Function	Viterbi	Forward-Backward
$a_{ij} = p$ if $\|i-j\| \leq d$, $a_{ij} = 0$ otherwise	Min-filter	Box sum
$a_{ij} \propto \exp(-\|i-j\|^2/2\sigma^2)$	L_2^2 dist. trans.	Gaussian approx.
$a_{ij} \propto \exp(-k\|i-j\|)$	L_1 dist. trans.	FFT
$a_{ij} = p$ if $\|i-j\| \leq d$, $a_{ij} = q$ otherwise	Combin. min-filter	Combin. box sum
$a_{ij} \propto \exp(-\|i-j\|^2/2\sigma^2)$ if $\|i-j\| \leq d$, $a_{ij} \propto \exp(-k\|i-j\|)$ otherwise	Combin. dist. trans.	FFT

Table 1: Some transition probabilities that can be handled efficiently using our techniques (see text for an explanation). All running times are $O(Tn)$ except those using the FFT which are $O(Tn \log n)$.

that the initial state of the system is s_i.

Let q_t denote the state of the system at time t, while o_t denotes the observed symbol at time t. Given a sequence of observations $O = (o_1, \ldots, o_T)$ there are three standard estimation (or inference) problems that have wide applications:

1. Find a state sequence $Q = (q_1, \ldots, q_T)$ maximizing $P(Q|O, \lambda)$.
2. Compute $P(O|\lambda)$, the probability of an observation sequence being generated by λ.
3. Compute the posterior probabilities of each state, $P(q_t = s_i|O, \lambda)$.

As is well known these problems can be solved in $O(Tn^2)$ time using the Viterbi algorithm for the first task and the forward-backward algorithm for the others. We show how to solve them more efficiently for a wide range of transition probabilities based on *differences* between states that are embedded in an underlying grid. This grid can be multi-dimensional, however in this paper we consider only the one-dimensional case. Table 1 lists some widely applicable transition probability distributions that can be handled by our methods. The algorithms for each distribution differ slightly and are explained in the subsequent sections. The distributions given in the bottom part of the table can be computed as combinations of the basic distributions in the top part. Other distributions can be obtained using these same combination techniques, as long as only a constant number of distributions are being combined.

An additional problem, which we do not explicitly consider here, is that of determining the best model λ given some set of observed sequences $\{O_1, \ldots, O_l\}$. However the most widely used technique for solving this problem, expectation maximization (EM), requires repeatedly running the forward-backward algorithm. Thus our algorithms also indirectly make the model learning problem more efficient.

2.1 Viterbi Algorithm

The Viterbi algorithm is used to find a maximum posterior probability state sequence, that is a sequence $Q = (q_1, \ldots, q_T)$ maximizing $P(Q|O, \lambda)$. The main computation is to determine the highest probability along a path, accounting for the observations and ending in a given state. While there are an exponential number of possible paths, the Viterbi algorithm uses a dynamic programming approach