise hidden Markov model (iHMM) represents the following irrelevance assessments: conditional on its mother variable, the non-parent non-descendants of any variable in the tree are epistemically irrelevant to this variable and its descendants.

Epistemic irrelevance Y is irrelevant to X whenever the belief model (lower prevision E) about X does not change when we learn something about Y :
 $(\forall x \in \mathcal{X}(\mathcal{Y}))(\forall y \in \mathcal{Y}) E(x) = E(x|y)$.



Maximal state sequences

Conditioning the joint model

Since we assume that all local lower probabilities are strictly positive (in recent work, we dropped this assumption), $P_i(\{x_{i+1}\}) > 0$ and the Generalised Bayes Rule yields a uniquely coherent value of $P_i(f|y_{i+1})$, which has (this is very useful) the same value as $P_i(f|y_{i+1})$.



Optimality criterion

We can express a strict preference \succ between state sequences $x_{1:n}$ and $x_{1:n}$ as follows: $x_{1:n} \succ x_{1:n} \Leftrightarrow P_i(I_{i+1}|x_{1:n}) - P_i(I_{i+1}|x_{1:n}) > 0$. This induces a strict partial order \succ on the set of state sequences $\mathcal{X}_{1:n}$, and we consider a sequence $x_{1:n}$ to be optimal when it is undominated, or maximal, in this strict partial order:

$$x_{1:n} \in \text{opt}(\mathcal{X}_{1:n}|y_{1:n}) \Leftrightarrow (\forall x_{1:n} \in \mathcal{X}_{1:n}) x_{1:n} \not\succeq x_{1:n}$$

$$\Leftrightarrow (\forall x_{1:n} \in \mathcal{X}_{1:n}) P_i(I_{i+1}|x_{1:n}) - P_i(I_{i+1}|x_{1:n}) \leq 0$$

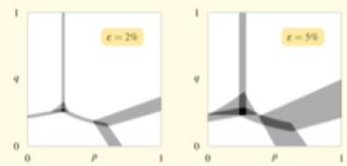
$$\Leftrightarrow (\forall x_{1:n} \in \mathcal{X}_{1:n}) P_i(I_{i+1}|x_{1:n}) - P_i(I_{i+1}|x_{1:n}) \leq 0$$

In an analogous manner we define the optimal subsequences:

$$x_{k:n} \in \text{opt}(\mathcal{X}_{k:n}|y_{k:n}) \Leftrightarrow (\forall x_{k:n} \in \mathcal{X}_{k:n}) P_i(I_{i+1}|x_{k:n}) - P_i(I_{i+1}|x_{k:n}) \leq 0$$

A first experiment

Motivation and description No algorithm, however cleverly designed, will be able to find all maximal sequences efficiently if there are too many. Because this number of maximal sequences is so important, we study its behaviour in more detail. We consider a binary, stationary iHMM with precise emission models. The imprecise marginal and transition models are generated by mixing precise models with a vacuous one, using a mixture coefficient ϵ . For a particular observation sequence of length three, we plot the number of maximal sequences as a function of the transition probabilities ρ and q . As this number grows from 1 to 4 the areas go from white to black.



Results We see that there are large regions of transition probability space where the number of maximal elements remains fairly small. The plots also display quite interesting behaviour. If we let the imprecision grow, by using higher ϵ , the areas with multiple maximal sequences become larger. They are expanded versions of the lines of indifference that occur in the precise case.

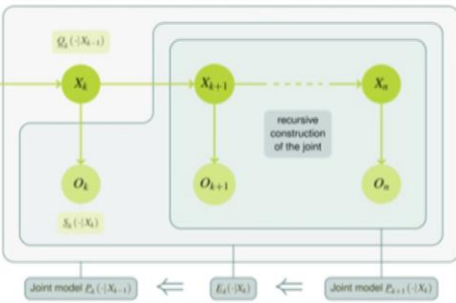
Jasper De Bock and Gert de Cooman

{Jasper.DeBock, Gert.deCooman}@UGent.be
 SYSTEms, Ghent University, Belgium



Independent natural extension
 $E_i(\cdot|X_i) = \bar{S}_i(\cdot|X_i) \otimes P_{i+1}(\cdot|X_i)$
 factoring! \Rightarrow very handy recursive expressions!

Marginal extension
 $E_i(\cdot|X_{i-1}) = \bar{Q}_i(\bar{S}_i(\cdot|X_i)|X_{i-1})$

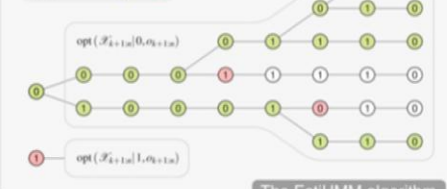


Principle of Optimality Using the recursive expressions for the joint model, we can derive an appropriate version of Bellman's Principle of Optimality:
 $x_{k:n} \in \text{opt}(\mathcal{X}_{k:n}|y_{k:n}, o_{k:n}) \Leftrightarrow x_{k+1:n} \in \text{opt}(\mathcal{X}_{k+1:n}|y_{k+1:n}, o_{k+1:n})$.

Backward-forward recursion We let k run backward from n to 1 . For each k and every $x_{k+1:n} \in \mathcal{X}_{k+1:n}$, we determine $\text{opt}(\mathcal{X}_{k:n}|y_{k:n}, o_{k:n})$ with our alternative procedure.

An alternative criterion If we limit ourselves to the possible sequences selected by the Principle of Optimality, we can reformulate the optimality criterion as follows:
 $x_{k:n} \in \text{opt}(\mathcal{X}_{k:n}|y_{k:n}, o_{k:n}) \Leftrightarrow \alpha_k^{\text{opt}}(k|x_{k:n}) \leq \alpha_k(x_{k:n})$.

Identify by choosing the forward running procedure demonstrated in the figure below. The sequences of $\text{opt}(\mathcal{X}_{k:n}|y_{k:n}, o_{k:n})$ are the elements of $\text{opt}(\mathcal{X}_{k:n}|y_{k:n}, o_{k:n})$ in this binary example, $\mathcal{X} = \{0, 1\}$.



Computational complexity

Theoretical analysis We prove that the computational complexity is at worst quadratic in the length of the Markov chain, cubic in the number of states, and roughly speaking linear in the number of maximal sequences (each backward step in the backward-forward loop has a linear complexity in the number of maximal elements at that stage).

Empirical confirmation In order to demonstrate that our algorithm is indeed quite efficient, we let it determine the maximal sequences for a random output sequence of length 100. The iHMM we use to determine the maximal sequences is generated by mixing precise local models with a vacuous one, using a mixture coefficient ϵ . For $\epsilon = 2\%$, there are 5 maximal sequences and it takes 0.2 seconds to calculate them. If we let ϵ grow to 5%, we get 764 maximal sequences and these can be determined in 32 seconds. This demonstrates that the complexity is indeed linear in the number of solutions and that the algorithm can efficiently calculate the maximal sequences even for long output sequences.

Maximal state sequences

State sequence prediction in imprecise hidden Markov models

We predict the state sequence by calculating a set of optimal sequences

Notion of optimality: **maximality**

Strict partial ordering:

$$\hat{x}_{1:n} \succ x_{1:n} \Leftrightarrow \underline{P}_1(\mathbb{I}_{\{\hat{x}_{1:n}\}} - \mathbb{I}_{\{x_{1:n}\}} | o_{1:n}) > 0.$$

Maximal state sequences:

$$\hat{x}_{1:n} \in \text{opt}(\mathcal{X}_{1:n} | o_{1:n})$$

$$\Leftrightarrow (\forall x_{1:n} \in \mathcal{X}_{1:n}) x_{1:n} \not\succeq \hat{x}_{1:n}$$

$$\Leftrightarrow (\forall x_{1:n} \in \mathcal{X}_{1:n}) \underline{P}_1(\mathbb{I}_{\{x_{1:n}\}} - \mathbb{I}_{\{\hat{x}_{1:n}\}} | o_{1:n}) \leq 0$$

State sequence prediction in imprecise hidden Markov models

Maximal state sequences

We predict the state sequence by calculating a set of optimal sequences

Notion of optimality: **maximality**

Strict partial ordering:

$$\hat{x}_{1:n} \succ x_{1:n} \Leftrightarrow \underline{P}_1(\mathbb{I}_{\{\hat{x}_{1:n}\}} - \mathbb{I}_{\{x_{1:n}\}} | o_{1:n}) > 0.$$

Maximal state sequences:

$$\hat{x}_{1:n} \in \text{opt}(\mathcal{X}_{1:n} | o_{1:n})$$

$$\Leftrightarrow (\forall x_{1:n} \in \mathcal{X}_{1:n}) x_{1:n} \not\succeq \hat{x}_{1:n}$$

$$\Leftrightarrow (\forall x_{1:n} \in \mathcal{X}_{1:n}) \underline{P}_1(\mathbb{I}_{\{x_{1:n}\}} - \mathbb{I}_{\{\hat{x}_{1:n}\}} | o_{1:n}) \leq 0$$

$$\Leftrightarrow (\forall x_{1:n} \in \mathcal{X}_{1:n}) \underline{P}_1(\mathbb{I}_{\{o_{1:n}\}} [\mathbb{I}_{\{x_{1:n}\}} - \mathbb{I}_{\{\hat{x}_{1:n}\}}]) \leq 0.$$

State sequence prediction in imprecise hidden Markov models

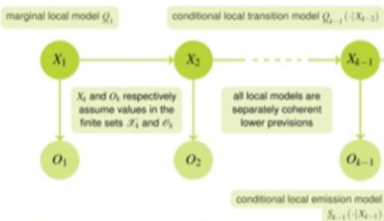
Maximal state sequences

State sequence prediction in imprecise hidden Markov models

Interpretation of the graphical model

Interpretation of the graphical structure Our imprecise hidden Markov model (iHMM) represents the following irrelevance assessments: conditional on its mother variable, the non-parent non-descendants of any variable in the tree are epistemically irrelevant to this variable and its descendants.

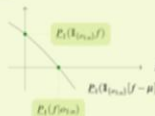
Epistemic irrelevance Y is irrelevant to X whenever the belief model (lower prevision E) about X does not change when we learn something about Y :
 $(\forall x \in \mathcal{X}(\mathcal{Y}))(\forall y \in \mathcal{Y}) E(x) = E(x|y)$.



Maximal state sequences

Conditioning the joint model

Since we assume that all local lower probabilities are strictly positive (in recent work, we dropped this assumption), $P_k(\{x_{k-1}\}) > 0$ and the Generalised Bayes Rule yields a uniquely coherent value of $P_k(f|y_{k-1})$, which has (this is very useful) the same sign as $P_k(\{x_{k-1}\})$.



Optimality criterion We can express a strict preference \succ between state sequences $x_{1:n}$ and $x_{1:n}$ as follows: $x_{1:n} \succ x_{1:n} \Leftrightarrow E_k(\{x_{1:n}\} - \{x_{1:n}\})_{k-1} > 0$. This induces a strict partial order \succ on the set of state sequences $\mathcal{X}_{1:n}$, and we consider a sequence $x_{1:n}$ to be optimal when it is undominated, or maximal, in this strict partial order:

$$x_{1:n} \in \text{opt}(\mathcal{X}_{1:n}) \Leftrightarrow (\forall x_{1:n} \in \mathcal{X}_{1:n}) x_{1:n} \not\succeq x_{1:n}$$

$$\Leftrightarrow (\forall x_{1:n} \in \mathcal{X}_{1:n}) E_k(\{x_{1:n}\} - \{x_{1:n}\})_{k-1} \leq 0$$

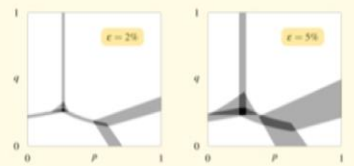
$$\Leftrightarrow (\forall x_{1:n} \in \mathcal{X}_{1:n}) E_k(\{x_{1:n}\} - \{x_{1:n}\})_{k-1} \leq 0$$

In an analogous manner we define the optimal subsequences:

$$x_{k:n} \in \text{opt}(\mathcal{X}_{k:n}) \Leftrightarrow (\forall x_{k:n} \in \mathcal{X}_{k:n}) E_k(\{x_{k:n}\} - \{x_{k:n}\})_{k-1} \leq 0$$

A first experiment

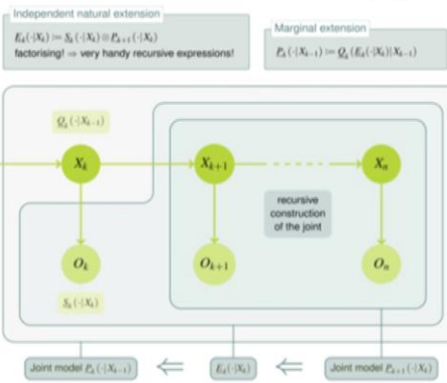
Motivation and description No algorithm, however cleverly designed, will be able to find all maximal sequences efficiently if there are too many. Because this number of maximal sequences is so important, we study its behaviour in more detail. We consider a binary, stationary iHMM with precise emission models. The imprecise marginal and transition models are generated by mixing precise models with a vacuous one, using a mixture coefficient ϵ . For a particular observation sequence of length three, we plot the number of maximal sequences as a function of the transition probabilities ρ and q . As this number grows from 1 to 4 the areas go from white to black.



Results We see that there are large regions of transition probability space where the number of maximal elements remains fairly small. The plots also display quite interesting behaviour. If we let the imprecision grow, by using higher ϵ , the areas with multiple maximal sequences become larger. They are expanded versions of the lines of indifference that occur in the precise case.

Jasper De Bock and Gert de Cooman

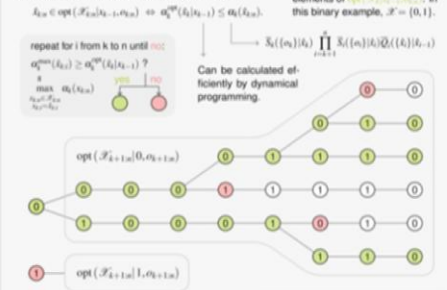
{Jasper.DeBock, Gert.deCooman}@UGent.be
 SYSTEms, Ghent University, Belgium



Principle of Optimality Using the recursive expressions for the joint model, we can derive an appropriate version of Bellman's Principle of Optimality:
 $x_{k:n} \in \text{opt}(\mathcal{X}_{k:n}|x_{1:k-1}, o_{1:k-1}) \Leftrightarrow x_{k+1:n} \in \text{opt}(\mathcal{X}_{k+1:n}|x_{1:k}, o_{1:k})$, which in turn implies that

$$\text{opt}(\mathcal{X}_{k:n}|x_{1:k-1}, o_{1:k-1}) \subseteq \bigcup_{x_k \in \mathcal{X}_k} \text{opt}(\mathcal{X}_{k+1:n}|x_{1:k}, o_{1:k})$$

An alternative criterion If we limit ourselves to the possible sequences selected by the Principle of Optimality, we can reformulate the optimality criterion as follows:
 $x_{k:n} \in \text{opt}(\mathcal{X}_{k:n}|x_{1:k-1}, o_{1:k-1}) \Leftrightarrow \alpha_k^{\text{opt}}(x_k|x_{1:k-1}) \leq \alpha_k(x_k)$.



The EstiHMM algorithm

Computational complexity

Theoretical analysis We prove that the computational complexity is at worst quadratic in the length of the Markov chain, cubic in the number of states, and roughly speaking linear in the number of maximal sequences (each backward step in the backward-forward loop has a linear complexity in the number of maximal elements at that stage).

Empirical confirmation In order to demonstrate that our algorithm is indeed quite efficient, we let it determine the maximal sequences for a random output sequence of length 100. The iHMM we use to determine the maximal sequences is generated by mixing precise local models with a vacuous one, using a mixture coefficient ϵ . For $\epsilon = 2\%$, there are 5 maximal sequences and it takes 0.2 seconds to calculate them. If we let ϵ grow to 5%, we get 764 maximal sequences and these can be determined in 32 seconds. This demonstrates that the complexity is indeed linear in the number of solutions and that the algorithm can efficiently calculate the maximal sequences even for long output sequences.

State sequence prediction in imprecise hidden Markov models

EstiHMM:
 an efficient algorithm to determine the maximal state sequences in an imprecise hidden Markov model

Principle of Optimality Using the recursive expressions for the joint model, we can derive an appropriate version of Bellman's Principle of Optimality:

$\hat{x}_{k:n} \in \text{opt}(\mathcal{X}_{k:n}|x_{k-1}, o_{k:n}) \Rightarrow \hat{x}_{k+1:n} \in \text{opt}(\mathcal{X}_{k+1:n}|\hat{x}_k, o_{k+1:n})$,
which in turn implies that

$$\text{opt}(\mathcal{X}_{k:n}|x_{k-1}, o_{k:n}) \subseteq \bigcup_{x_k \in \mathcal{X}_k} x_k \oplus \text{opt}(\mathcal{X}_{k+1:n}|x_k, o_{k+1:n}).$$

An alternative criterion If we limit ourselves to the possible sequences selected by the Principle of Optimality, we can reformulate the optimality criterion as follows:

$$\hat{x}_{k:n} \in \text{opt}(\mathcal{X}_{k:n}|x_{k-1}, o_{k:n}) \Leftrightarrow \alpha_k^{\text{opt}}(\hat{x}_k|x_{k-1}) \leq \alpha_k(\hat{x}_{k:n}).$$

Backward-forward recursion

We let k run *backward* from n to 1. For each k and every $x_{k-1} \in \mathcal{X}_{k-1}$, we determine $\text{opt}(\mathcal{X}_k|x_{k-1}, o_{k:n})$ with our alternative optimality criterion. We prove that we can do this efficiently by executing the *forward running procedure* demonstrated in the figure below. The sequences of **green nodes** are the elements of $\text{opt}(\mathcal{X}_k|x_{k-1}, o_{k:n})$. In this binary example, $\mathcal{X} = \{0, 1\}$.

repeat for i from k to n until **no**:

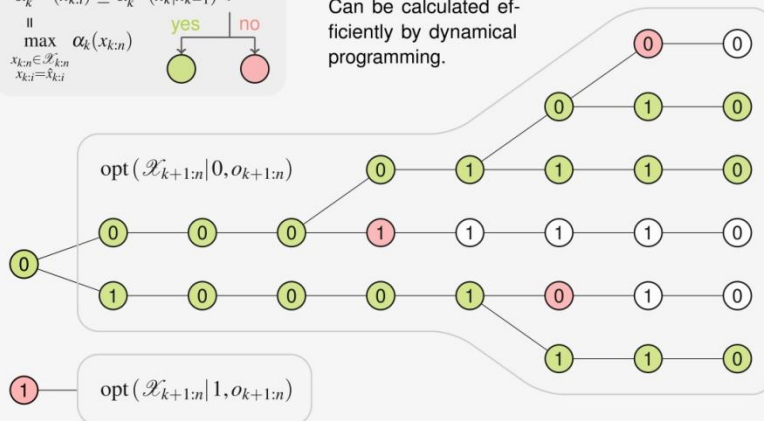
$$\alpha_k^{\text{max}}(\hat{x}_{k:i}) \geq \alpha_k^{\text{opt}}(\hat{x}_k|x_{k-1}) ?$$

$$\text{if } \max_{\substack{x_{k:n} \in \mathcal{X}_{k:n} \\ x_{k-1} = \hat{x}_{k-1}}} \alpha_k(x_{k:n}) \text{ yes | no}$$



Can be calculated efficiently by dynamical programming.

$$\bar{S}_k(\{o_k\}|\hat{x}_k) \prod_{i=k+1}^n \bar{S}_i(\{o_i\}|\hat{x}_i) \bar{Q}_i(\{\hat{x}_i\}|\hat{x}_{i-1})$$



The EstiHMM algorithm

Computational complexity

Theoretical analysis We prove that the computational complexity is at worst quadratic in the length of the Markov chain, cubic in the number of states, and roughly speaking linear in the number of maximal sequences (each backward step in the backward-forward loop has a linear complexity in the number of maximal elements at that stage).

Empirical confirmation In order to demonstrate that our algorithm is indeed quite efficient, we let it determine the maximal sequences for a random output sequence of length 100. The iHMM we use to determine the maximal sequences is generated by mixing precise local models with a vacuous one, using a mixture coefficient ϵ . For $\epsilon = 2\%$, there are 5 maximal sequences and it takes 0.2 seconds to calculate them. If we let ϵ grow to 5%, we get 764 maximal sequences and these can be determined in 32 seconds. This demonstrates that the complexity is indeed linear in the number of solutions and that the algorithm can efficiently calculate the maximal sequences even for long output sequences.

State sequence prediction in imprecise hidden Markov models

EstiHMM:
an efficient algorithm to determine the maximal state sequences in an imprecise hidden Markov model

EstiHMM: an efficient algorithm to determine the maximal sequences

Principle of Optimality Using the recursive expressions for the joint model, we can derive an appropriate version of Bellman's Principle of Optimality:

$\hat{x}_{k:n} \in \text{opt}(\mathcal{X}_{k:n}|x_{k-1}, o_{k:n}) \Rightarrow \hat{x}_{k+1:n} \in \text{opt}(\mathcal{X}_{k+1:n}|\hat{x}_k, o_{k+1:n})$,
which in turn implies that

$$\text{opt}(\mathcal{X}_{k:n}|x_{k-1}, o_{k:n}) \subseteq \bigcup_{x_k \in \mathcal{X}_k} x_k \oplus \text{opt}(\mathcal{X}_{k+1:n}|x_k, o_{k+1:n}).$$

An alternative criterion If we limit ourselves to the possible sequences selected by the Principle of Optimality, we can reformulate the optimality criterion as follows:

$$\hat{x}_{k:n} \in \text{opt}(\mathcal{X}_{k:n}|x_{k-1}, o_{k:n}) \Leftrightarrow \alpha_k^{\text{opt}}(\hat{x}_k|x_{k-1}) \leq \alpha_k(\hat{x}_{k:n}).$$

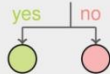
Backward-forward recursion

We let k run *backward* from n to 1 . For each k and every $x_{k-1} \in \mathcal{X}_{k-1}$, we determine $\text{opt}(\mathcal{X}_k|x_{k-1}, o_{k:n})$ with our alternative optimality criterion. We prove that we can do this efficiently by executing the *forward running procedure* demonstrated in the figure below. The sequences of **green nodes** are the elements of $\text{opt}(\mathcal{X}_k|x_{k-1}, o_{k:n})$. In this binary example, $\mathcal{X} = \{0, 1\}$.

repeat for i from k to n until **no**:

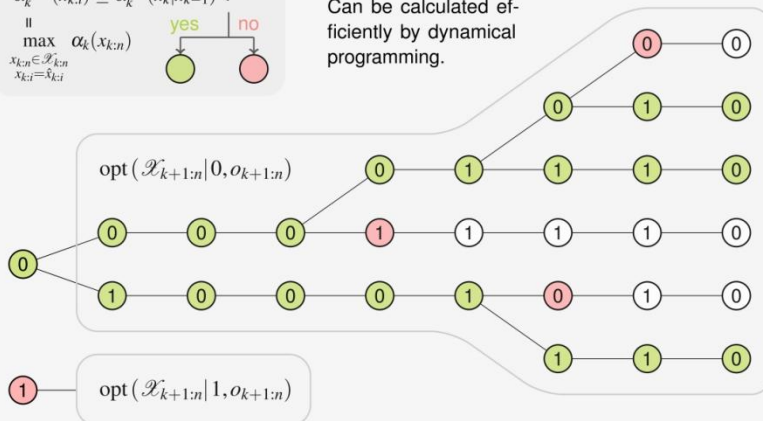
$$\alpha_k^{\text{max}}(\hat{x}_{k:i}) \geq \alpha_k^{\text{opt}}(\hat{x}_k|x_{k-1}) ?$$

$$\text{if } \max_{\substack{x_{k:n} \in \mathcal{X}_{k:n} \\ x_k = \hat{x}_k}} \alpha_k(x_{k:n}) \begin{cases} \text{yes} \\ \text{no} \end{cases}$$



Can be calculated efficiently by dynamical programming.

$$\bar{S}_k(\{o_k\}|\hat{x}_k) \prod_{i=k+1}^n \bar{S}_i(\{o_i\}|\hat{x}_i) \bar{Q}_i(\{\hat{x}_i\}|\hat{x}_{i-1})$$



The EstiHMM algorithm

Computational complexity

Theoretical analysis We prove that the computational complexity is at worst quadratic in the length of the Markov chain, cubic in the number of states, and roughly speaking linear in the number of maximal sequences (each backward step in the backward-forward loop has a linear complexity in the number of maximal elements at that stage).

Empirical confirmation In order to demonstrate that our algorithm is indeed quite efficient, we let it determine the maximal sequences for a random output sequence of length 100. The iHMM we use to determine the maximal sequences is generated by mixing precise local models with a vacuous one, using a mixture coefficient ϵ . For $\epsilon = 2\%$, there are 5 maximal sequences and it takes 0.2 seconds to calculate them. If we let ϵ grow to 5%, we get 764 maximal sequences and these can be determined in 32 seconds. This demonstrates that the complexity is indeed linear in the number of solutions and that the algorithm can efficiently calculate the maximal sequences even for long output sequences.

Principle of Optimality Using the recursive expressions for the joint model, we can derive an appropriate version of Bellman's Principle of Optimality:

$$\hat{x}_{k:n} \in \text{opt}(\mathcal{X}_{k:n}|x_{k-1}, o_{k:n}) \Rightarrow \hat{x}_{k+1:n} \in \text{opt}(\mathcal{X}_{k+1:n}|\hat{x}_k, o_{k+1:n}),$$

which in turn implies that

$$\text{opt}(\mathcal{X}_{k:n}|x_{k-1}, o_{k:n}) \subseteq \bigcup_{x_k \in \mathcal{X}_k} x_k \oplus \text{opt}(\mathcal{X}_{k+1:n}|x_k, o_{k+1:n}).$$

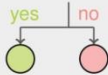
An alternative criterion If we limit ourselves to the possible sequences selected by the Principle of Optimality, we can reformulate the optimality criterion as follows:

$$\hat{x}_{k:n} \in \text{opt}(\mathcal{X}_{k:n}|x_{k-1}, o_{k:n}) \Leftrightarrow \alpha_k^{\text{opt}}(\hat{x}_k|x_{k-1}) \leq \alpha_k(\hat{x}_{k:n}).$$

repeat for i from k to n until **no**:

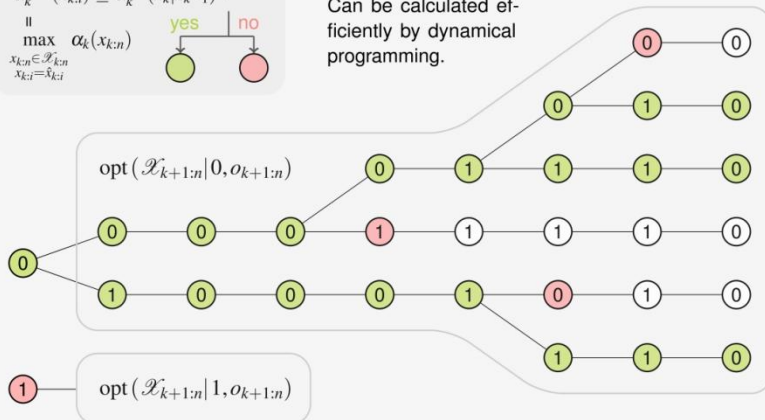
$$\alpha_k^{\text{max}}(\hat{x}_{k:i}) \geq \alpha_k^{\text{opt}}(\hat{x}_k|x_{k-1})?$$

$$\text{if } \max_{\substack{x_{k:n} \in \mathcal{X}_{k:n} \\ x_{k-1} = \hat{x}_{k-1}}} \alpha_k(x_{k:n}) \begin{cases} \text{yes} \\ \text{no} \end{cases}$$



Can be calculated efficiently by dynamical programming.

$$\bar{S}_k(\{o_k\}|\hat{x}_k) \prod_{i=k+1}^n \bar{S}_i(\{o_i\}|\hat{x}_i) \bar{Q}_i(\{\hat{x}_i\}|\hat{x}_{i-1})$$



The EstiHMM algorithm

Computational complexity

Theoretical analysis We prove that the computational complexity is at worst quadratic in the length of the Markov chain, cubic in the number of states, and roughly speaking linear in the number of maximal sequences (each backward step in the backward-forward loop has a linear complexity in the number of maximal elements at that stage).

Empirical confirmation In order to demonstrate that our algorithm is indeed quite efficient, we let it determine the maximal sequences for a random output sequence of length 100. The iHMM we use to determine the maximal sequences is generated by mixing precise local models with a vacuous one, using a mixture coefficient ϵ . For $\epsilon = 2\%$, there are 5 maximal sequences and it takes 0.2 seconds to calculate them. If we let ϵ grow to 5%, we get 764 maximal sequences and these can be determined in 32 seconds. This demonstrates that the complexity is indeed linear in the number of solutions and that the algorithm can efficiently calculate the maximal sequences even for long output sequences.

EstiHMM: an efficient algorithm to determine the maximal sequences

- Principle of optimality

Principle of Optimality Using the recursive expressions for the joint model, we can derive an appropriate version of Bellman's Principle of Optimality:

$$\hat{x}_{k:n} \in \text{opt}(\mathcal{X}_{k:n}|x_{k-1}, o_{k:n}) \Rightarrow \hat{x}_{k+1:n} \in \text{opt}(\mathcal{X}_{k+1:n}|\hat{x}_k, o_{k+1:n}),$$

which in turn implies that

$$\text{opt}(\mathcal{X}_{k:n}|x_{k-1}, o_{k:n}) \subseteq \bigcup_{x_k \in \mathcal{X}_k} x_k \oplus \text{opt}(\mathcal{X}_{k+1:n}|x_k, o_{k+1:n}).$$

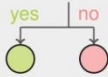
An alternative criterion If we limit ourselves to the possible sequences selected by the Principle of Optimality, we can reformulate the optimality criterion as follows:

$$\hat{x}_{k:n} \in \text{opt}(\mathcal{X}_{k:n}|x_{k-1}, o_{k:n}) \Leftrightarrow \alpha_k^{\text{opt}}(\hat{x}_k|x_{k-1}) \leq \alpha_k(\hat{x}_{k:n}).$$

repeat for i from k to n until no:

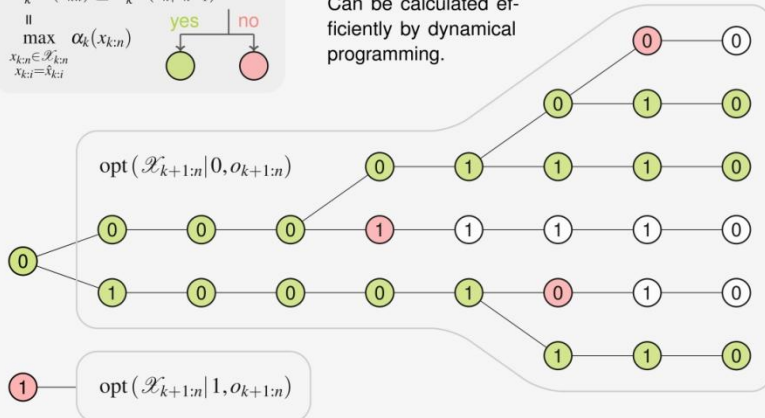
$$\alpha_k^{\text{max}}(\hat{x}_{k:i}) \geq \alpha_k^{\text{opt}}(\hat{x}_k|x_{k-1}) ?$$

$$\text{if } \max_{\substack{x_{k:n} \in \mathcal{X}_{k:n} \\ x_{k-1} = \hat{x}_{k-1}}} \alpha_k(x_{k:n}) \begin{cases} \text{yes} \\ \text{no} \end{cases}$$



Can be calculated efficiently by dynamical programming.

$$\bar{S}_k(\{o_k\}|\hat{x}_k) \prod_{i=k+1}^n \bar{S}_i(\{o_i\}|\hat{x}_i) \bar{Q}_i(\{\hat{x}_i\}|\hat{x}_{i-1})$$



The EstiHMM algorithm

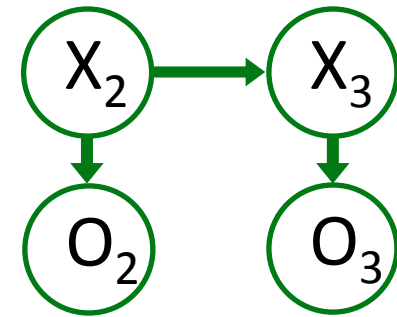
Computational complexity

Theoretical analysis We prove that the computational complexity is at worst quadratic in the length of the Markov chain, cubic in the number of states, and roughly speaking linear in the number of maximal sequences (each backward step in the backward-forward loop has a linear complexity in the number of maximal elements at that stage).

Empirical confirmation In order to demonstrate that our algorithm is indeed quite efficient, we let it determine the maximal sequences for a random output sequence of length 100. The iHMM we use to determine the maximal sequences is generated by mixing precise local models with a vacuous one, using a mixture coefficient ϵ . For $\epsilon = 2\%$, there are 5 maximal sequences and it takes 0.2 seconds to calculate them. If we let ϵ grow to 5%, we get 764 maximal sequences and these can be determined in 32 seconds. This demonstrates that the complexity is indeed linear in the number of solutions and that the algorithm can efficiently calculate the maximal sequences even for long output sequences.

EstiHMM: an efficient algorithm to determine the maximal sequences

- Principle of optimality



Principle of Optimality Using the recursive expressions for the joint model, we can derive an appropriate version of Bellman's Principle of Optimality:

$$\hat{x}_{k:n} \in \text{opt}(\mathcal{X}_{k:n}|x_{k-1}, o_{k:n}) \Rightarrow \hat{x}_{k+1:n} \in \text{opt}(\mathcal{X}_{k+1:n}|\hat{x}_k, o_{k+1:n}),$$

which in turn implies that

$$\text{opt}(\mathcal{X}_{k:n}|x_{k-1}, o_{k:n}) \subseteq \bigcup_{x_k \in \mathcal{X}} x_k \oplus \text{opt}(\mathcal{X}_{k+1:n}|x_k, o_{k+1:n}).$$

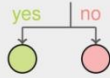
An alternative criterion If we limit ourselves to the possible sequences selected by the Principle of Optimality, we can reformulate the optimality criterion as follows:

$$\hat{x}_{k:n} \in \text{opt}(\mathcal{X}_{k:n}|x_{k-1}, o_{k:n}) \Leftrightarrow \alpha_k^{\text{opt}}(\hat{x}_k|x_{k-1}) \leq \alpha_k(\hat{x}_{k:n}).$$

repeat for i from k to n until no:

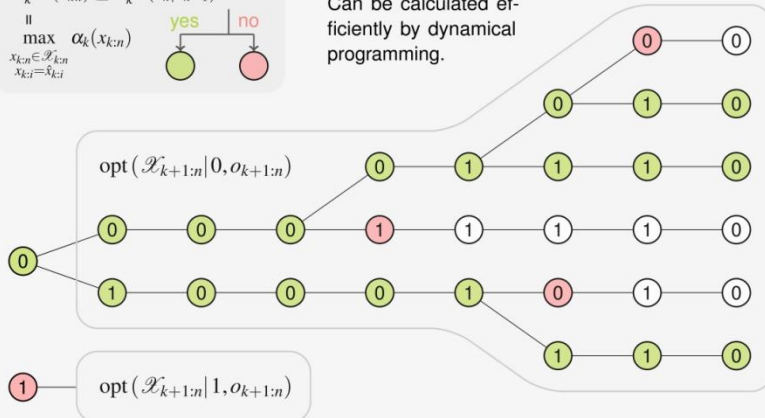
$$\alpha_k^{\text{max}}(\hat{x}_{k:i}) \geq \alpha_k^{\text{opt}}(\hat{x}_k|x_{k-1})?$$

$$\text{if } \max_{\substack{x_{k:n} \in \mathcal{X}_{k:n} \\ x_{k-1} = \hat{x}_{k-1}}} \alpha_k(x_{k:n}) \begin{cases} \text{yes} \\ \text{no} \end{cases}$$



Can be calculated efficiently by dynamical programming.

$$\bar{S}_k(\{o_k\}|\hat{x}_k) \prod_{i=k+1}^n \bar{S}_i(\{o_i\}|\hat{x}_i) \bar{Q}_i(\{\hat{x}_i\}|\hat{x}_{i-1})$$



The EstiHMM algorithm

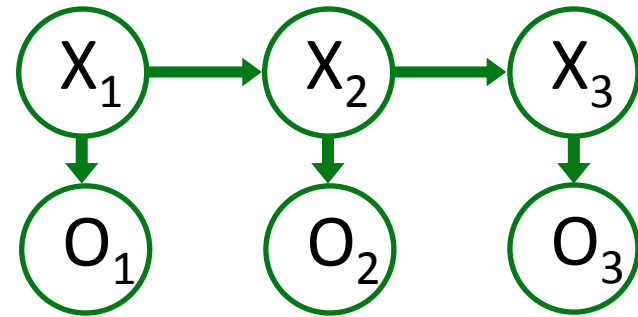
Computational complexity

Theoretical analysis We prove that the computational complexity is at worst quadratic in the length of the Markov chain, cubic in the number of states, and roughly speaking linear in the number of maximal sequences (each backward step in the backward-forward loop has a linear complexity in the number of maximal elements at that stage).

Empirical confirmation In order to demonstrate that our algorithm is indeed quite efficient, we let it determine the maximal sequences for a random output sequence of length 100. The iHMM we use to determine the maximal sequences is generated by mixing precise local models with a vacuous one, using a mixture coefficient ϵ . For $\epsilon = 2\%$, there are 5 maximal sequences and it takes 0.2 seconds to calculate them. If we let ϵ grow to 5%, we get 764 maximal sequences and these can be determined in 32 seconds. This demonstrates that the complexity is indeed linear in the number of solutions and that the algorithm can efficiently calculate the maximal sequences even for long output sequences.

EstiHMM: an efficient algorithm to determine the maximal sequences

- Principle of optimality



Principle of Optimality Using the recursive expressions for the joint model, we can derive an appropriate version of Bellman's Principle of Optimality:

$$\hat{x}_{k:n} \in \text{opt}(\mathcal{X}_{k:n}|x_{k-1}, o_{k:n}) \Rightarrow \hat{x}_{k+1:n} \in \text{opt}(\mathcal{X}_{k+1:n}|\hat{x}_k, o_{k+1:n}),$$

which in turn implies that

$$\text{opt}(\mathcal{X}_{k:n}|x_{k-1}, o_{k:n}) \subseteq \bigcup x_k \oplus \text{opt}(\mathcal{X}_{k+1:n}|x_k, o_{k+1:n}).$$

An alternative criterion If we limit ourselves to the possible sequences selected by the Principle of Optimality, we can reformulate the optimality criterion as follows:

$$\hat{x}_{k:n} \in \text{opt}(\mathcal{X}_{k:n}|x_{k-1}, o_{k:n}) \Leftrightarrow \alpha_k^{\text{opt}}(\hat{x}_k|x_{k-1}) \leq \alpha_k(\hat{x}_{k:n}).$$

Backward-forward recursion

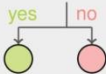
We let k run *backward* from n to 1 . For each k and every $x_{k-1} \in \mathcal{X}_{k-1}$, we determine $\text{opt}(\mathcal{X}_k|x_{k-1}, o_{k:n})$ with our alternative optimality criterion. We prove that we can do this efficiently by executing the *forward running procedure* demonstrated in the figure below. The sequences of **green nodes** are the elements of $\text{opt}(\mathcal{X}_k|x_{k-1}, o_{k:n})$. In this binary example, $\mathcal{X} = \{0, 1\}$.

repeat for i from k to n until **no**:

$$\alpha_k^{\text{max}}(\hat{x}_{k:i}) \geq \alpha_k^{\text{opt}}(\hat{x}_k|x_{k-1}) ?$$

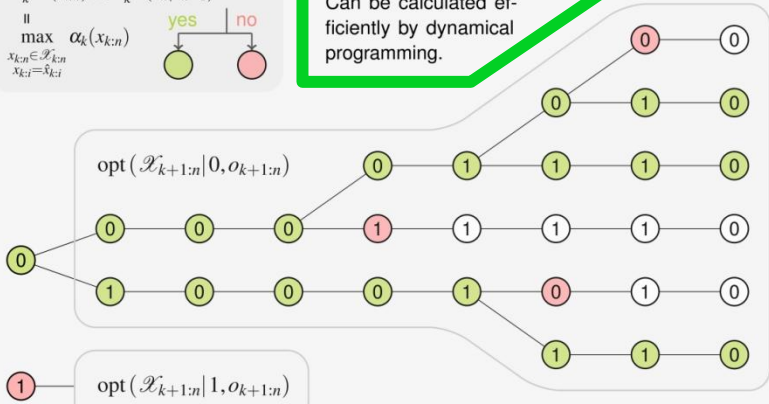
$$\text{if } \max_{x_{k:n} \in \mathcal{X}_{k:n}} \alpha_k(x_{k:n})$$

$$x_{k:n} = \hat{x}_{k:n}$$



Can be calculated efficiently by dynamical programming.

$$\bar{S}_k(\{o_k\}|\hat{x}_k) \prod_{i=k+1}^n \bar{S}_i(\{o_i\}|\hat{x}_i) \bar{Q}_i(\{\hat{x}_i\}|\hat{x}_{i-1})$$



The EstiHMM algorithm

Computational complexity

Theoretical analysis We prove that the computational complexity is at worst quadratic in the length of the Markov chain, cubic in the number of states, and roughly speaking linear in the number of maximal sequences (each backward step in the backward-forward loop has a linear complexity in the number of maximal elements at that stage).

Empirical confirmation In order to demonstrate that our algorithm is indeed quite efficient, we let it determine the maximal sequences for a random output sequence of length 100. The iHMM we use to determine the maximal sequences is generated by mixing precise local models with a vacuous one, using a mixture coefficient ϵ . For $\epsilon = 2\%$, there are 5 maximal sequences and it takes 0.2 seconds to calculate them. If we let ϵ grow to 5%, we get 764 maximal sequences and these can be determined in 32 seconds. This demonstrates that the complexity is indeed linear in the number of solutions and that the algorithm can efficiently calculate the maximal sequences even for long output sequences.

EstiHMM: an efficient algorithm to determine the maximal sequences

- Principle of optimality
- Deriving an alternative optimality criterion

$$\hat{x}_{1:n} \in \text{opt}(\mathcal{X}_{1:n}|o_{1:n})$$



$$(\forall x_{1:n} \in \mathcal{X}_{1:n}) \underline{P}_1(\mathbb{I}_{\{o_{1:n}\}}[\mathbb{I}_{\{x_{1:n}\}} - \mathbb{I}_{\{\hat{x}_{1:n}\}}]) \leq 0.$$

Principle of Optimality Using the recursive expressions for the joint model, we can derive an appropriate version of Bellman's Principle of Optimality:

$$\hat{x}_{k:n} \in \text{opt}(\mathcal{X}_{k:n}|x_{k-1}, o_{k:n}) \Rightarrow \hat{x}_{k+1:n} \in \text{opt}(\mathcal{X}_{k+1:n}|\hat{x}_k, o_{k+1:n}),$$

which in turn implies that

$$\text{opt}(\mathcal{X}_{k:n}|x_{k-1}, o_{k:n}) \subseteq \bigcup x_k \oplus \text{opt}(\mathcal{X}_{k+1:n}|x_k, o_{k+1:n}).$$

An alternative criterion If we limit ourselves to the possible sequences selected by the Principle of Optimality, we can reformulate the optimality criterion as follows:

$$\hat{x}_{k:n} \in \text{opt}(\mathcal{X}_{k:n}|x_{k-1}, o_{k:n}) \Leftrightarrow \alpha_k^{\text{opt}}(\hat{x}_k|x_{k-1}) \leq \alpha_k(\hat{x}_{k:n}).$$

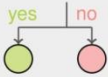
Backward-forward recursion

We let k run *backward* from n to 1 . For each k and every $x_{k-1} \in \mathcal{X}_{k-1}$, we determine $\text{opt}(\mathcal{X}_k|x_{k-1}, o_{k:n})$ with our alternative optimality criterion. We prove that we can do this efficiently by executing the *forward running procedure* demonstrated in the figure below. The sequences of **green nodes** are the elements of $\text{opt}(\mathcal{X}_k|x_{k-1}, o_{k:n})$. In this binary example, $\mathcal{X} = \{0, 1\}$.

repeat for i from k to n until **no**:

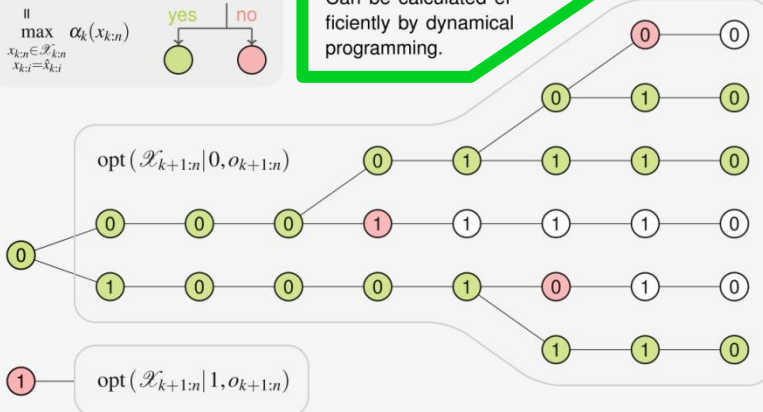
$$\alpha_k^{\text{max}}(\hat{x}_{k:i}) \geq \alpha_k^{\text{opt}}(\hat{x}_k|x_{k-1})?$$

$$\text{if } \max_{\substack{x_{k:n} \in \mathcal{X}_{k:n} \\ x_{k-1} = \hat{x}_{k-1}}} \alpha_k(x_{k:n})$$



Can be calculated efficiently by dynamical programming.

$$\bar{S}_k(\{o_k\}|\hat{x}_k) \prod_{i=k+1}^n \bar{S}_i(\{o_i\}|\hat{x}_i) \bar{Q}_i(\{\hat{x}_i\}|\hat{x}_{i-1})$$



The EstiHMM algorithm

Computational complexity

Theoretical analysis We prove that the computational complexity is at worst quadratic in the length of the Markov chain, cubic in the number of states, and roughly speaking linear in the number of maximal sequences (each backward step in the backward-forward loop has a linear complexity in the number of maximal elements at that stage).

Empirical confirmation In order to demonstrate that our algorithm is indeed quite efficient, we let it determine the maximal sequences for a random output sequence of length 100. The iHMM we use to determine the maximal sequences is generated by mixing precise local models with a vacuous one, using a mixture coefficient ϵ . For $\epsilon = 2\%$, there are 5 maximal sequences and it takes 0.2 seconds to calculate them. If we let ϵ grow to 5%, we get 764 maximal sequences and these can be determined in 32 seconds. This demonstrates that the complexity is indeed linear in the number of solutions and that the algorithm can efficiently calculate the maximal sequences even for long output sequences.

EstiHMM: an efficient algorithm to determine the maximal sequences

- Principle of optimality
- Deriving an alternative optimality criterion

$$\hat{x}_{1:n} \in \text{opt}(\mathcal{X}_{1:n}|o_{1:n})$$



~~$$(\forall x_{1:n} \in \mathcal{X}_{1:n}) \underline{P}_1(\mathbb{I}_{\{x_{1:n}\}} - \mathbb{I}_{\{\hat{x}_{1:n}\}}) \leq 0.$$~~

$$\alpha_k^{\text{opt}}(\hat{x}_k|x_{k-1}) \leq \alpha_k(\hat{x}_{k:n}).$$

EstiHMM: an efficient algorithm to determine the maximal sequences

- Principle of optimality
- Deriving an alternative optimality criterion
- **A recursive approach**

Principle of Optimality Using the recursive expressions for the joint model, we can derive an appropriate version of Bellman's Principle of Optimality:

$$\hat{x}_{k:n} \in \text{opt}(\mathcal{X}_{k:n}|x_{k-1}, o_{k:n}) \Rightarrow \hat{x}_{k+1:n} \in \text{opt}(\mathcal{X}_{k+1:n}|\hat{x}_k, o_{k+1:n}),$$

which in turn implies that

$$\text{opt}(\mathcal{X}_{k:n}|x_{k-1}, o_{k:n}) \subseteq \bigcup_{x_k \in \mathcal{X}_k} x_k \oplus \text{opt}(\mathcal{X}_{k+1:n}|x_k, o_{k+1:n}).$$

An alternative criterion If we limit ourselves to the possible sequences selected by the Principle of Optimality, we can reformulate the optimality criterion as follows:

$$\hat{x}_{k:n} \in \text{opt}(\mathcal{X}_{k:n}|x_{k-1}, o_{k:n}) \Leftrightarrow \alpha_k^{\text{opt}}(\hat{x}_k|x_{k-1}) \leq \alpha_k(\hat{x}_{k:n}).$$

Backward-forward recursion

We let k run *backward* from n to 1 . For each k and every $x_{k-1} \in \mathcal{X}_{k-1}$, we determine $\text{opt}(\mathcal{X}_k|x_{k-1}, o_{k:n})$ with our alternative optimality criterion. We prove that we can do this efficiently by executing the *forward running procedure* demonstrated in the figure below. The sequences of **green nodes** are the elements of $\text{opt}(\mathcal{X}_k|x_{k-1}, o_{k:n})$. In this binary example, $\mathcal{X} = \{0, 1\}$.

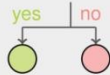
repeat for i from k to n until **no**:

$$\alpha_k^{\text{max}}(\hat{x}_{k:i}) \geq \alpha_k^{\text{opt}}(\hat{x}_k|x_{k-1})?$$

if **yes**

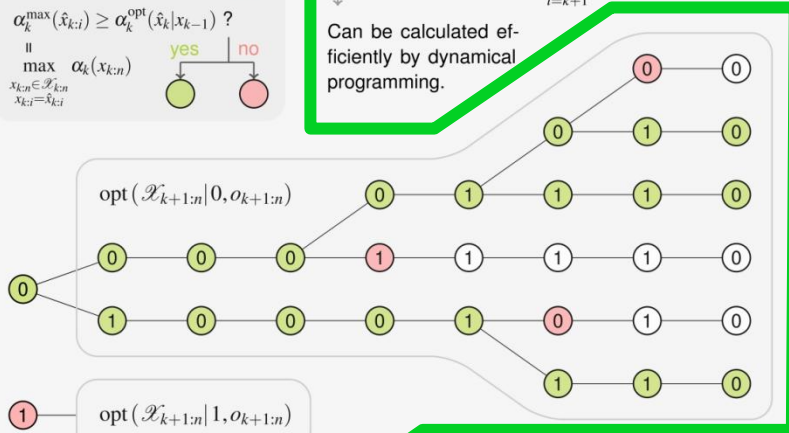
$$\max_{x_{k:n} \in \mathcal{X}_{k:n}} \alpha_k(x_{k:n})$$

$x_{k:n} = \hat{x}_{k:n}$



Can be calculated efficiently by dynamical programming.

$$\bar{S}_k(\{o_k\}|\hat{x}_k) \prod_{i=k+1}^n \bar{S}_i(\{o_i\}|\hat{x}_i) \bar{Q}_i(\{\hat{x}_i\}|\hat{x}_{i-1})$$



The EstiHMM algorithm

Computational complexity

Theoretical analysis We prove that the computational complexity is at worst quadratic in the length of the Markov chain, cubic in the number of states, and roughly speaking linear in the number of maximal sequences (each backward step in the backward-forward loop has a linear complexity in the number of maximal elements at that stage).

Empirical confirmation In order to demonstrate that our algorithm is indeed quite efficient, we let it determine the maximal sequences for a random output sequence of length 100. The iHMM we use to determine the maximal sequences is generated by mixing precise local models with a vacuous one, using a mixture coefficient ϵ . For $\epsilon = 2\%$, there are 5 maximal sequences and it takes 0.2 seconds to calculate them. If we let ϵ grow to 5%, we get 764 maximal sequences and these can be determined in 32 seconds. This demonstrates that the complexity is indeed linear in the number of solutions and that the algorithm can efficiently calculate the maximal sequences even for long output sequences.

EstiHMM: an efficient algorithm to determine the maximal sequences

- Principle of optimality
- Deriving an alternative optimality criterion
- A recursive approach

Complexity

- Theoretical analysis

Principle of Optimality Using the recursive expressions for the joint model, we can derive an appropriate version of Bellman's Principle of Optimality:

$$\hat{x}_{k:n} \in \text{opt}(\mathcal{X}_{k:n}|x_{k-1}, o_{k:n}) \Rightarrow \hat{x}_{k+1:n} \in \text{opt}(\mathcal{X}_{k+1:n}|\hat{x}_k, o_{k+1:n}),$$

which in turn implies that

$$\text{opt}(\mathcal{X}_{k:n}|x_{k-1}, o_{k:n}) \subseteq \bigcup_{x_k \in \mathcal{X}_k} x_k \oplus \text{opt}(\mathcal{X}_{k+1:n}|x_k, o_{k+1:n}).$$

An alternative criterion If we limit ourselves to the possible sequences selected by the Principle of Optimality, we can reformulate the optimality criterion as follows:

$$\hat{x}_{k:n} \in \text{opt}(\mathcal{X}_{k:n}|x_{k-1}, o_{k:n}) \Leftrightarrow \alpha_k^{\text{opt}}(\hat{x}_k|x_{k-1}) \leq \alpha_k(\hat{x}_{k:n}).$$

Backward-forward recursion

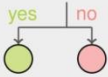
We let k run *backward* from n to 1 . For each k and every $x_{k-1} \in \mathcal{X}_{k-1}$, we determine $\text{opt}(\mathcal{X}_k|x_{k-1}, o_{k:n})$ with our alternative optimality criterion. We prove that we can do this efficiently by executing the *forward running procedure* demonstrated in the figure below. The sequences of **green nodes** are the elements of $\text{opt}(\mathcal{X}_k|x_{k-1}, o_{k:n})$. In this binary example, $\mathcal{X} = \{0, 1\}$.

repeat for i from k to n until **no**:

$$\alpha_k^{\text{max}}(\hat{x}_{k:i}) \geq \alpha_k^{\text{opt}}(\hat{x}_k|x_{k-1}) ?$$

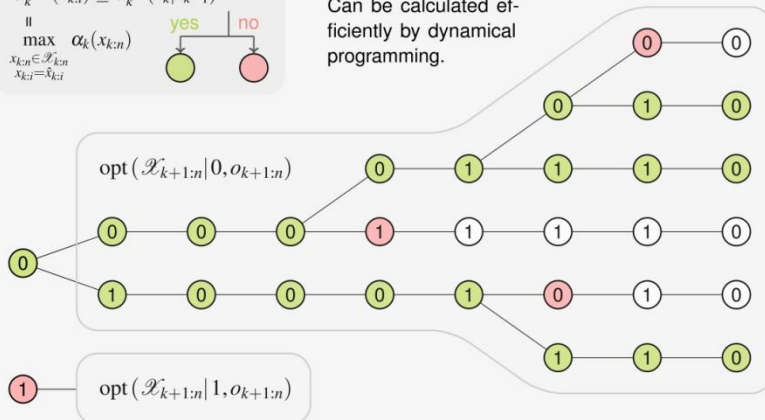
$$\text{if } \max_{x_{k:n} \in \mathcal{X}_{k:n}} \alpha_k(x_{k:n})$$

$$x_{k:n} = \hat{x}_{k:n}$$



Can be calculated efficiently by dynamical programming.

$$\bar{S}_k(\{o_k\}|\hat{x}_k) \prod_{i=k+1}^n \bar{S}_i(\{o_i\}|\hat{x}_i) \bar{Q}_i(\{\hat{x}_i\}|\hat{x}_{i-1})$$



The EstiHMM algorithm

Computational complexity

Theoretical analysis We prove that the computational complexity is at worst quadratic in the length of the Markov chain, cubic in the number of states, and roughly speaking linear in the number of maximal sequences (each backward step in the backward-forward loop has a linear complexity in the number of maximal elements at that stage).

Empirical confirmation In order to demonstrate that our algorithm is indeed quite efficient, we let it determine the maximal sequences for a random output sequence of length 100. The iHMM we use to determine the maximal sequences is generated by mixing precise local models with a vacuous one, using a mixture coefficient ϵ . For $\epsilon = 2\%$, there are 5 maximal sequences and it takes 0.2 seconds to calculate them. If we let ϵ grow to 5%, we get 764 maximal sequences and these can be determined in 32 seconds. This demonstrates that the complexity is indeed linear in the number of solutions and that the algorithm can efficiently calculate the maximal sequences even for long output sequences.

Principle of Optimality Using the recursive expressions for the joint model, we can derive an appropriate version of Bellman's Principle of Optimality:

$$\hat{x}_{k:n} \in \text{opt}(\mathcal{X}_{k:n}|x_{k-1}, o_{k:n}) \Rightarrow \hat{x}_{k+1:n} \in \text{opt}(\mathcal{X}_{k+1:n}|\hat{x}_k, o_{k+1:n}),$$

which in turn implies that

$$\text{opt}(\mathcal{X}_{k:n}|x_{k-1}, o_{k:n}) \subseteq \bigcup_{x_k \in \mathcal{X}_k} x_k \oplus \text{opt}(\mathcal{X}_{k+1:n}|x_k, o_{k+1:n}).$$

An alternative criterion If we limit ourselves to the possible sequences selected by the Principle of Optimality, we can reformulate the optimality criterion as follows:

$$\hat{x}_{k:n} \in \text{opt}(\mathcal{X}_{k:n}|x_{k-1}, o_{k:n}) \Leftrightarrow \alpha_k^{\text{opt}}(\hat{x}_k|x_{k-1}) \leq \alpha_k(\hat{x}_{k:n}).$$

Backward-forward recursion

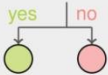
We let k run *backward* from n to 1 . For each k and every $x_{k-1} \in \mathcal{X}_{k-1}$, we determine $\text{opt}(\mathcal{X}_k|x_{k-1}, o_{k:n})$ with our alternative optimality criterion. We prove that we can do this efficiently by executing the *forward running procedure* demonstrated in the figure below. The sequences of **green nodes** are the elements of $\text{opt}(\mathcal{X}_k|x_{k-1}, o_{k:n})$. In this binary example, $\mathcal{X} = \{0, 1\}$.

repeat for i from k to n until **no**:

$$\alpha_k^{\text{max}}(\hat{x}_{k:i}) \geq \alpha_k^{\text{opt}}(\hat{x}_k|x_{k-1}) ?$$

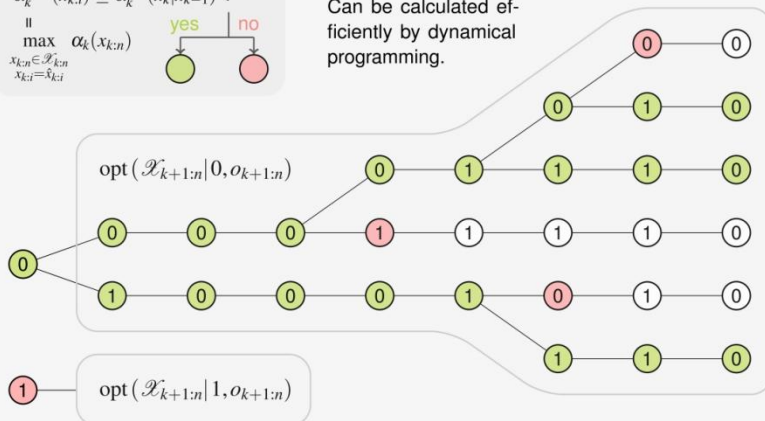
$$\text{if } \max_{x_{k:n} \in \mathcal{X}_{k:n}} \alpha_k(x_{k:n})$$

$$x_{k:n} = \hat{x}_{k:n}$$



Can be calculated efficiently by dynamical programming.

$$\bar{S}_k(\{o_k\}|\hat{x}_k) \prod_{i=k+1}^n \bar{S}_i(\{o_i\}|\hat{x}_i) \bar{Q}_i(\{\hat{x}_i\}|\hat{x}_{i-1})$$



The EstiHMM algorithm

Computational complexity

Theoretical analysis We prove that the computational complexity is at worst quadratic in the length of the Markov chain, cubic in the number of states, and roughly speaking linear in the number of maximal sequences (each backward step in the backward-forward loop has a linear complexity in the number of maximal elements at that stage).

Empirical confirmation In order to demonstrate that our algorithm is indeed quite efficient, we let it determine the maximal sequences for a random output sequence of length 100. The iHMM we use to determine the maximal sequences is generated by mixing precise local models with a vacuous one, using a mixture coefficient ϵ . For $\epsilon = 2\%$, there are 5 maximal sequences and it takes 0.2 seconds to calculate them. If we let ϵ grow to 5%, we get 764 maximal sequences and these can be determined in 32 seconds. This demonstrates that the complexity is indeed linear in the number of solutions and that the algorithm can efficiently calculate the maximal sequences even for long output sequences.

EstiHMM: an efficient algorithm to determine the maximal sequences

- Principle of optimality
- Deriving an alternative optimality criterion
- A recursive approach

Complexity

- Theoretical analysis
- **Linear in the number of maximal sequences**

Principle of Optimality Using the recursive expressions for the joint model, we can derive an appropriate version of Bellman's Principle of Optimality:

$$\hat{x}_{k:n} \in \text{opt}(\mathcal{X}_{k:n}|x_{k-1}, o_{k:n}) \Rightarrow \hat{x}_{k+1:n} \in \text{opt}(\mathcal{X}_{k+1:n}|\hat{x}_k, o_{k+1:n}),$$

which in turn implies that

$$\text{opt}(\mathcal{X}_{k:n}|x_{k-1}, o_{k:n}) \subseteq \bigcup_{x_k \in \mathcal{X}_k} x_k \oplus \text{opt}(\mathcal{X}_{k+1:n}|x_k, o_{k+1:n}).$$

An alternative criterion If we limit ourselves to the possible sequences selected by the Principle of Optimality, we can reformulate the optimality criterion as follows:

$$\hat{x}_{k:n} \in \text{opt}(\mathcal{X}_{k:n}|x_{k-1}, o_{k:n}) \Leftrightarrow \alpha_k^{\text{opt}}(\hat{x}_k|x_{k-1}) \leq \alpha_k(\hat{x}_{k:n}).$$

Backward-forward recursion

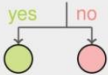
We let k run *backward* from n to 1 . For each k and every $x_{k-1} \in \mathcal{X}_{k-1}$, we determine $\text{opt}(\mathcal{X}_k|x_{k-1}, o_{k:n})$ with our alternative optimality criterion. We prove that we can do this efficiently by executing the *forward running procedure* demonstrated in the figure below. The sequences of **green nodes** are the elements of $\text{opt}(\mathcal{X}_k|x_{k-1}, o_{k:n})$. In this binary example, $\mathcal{X} = \{0, 1\}$.

repeat for i from k to n until **no**:

$$\alpha_k^{\text{max}}(\hat{x}_{k:i}) \geq \alpha_k^{\text{opt}}(\hat{x}_k|x_{k-1}) ?$$

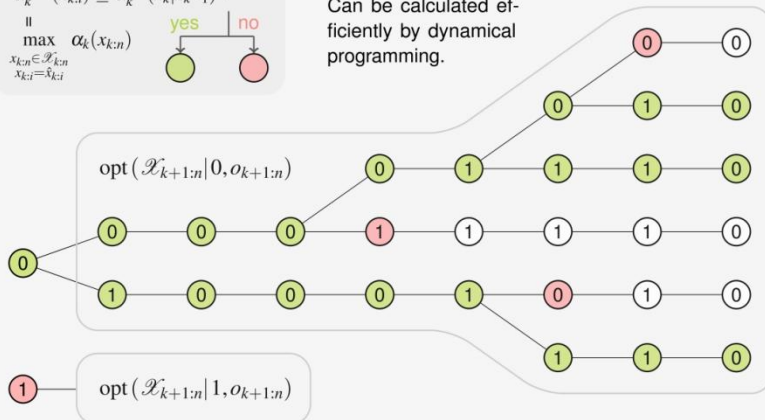
$$\text{if } \max_{x_{k:n} \in \mathcal{X}_{k:n}} \alpha_k(x_{k:n})$$

$$x_{k:n} = \hat{x}_{k:n}$$



Can be calculated efficiently by dynamical programming.

$$\bar{S}_k(\{o_k\}|\hat{x}_k) \prod_{i=k+1}^n \bar{S}_i(\{o_i\}|\hat{x}_i) \bar{Q}_i(\{\hat{x}_i\}|\hat{x}_{i-1})$$



The EstiHMM algorithm

Computational complexity

Theoretical analysis We prove that the computational complexity is at worst quadratic in the length of the Markov chain, cubic in the number of states, and roughly speaking linear in the number of maximal sequences (each backward step in the backward-forward loop has a linear complexity in the number of maximal elements at that stage).

Empirical confirmation In order to demonstrate that our algorithm is indeed quite efficient, we let it determine the maximal sequences for a random output sequence of length 100. The iHMM we use to determine the maximal sequences is generated by mixing precise local models with a vacuous one, using a mixture coefficient ϵ . For $\epsilon = 2\%$, there are 5 maximal sequences and it takes 0.2 seconds to calculate them. If we let ϵ grow to 5%, we get 764 maximal sequences and these can be determined in 32 seconds. This demonstrates that the complexity is indeed linear in the number of solutions and that the algorithm can efficiently calculate the maximal sequences even for long output sequences.

EstiHMM: an efficient algorithm to determine the maximal sequences

- Principle of optimality
- Deriving an alternative optimality criterion
- A recursive approach

Complexity

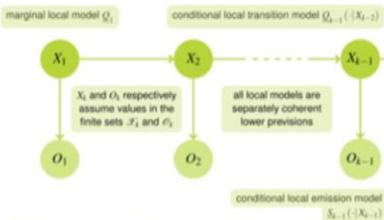
- Theoretical analysis
- Linear in the number of maximal sequences
- **Empirical confirmation**

State sequence prediction in imprecise hidden Markov models

Interpretation of the graphical model

Interpretation of the graphical structure Our imprecise hidden Markov model (iHMM) represents the following irrelevance assessments: conditional on its mother variable, the non-parent non-descendants of any variable in the tree are epistemically irrelevant to this variable and its descendants.

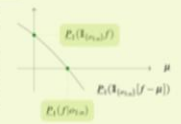
Epistemic irrelevance Y is irrelevant to X whenever the belief model (lower prevision P) about X does not change when we learn something about Y :
 $(\forall x \in \mathcal{X}(\mathcal{Y}))(\forall y \in \mathcal{Y}) P(x) = P(x|y)$.



Maximal state sequences

Conditioning the joint model

Since we assume that all local lower probabilities are strictly positive (in recent work, we dropped this assumption), $P_i(\{o_{i+1}\}) > 0$ and the Generalised Bayes Rule yields a uniquely coherent value of $P_i(f|o_{i+1})$, which has (this is very useful) the same sign as $P_i(\{i_{i+1}|f\})$.



Optimality criterion

We can express a strict preference \succ between state sequences $i_{1:n}$ and $i'_{1:n}$ as follows: $i_{1:n} \succ i'_{1:n} \Leftrightarrow P_i(\{i_{1:n}\} - \{i'_{1:n}\} | o_{1:n}) > 0$. This induces a strict partial order \succ on the set of state sequences $\mathcal{I}_{1:n}$, and we consider a sequence $i_{1:n}$ to be optimal when it is undominated, or maximal, in this strict partial order:

$$i_{1:n} \in \text{opt}(\mathcal{I}_{1:n} | o_{1:n}) \Leftrightarrow (\forall i'_{1:n} \in \mathcal{I}_{1:n}) i_{1:n} \not\succeq i'_{1:n}$$

$$\Leftrightarrow (\forall i'_{1:n} \in \mathcal{I}_{1:n}) P_i(\{i_{1:n}\} - \{i'_{1:n}\} | o_{1:n}) \leq 0$$

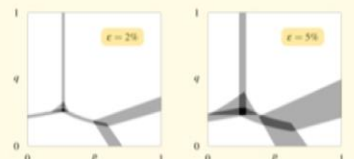
$$\Leftrightarrow (\forall i'_{1:n} \in \mathcal{I}_{1:n}) P_i(\{i'_{1:n}\} - \{i_{1:n}\} | o_{1:n}) \leq 0$$

In an analogous manner we define the optimal subsequences:

$$i_{k:n} \in \text{opt}(\mathcal{I}_{k:n} | o_{k:n}) \Leftrightarrow (\forall i'_{k:n} \in \mathcal{I}_{k:n}) P_i(\{i_{k:n}\} - \{i'_{k:n}\} | o_{k:n}) \leq 0$$

A first experiment

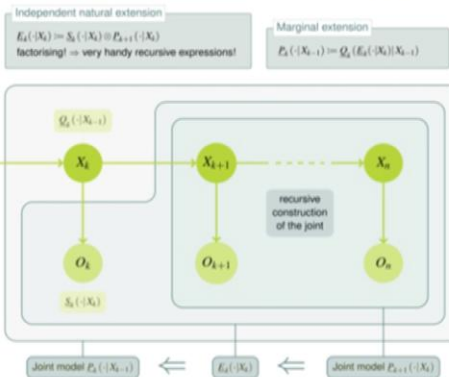
Motivation and description No algorithm, however cleverly designed, will be able to find all maximal sequences efficiently if there are too many. Because this number of maximal sequences is so important, we study its behaviour in more detail. We consider a binary, stationary iHMM with precise emission models. The imprecise marginal and transition models are generated by mixing precise models with a vacuous one, using a mixture coefficient ϵ . For a particular observation sequence of length three, we plot the number of maximal sequences as a function of the transition probabilities p and q . As this number grows from 1 to 4 the areas go from white to black.



Results We see that there are large regions of transition probability space where the number of maximal elements remains fairly small. The plots also display quite interesting behaviour. If we let the imprecision grow, by using higher ϵ , the areas with multiple maximal sequences become larger. They are expanded versions of the lines of indifference that occur in the precise case.

Jasper De Bock and Gert de Cooman

{Jasper.DeBock, Gert.deCooman}@UGent.be
 SYSTEms, Ghent University, Belgium



Principle of Optimality Using the recursive expressions for the joint model, we can derive an appropriate version of Bellman's Principle of Optimality:
 $i_{k:n} \in \text{opt}(\mathcal{I}_{k:n} | o_{k:n}) \Leftrightarrow i_{k+1:n} \in \text{opt}(\mathcal{I}_{k+1:n} | o_{k+1:n})$, which in turn implies that
 $\text{opt}(\mathcal{I}_{k:n} | o_{k:n}) \subseteq \bigcup_{o_{k+1}} \text{opt}(\mathcal{I}_{k+1:n} | o_{k+1:n})$.

An alternative criterion If we limit ourselves to the possible sequences selected by the Principle of Optimality, we can reformulate the optimality criterion as follows:
 $i_{k:n} \in \text{opt}(\mathcal{I}_{k:n} | o_{k:n}) \Leftrightarrow \alpha_k^{opt}(i_k | o_k) \leq \alpha_k(i_k)$.

Backward-forward recursion We let k run backward from n to 1. For each k and every $i_{k+1:n} \in \mathcal{I}_{k+1:n}$, we determine $\text{opt}(\mathcal{I}_{k+1:n} | o_{k+1:n})$ with our alternative optimality criterion. We prove that we can do this efficiently by executing the forward running procedure demonstrated in the figure below. The sequences of $\text{opt}(\mathcal{I}_{k+1:n} | o_{k+1:n})$ are the elements of $\text{opt}(\mathcal{I}_{k+1:n} | o_{k+1:n})$ in this binary example, $\mathcal{X} = \{0, 1\}$.

repeat for l from k to n until **end**
 $\alpha_k^{opt}(i_k) \geq \alpha_k^{opt}(i_k | i_{k+1:n})$?
 $\alpha_k(i_k) \geq \alpha_k(i_k)$?
 Can be calculated efficiently by dynamical programming.

$\bar{\alpha}_k(\{o_k\} | i_k) = \prod_{t=k+1}^n \bar{\alpha}_t(\{o_t\} | i_t | i_{t+1:n})$

$\text{opt}(\mathcal{I}_{k+1:n} | o_{k+1:n})$

$\text{opt}(\mathcal{I}_{k+1:n} | o_{k+1:n})$

The EstiHMM algorithm

Computational complexity

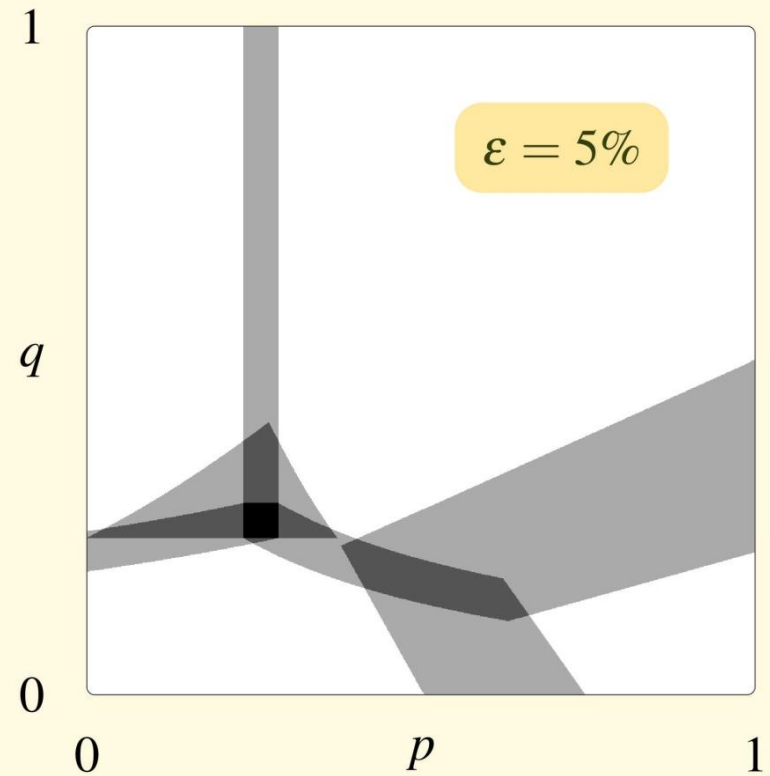
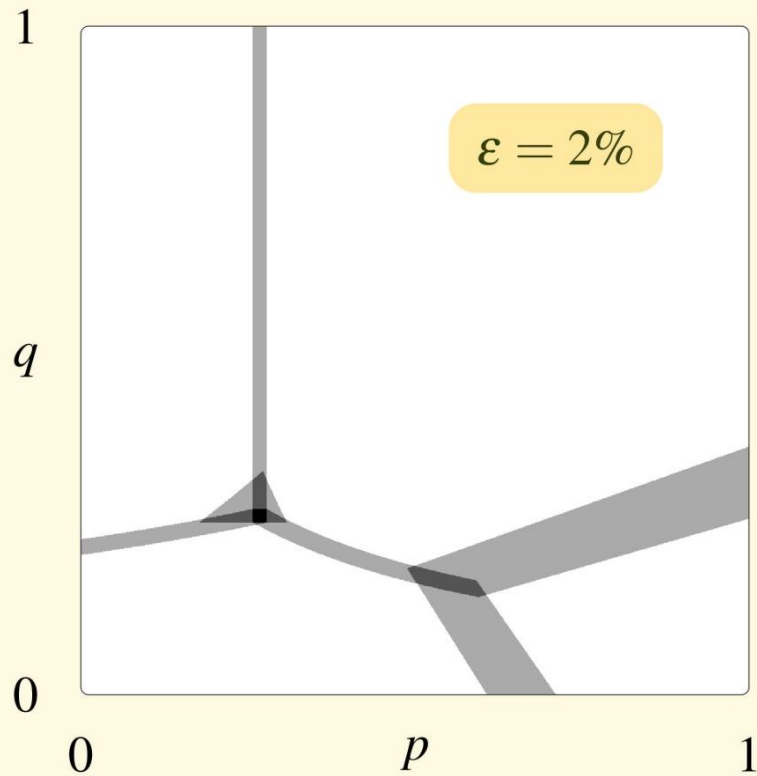
Theoretical analysis We prove that the computational complexity is at worst quadratic in the length of the Markov chain, cubic in the number of states, and roughly speaking linear in the number of maximal sequences (each backward step in the backward-forward loop has a linear complexity in the number of maximal elements at that stage).

Empirical confirmation In order to demonstrate that our algorithm is indeed quite efficient, we let it determine the maximal sequences for a random output sequence of length 100. The iHMM we use to determine the maximal sequences is generated by mixing precise local models with a vacuous one, using a mixture coefficient ϵ . For $\epsilon = 2\%$, there are 5 maximal sequences and it takes 0.2 seconds to calculate them. If we let ϵ grow to 5%, we get 764 maximal sequences and these can be determined in 32 seconds. This demonstrates that the complexity is indeed linear in the number of solutions and that the algorithm can efficiently calculate the maximal sequences even for long output sequences.

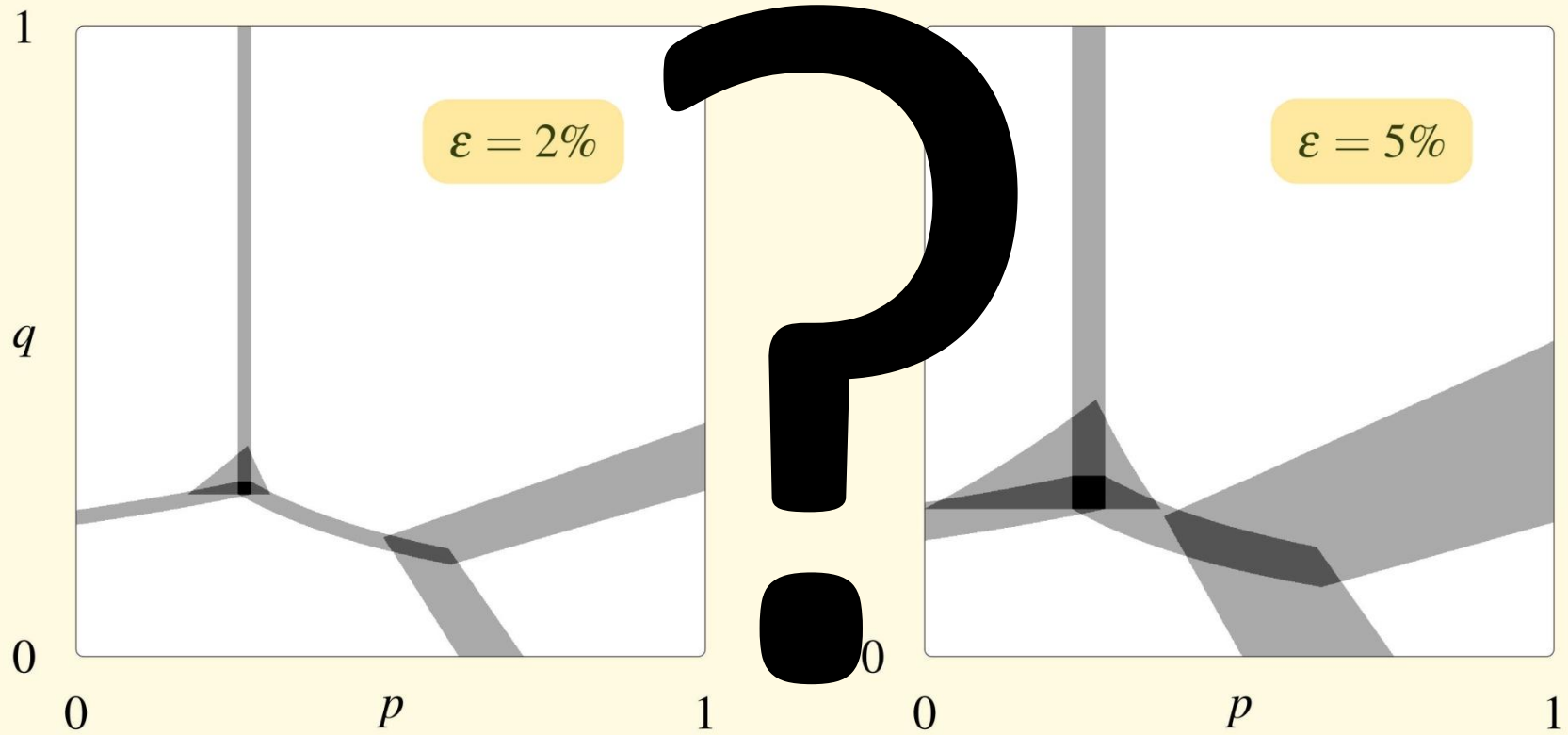
State sequence prediction in imprecise hidden Markov models

A first experiment

A first experiment



A first experiment



A first experiment

