# GRAPHICAL MARKOV MODELS
# EXAMINATION WS2010 (25+6P)

**Assignment 1. (4p)**
A Markov model for $K$-valued sequences $s \in K^n$ of length $n$ is given by the probability distribution $p(s_1)$ for the first state of the sequence and the transition probabilities $p(s_i|s_{i-1})$, $i = 2, \ldots, n$. Assume that the set of states $K$ is completely ordered and define $M$ to be the set of all non-decreasing sequences

$$M = \big\{ s \in K^n \;\big|\; s_i \geqslant s_{i-1} \;, \forall i = 2, \ldots, n \big\}$$

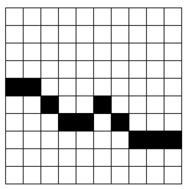
Describe an efficient algorithm for calculating the probability of $p(M) = \sum_{s \in M} p(s)$. What complexity has it?

**Assignment 2. (8p)**
Consider a Hidden Markov model for pairs of sequences $x = (x_1, \ldots, x_n)$ and $s = (s_1, \ldots, s_n)$ given by its parameters $p(s_1)$, $p(s_i|s_{i-1})$ and $p(x_i|s_i)$.
**a)** What is the optimal Bayes decision for the sequence of hidden states $s$ given the observed sequence $x$ if the loss function is the Hamming distance

$$C(s, s') = \sum_{i=1}^{n} \big[ 1 - \delta(s_i, s'_i) \big],$$

where $\delta$ is the Kronecker symbol? Explain an efficient algorithm for solving this task.
**b)** Suppose now, that the model is not strictly positive – some transitions $k \mapsto k'$ have zero probability for certain sequence positions $i$. How does the optimal Bayes decision change if it is required in addition, that the inferred sequence of hidden states should have non zero probability? Describe an efficient algorithm for this decision and give its complexity.

**Assignment 3. (8p)**
Consider the following image language $\mathcal{L}$ for rectangular b/w images. An image belongs to $\mathcal{L}$ if it contains a one pixel wide black boundary which separates a white region below the boundary from a white region above it. The boundary itself should be 8-connected (see figure).



**a)** Express the language $\mathcal{L}$ by a local conjunctive predicate defined for appropriately chosen non-terminal symbols. Explain the graphical structure and describe the local predicates.

**b)** Suppose you are given a grey-valued image $x$ with grey values in the range $[0, 1]$ and the task is to interpret it in terms of a b/w image $s \in \mathcal{L}$. The corresponding cost is additive:

$$C(x, s) = \sum_{i \in D} |x_i - s_i|,$$

where $D$ is the image domain and $s_i$ has values 0 and 1 for black and white respectively. The task is to find the interpretation $s \in \mathcal{L}$ with the lowest cost. Use the non-terminal symbols and the local conjunctive predicate from a) and formulate this task as a (min, +)-problem. Is this task submodular?

**Assignment 4. (5p)**

Suppose you are given two finite sets $\mathcal{S}_1$, $\mathcal{S}_2$ of $K$-valued sequences of length $n$. The task is to learn two Markov chain models $p_1$, $p_2$ such that the following holds

$$p_1(s) > p_2(s) \;\; \forall s \in \mathcal{S}_1,$$
$$p_2(s) > p_1(s) \;\; \forall s \in \mathcal{S}_2.$$

Explain how to solve this learning task by the perceptron algorithm.

**Assignment 5. (6p)***

A Markov chain model for $K$-valued sequences of length $n$ is defined by its parameters $p(s_1)$ and $p(s_i|s_{i-1})$. These parameters are unknown and should be estimated given a sample of training data $\mathcal{T}$. The latter consists of *parts* of sequences only: each example $s^j \in \mathcal{T}$ is given for a certain subset of positions $I^j \subset \{1, 2, \ldots, n\}$ i.e. $s^j \colon I^j \to K$. How would you learn the model parameters according to the Maximum Likelihood principle?