Approximate Inference in Hidden Markov Models Using Iterative Active State Selection

C. M. Vithanage, *Student Member, IEEE*, C. Andrieu, and R. J. Piechocki, *Member, IEEE*

***Abstract*—The inferential task of computing the marginal posterior probability mass functions of state variables and pairs of consecutive state variables of a hidden Markov model is considered. This can be exactly and efficiently performed using a message passing scheme such as the Bahl–Cocke–Jelinek–Raviv (BCJR) algorithm. We present a novel iterative reduced complexity variation of the BCJR algorithm that uses reduced support approximations for the forward and backward messages, as in the M-BCJR algorithm. Forward/backward message computation is based on the concept of expectation propagation, which results in an algorithm similar to the M-BCJR algorithm with the active state selection criterion being changed from the filtered distribution of state variables to beliefs of state variables. By allowing possibly different supports for the forward and backward messages, we derive identical forward and backward recursions that can be iterated. Simulation results of application for trellis-based equalization of a wireless communication system confirm the improved performance over the M-BCJR algorithm.**

***Index Terms*—Deterministic algorithms, equalizers, hidden Markov models (HMMs), message passing, state space methods.**

I. INTRODUCTION

COMPUTATION of marginal posterior probability mass functions of state variables and/or pairs of consecutive state variables of a hidden Markov model (HMM) is necessary in many applications, such as optimal symbol detection in frequency-selective fading channels and training of an HMM-based speech recognizer. These can be efficiently computed using a message passing scheme such as the sum-product algorithm [1] working on the corresponding factor graph or equivalently as used in communications research, using the Bahl–Cocke–Jelinek–Raviv (BCJR) algorithm [2]. Though efficient, the BCJR algorithm is still intractable for real-time application in many practical situations. One method of reducing the complexity of this optimal algorithm at the expense of some reduction in performance is to allow the forward and backward messages to be nonzero only for some configurations (say M) of the state variables, as in the M-BCJR algorithm of [3]. The M-BCJR algorithm selects the support of the forward and backward messages on the state variables using approximations for the forward messages, which in turn are related to the filtered probability distributions of the state variables. In this letter, we apply the concept of expectation propagation (EP) (originally proposed in [4] and also applied for signal detection in [5]) as presented for dynamic Bayesian networks in [6] to build up simplified approximate forward/backward messages, which is analogous to building up the messages as in the forward recursion of the M-BCJR algorithm with a different active state selection criterion. We will present the proposed algorithm in the context of equalization of frequency-selective channels in the following sections. Specifically, we will consider a turbo equalizer where the iterations of forward backward recursions of the proposed algorithm can be naturally incorporated into the turbo equalization iterations.

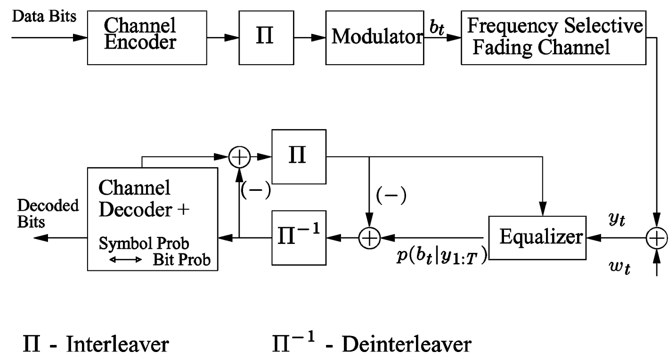

Fig. 1. System model.

II. SYSTEM MODEL

For simplicity, let us consider a coded single transmit antenna–single receive antenna wireless communication system, transmitting into a frequency-selective channel with memory L, as shown in Fig. 1.

Let the fading coefficient of the dth channel tap be denoted by $h(d)$ for $d \in \{0, \ldots, L\}$. Let us also denote the received signal and the transmitted antenna symbol at the tth time instant by y_t and b_t, respectively, and assume that each symbol b_t is chosen from set B. The received signal y_t consists of the convolution of the channel impulse response and a sequence of symbols transmitted up to $L+1$ time instants and an additive white Gaussian noise w_t

$$y_t = \sum_{d=0}^{L} h(d) b_{t-d} + w_t. \quad (1)$$

We use the notation $v_{t:t'}$ to refer to the sequence of some time-indexed variables or sets from time t to t' (with $t \leq t'$) and $(\bullet)^\dagger$ to denote the vector transpose operation. Within the framework of turbo equalization, the task of the optimal equalizer at each turbo iteration is to consider the channel output sequence $y_{1:T}$ (where T is the frame length) and

Manuscript received June 17, 2005; revised September 13, 2005. This work was supported by Toshiba Research Europe Limited, Bristol, U.K. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Hongbin Li.

C. M. Vithanage and C. Andrieu are with the Department of Mathematics, University of Bristol, Bristol BS8 1TW, U.K. (e-mail: c.vithanage@bristol.ac.uk; c.andrieu@bristol.ac.uk).

R. J. Piechocki is with the Centre for Communications Research, University of Bristol, Bristol BS8 1UB, U.K. (e-mail: r.j.piechocki@bristol.ac.uk).

Digital Object Identifier 10.1109/LSP.2005.861600

some prior distribution on b_t (for $t \in \{1,\ldots,T\}$) provided by the extrinsic output of the channel decoder and compute the marginal posterior mass functions $p(b_t|y_{1:T})$ for each $t \in \{1,\ldots,T\}$. For this task, it is convenient to define a state variable consisting of candidates for transmitted symbols of L time instants as $\mathbf{s}_t = (b_t,\ldots,b_{t-L+1})^\dagger$. We will use the terms *state* of $\mathbf{s}_t$ and *support* of $p(\mathbf{s}_t|\bullet)$ to indicate a possible configuration of $\mathbf{s}_t$ and the set of configurations of $\mathbf{s}_t$ such that $p(\mathbf{s}_t|\bullet) > 0$, respectively. Thus, each $\mathbf{s}_t$ takes values from a set $S = B^L$ and has, say, $|S|$ possible states. Taking $\mathbf{s}_0$ to be an initial state variable, it can be observed now that we have the conditional independences $p(\mathbf{s}_t|\mathbf{s}_{0:t}) = p(\mathbf{s}_t|\mathbf{s}_{t-1})$ and $p(y_t|\mathbf{s}_{1:t}) = p(y_t|\mathbf{s}_{t-1},\mathbf{s}_t)$ for each $t \in \{1,\ldots,T\}$. Thus, the variables concerned can be represented by an HMM. It can also be seen that the marginal posterior distributions of the transmitted symbols $p(b_t|y_{1:T})$ can be derived directly from the distributions of $p(\mathbf{s}_{t-1},\mathbf{s}_t|y_{1:T})$ for each $t \in \{1,\ldots,T\}$. Thus, the problem of optimal equalization can be formulated as the problem of efficient computation of the marginal posterior distributions of pairs of consecutive state variables of an HMM, and we can apply the BCJR algorithm, as well as reduced complexity variants such as M-BCJR and EP, to obtain the exact and approximate solutions as follows.

A. BCJR, M-BCJR, and EP Algorithms

The BCJR algorithm involves a forward recursion through the Markov chain of the state variables, where the forward messages $\alpha(\mathbf{s}_t)$ are recursively computed by $\alpha(\mathbf{s}_t) = \sum_{\mathbf{s}_{t-1}\in S} \alpha(\mathbf{s}_{t-1})\gamma(\mathbf{s}_{t-1},\mathbf{s}_t)$, with $\gamma(\mathbf{s}_{t-1},\mathbf{s}_t) = p(\mathbf{s}_t|\mathbf{s}_{t-1})p(y_t|\mathbf{s}_{t-1},\mathbf{s}_t)$, and a similar backward recursion in which the backward messages $\beta(\mathbf{s}_t)$ are recursively computed as $\beta(\mathbf{s}_t) = \sum_{\mathbf{s}_{t+1}\in S} \beta(\mathbf{s}_{t+1})\gamma(\mathbf{s}_t,\mathbf{s}_{t+1})$. A final combination of the computed messages results in the desired probability mass functions

$$p(\mathbf{s}_t|y_{1:T}) \propto \alpha(\mathbf{s}_t)\beta(\mathbf{s}_t) \tag{2}$$

$$p(\mathbf{s}_{t-1},\mathbf{s}_t|y_{1:T}) \propto \alpha(\mathbf{s}_{t-1})\gamma(\mathbf{s}_{t-1},\mathbf{s}_t)\beta(\mathbf{s}_t). \tag{3}$$

The M-BCJR algorithm is a reduced complexity variation of this, where the forward messages of each time t are computed for $M(<|S|)$ possible states (which are termed as the active states) only. The messages of the inactive states are set to zero. Let us use the notation Ω_t to denote the set of selected active states of time t. At each time instant t, given the approximate forward messages of time $t-1$ as $\tilde{\alpha}(\mathbf{s}_{t-1}); \mathbf{s}_{t-1} \in \Omega_{t-1}$, the candidate set for active states of time t (say, Ξ_t) is given by the union of the supports in $\mathbf{s}_t$ in $\gamma(\mathbf{s}_{t-1},\mathbf{s}_t)$ for $\mathbf{s}_{t-1} \in \Omega_{t-1}$. The cardinality of Ξ_t, $|\Xi_t| \leq |S|$, with the inequality being the usual case. The M-BCJR algorithm computes temporary forward messages for states in Ξ_t as

$$\hat{\alpha}(\mathbf{s}_t) = \sum_{\mathbf{s}_{t-1}\in\Omega_{t-1}} \tilde{\alpha}(\mathbf{s}_{t-1})\gamma(\mathbf{s}_{t-1},\mathbf{s}_t) \tag{4}$$

and selects the M states with the largest messages and makes them the set Ω_t. The messages of states in Ω_t are kept intact, and the other states are considered inactive

$$\begin{cases} \tilde{\alpha}(\mathbf{s}_t) = \hat{\alpha}(\mathbf{s}_t), \mathbf{s}_t \in \Omega_t \\ \tilde{\alpha}(\mathbf{s}_t) = 0, \quad \mathbf{s}_t \notin \Omega_t. \end{cases} \tag{5}$$

This recursive selection of the active states results in the selection of a state set sequence $\Omega_{1:T}$ through the trellis. The backward recursion as well as the final marginal posterior probabilities are only computed for the states in $\Omega_{1:T}$.

The EP algorithm of [6] considers message passing in a general dynamic Bayesian network, i.e., the requirement of the state variables being discrete random variables is relaxed. Its direct formulation for an HMM is as follows. As this algorithm has symmetrical forward and backward recursions, we only consider the forward recursion at time t in the $(i+1)^{st}$ iteration of the algorithm, assuming the availability of the approximated forward messages $\tilde{\alpha}^{i+1}(\mathbf{s}_{t-1})$ and backward messages $\tilde{\beta}^i(\mathbf{s}_t)$. Using (3), we can first build an approximation of the posterior marginal distribution of state variable $\mathbf{s}_t$ (or our belief about the posterior marginal distribution of $\mathbf{s}_t$) as

$$\tilde{p}(\mathbf{s}_t|y_{1:T}) \propto \sum_{\mathbf{s}_{t-1}} \tilde{\alpha}^{i+1}(\mathbf{s}_{t-1})\gamma(\mathbf{s}_{t-1},\mathbf{s}_t)\tilde{\beta}^i(\mathbf{s}_t). \tag{6}$$

Thereafter, EP suggests to approximate $\tilde{p}(\mathbf{s}_t|y_{1:T})$ using some distribution $q(\mathbf{s}_t)$, which belongs to a chosen exponential family of distributions, by minimizing the Kullback–Leibler divergence

$$D(\tilde{p}\|q) = \sum_{\mathbf{s}_t} \tilde{p}(\mathbf{s}_t|y_{1:T}) \log \frac{\tilde{p}(\mathbf{s}_t|y_{1:T})}{q(\mathbf{s}_t)}. \tag{7}$$

Finally, the new forward messages are derived from (2) as

$$\tilde{\alpha}^{i+1}(\mathbf{s}_t) = \frac{q(\mathbf{s}_t)}{\tilde{\beta}^i(\mathbf{s}_t)}. \tag{8}$$

These symmetrical forward and backward recursions are iterated several times until the algorithm reaches a fixed point. If the forward and backward messages are initialized to be in the particular exponential family of distributions to which $q(\mathbf{s}_t)$ belongs, they are ensured to stay within that family for every t. The approximation of the beliefs and the messages by a member of an exponential family (say, by a k-parameter exponential family) allows the corresponding distributions to be expressed using only k parameters and results in the reduction of complexity compared to an exact algorithm.

III. Proposed Algorithm

Two main ideas in the EP algorithm are to select a particular family of distributions from which to choose the approximating distribution $q(\mathbf{s}_t)$ and to choose a distribution from this family by minimizing the divergence $D(\tilde{p}\|q)$. In this algorithm, for the approximating distributions of the beliefs of the state variables, we choose to use discrete distributions with supports of only $M(<|S|)$ states, i.e, given some discrete distribution $\tilde{p}(\mathbf{s}_t|y_{1:T})$ with an arbitrary support, we will approximate it using a discrete distribution $q(\mathbf{s}_t)$, which has a support of a maximum of M elements. Also deviating from the conventional EP algorithm, for the selection of the parameters of $q(\mathbf{s}_t)$, we will minimize the divergence $D(q\|\tilde{p})$. This step is in favor of the resulting computational simplicity and corresponds to α-EP of [7] with $\alpha = -1$. As shown in the Appendix, such a distribution $q(\mathbf{s}_t)$ is easily found by considering the M largest probabilities of the probability mass function $\tilde{p}(\mathbf{s}_t|y_{1:T})$. We also restrict the forward/backward messages to be in the same family of distributions as the beliefs, i.e., discrete distributions with a maximum support of M elements.

This is similar to the active state selection and computation of messages for the active states in the M-BCJR algorithm if we considered our approximations of the posterior marginal distributions to be the criterion for active state selection. M-BCJR simply uses an approximation of the filtered probability distribution of states to make the active state selection. Also, as in the EP algorithm, we seek to have identical forward/backward recursions with possibly different active states (also similar, for example, to [8]), and the recursions of the proposed algorithm are to be iterated for improved performance. Assuming the existence of backward messages ($\tilde{\beta}^i(\mathbf{s}_t)$ for $t \in \{0, \ldots, T\}$) and the set of active states ($\Omega^i_{t,bwd}$ for $t \in \{0, \ldots, T\}$) of the backward recursion of the ith iteration of the algorithm, the forward recursion of the $(i+1)$st iteration proceeds as described below. The backward recursion is identical, and these two recursions are iterated. We are assuming an initialization of $\mathbf{s}_0 = x_0$ and denoting the set of active states of time $t-1$ and the set of states contending to be active states at time t by $\Omega^{i+1}_{t-1,fwd}$ and $\Xi^{i+1}_{t,fwd}$, respectively. The parameter η is a positive minimum set to the forward/backward messages.

$\mathcal{A}$ Set $t = 1$, $\Omega^{i+1}_{0,fwd} = \{x_0\}$, $\tilde{\alpha}^{i+1}(\mathbf{s}_0 = x_0) = 1$ and $\tilde{\alpha}^{i+1}(\mathbf{s}_0 \neq x_0) = 0$.

$\mathcal{B}$ Compute

$$\hat{\alpha}^{i+1}(\mathbf{s}_t) = \sum_{\mathbf{s}_{t-1} \in \Omega^{i+1}_{t-1,fwd}} \tilde{\alpha}^{i+1}(\mathbf{s}_{t-1}) \gamma(\mathbf{s}_{t-1}, \mathbf{s}_t)$$

for $\mathbf{s}_t \in \Xi^{i+1}_{t,fwd}$.

$\mathcal{C}$ Compute

$$\tilde{p}(\mathbf{s}_t | y_{1:T}) \propto \hat{\alpha}^{i+1}(\mathbf{s}_t) \tilde{\beta}^i(\mathbf{s}_t)$$

for $\mathbf{s}_t \in \Xi^{i+1}_{t,fwd}$ with the assumption $\tilde{\beta}^i(\mathbf{s}_t) = \eta$ for $\mathbf{s}_t \in \Xi^{i+1}_{t,fwd} \setminus \Omega^i_{t,bwd}$.

$\mathcal{D}$ Select the M largest components of $\tilde{p}(\mathbf{s}_t | y_{1:T})$ and thereby determine the set $\Omega^{i+1}_{t,fwd}$.

$\mathcal{E}$ Update the forward messages

$$\tilde{\alpha}^{i+1}(\mathbf{s}_t) = \hat{\alpha}^{i+1}(\mathbf{s}_t) \text{ for } \mathbf{s}_t \in \Omega^{i+1}_{t,fwd}$$

$\mathcal{F}$ if $t < T$, increase t by one and go to step $\mathcal{B}$; otherwise, end forward recursion.

The idea of setting the threshold value η for the forward/backward messages of the inactive states is to enable possibly different active state selections in each of the recursions. As shown in the next section, the algorithm is not critically sensitive to the value of η. After the final (say, Ith) iteration, approximation of the marginal posterior probabilities of the state variables and pairs of consecutive state variables can be derived as

$$\tilde{p}(\mathbf{s}_t | y_{1:T}) = \frac{1}{\widehat{Z}} \tilde{\alpha}^{\mathrm{I}}(\mathbf{s}_t) \tilde{\beta}^{\mathrm{I}}(\mathbf{s}_t) \tag{9}$$

$$\tilde{p}(\mathbf{s}_{t-1}, \mathbf{s}_t | y_{1:T}) = \frac{1}{\breve{Z}} \tilde{\alpha}^{\mathrm{I}}(\mathbf{s}_{t-1}) \gamma(\mathbf{s}_{t-1}, \mathbf{s}_t) \tilde{\beta}^{\mathrm{I}}(\mathbf{s}_t) \tag{10}$$

for $\mathbf{s}_t \in \Omega^{\mathrm{I}}_{t,fwd} \cup \Omega^{\mathrm{I}}_{t,bwd}$ and $\mathbf{s}_{t-1} \in \Omega^{\mathrm{I}}_{t-1,fwd} \cup \Omega^{\mathrm{I}}_{t-1,bwd}$ with the threshold of value η for the messages of inactive states. The coefficients $\widehat{Z}$ and $\breve{Z}$ ensure that the above mass functions sum to one.

IV. Simulation Results

Fig. 2 shows simulation of an 8PSK transmission into a frequency-selective quasi-static Rayleigh-fading channel with turbo equalization performed at the receiver. The channels considered were of five taps with a uniform energy distribution among the channel taps. The frame length contained 192 data bits that were encoded by a rate half $(5, 7)_8$ convolutional code. The optimal equalizer of this system has a total of 4096 states. We have compared the performance of BCJR, M-BCJR, and the proposed (EP-MBCJR) scheme. For the EP-MBCJR, we have performed two iterations of the algorithm in the first global turbo iteration and only one iteration afterward. Hence, the compared M-BCJR and EP-MBCJR systems have similar complexities. After the first global iteration, the algorithm uses the backward messages and active states of the previous global iteration in building up the beliefs of the forward recursion. The proposed algorithm clearly outperforms the conventional M-BCJR algorithm and achieves a frame error rate within 1 dB of the optimal at a fraction of the complexity of the optimal algorithm.

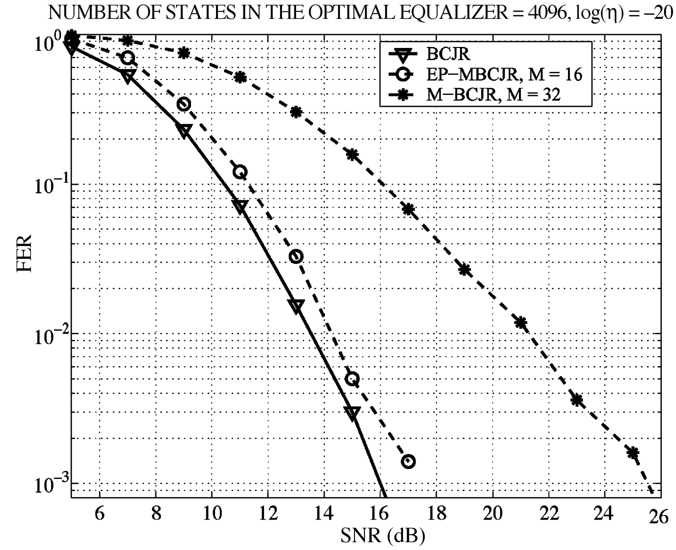

Fig. 2. 8PSK transmission into a five-tap uniform delay profile channel with five turbo equalization iterations.

We have investigated the active states change with iterations and the sensitivity of the proposed algorithm on the threshold η. Fig. 3 shows the simulation of the same transmitter and channel but with the receiver performing serially concatenated equalization and channel decoding. Two iterations of EP-MBCJR were performed at the equalizer in parts (b) and (c). It can be seen that the algorithm stabilizes quickly, giving a final set of active states not totally different to the M-BCJR algorithm, and that the system exhibits stable performance in terms of bit-error rates for a wide range of η.

V. Conclusion

We have applied the concept of EP to the estimation of marginal posterior probabilities of state variables and pairs of state variables of an HMM. The derived algorithm is similar to the conventional M-BCJR algorithm with the active state selection criterion being changed from the filtered probability distribution of state variables to beliefs of state variables. Simulation results of trellis-based equalization showed that the new algorithm outperforms the M-BCJR algorithm in terms of error rate performance for comparable implementation complexity.

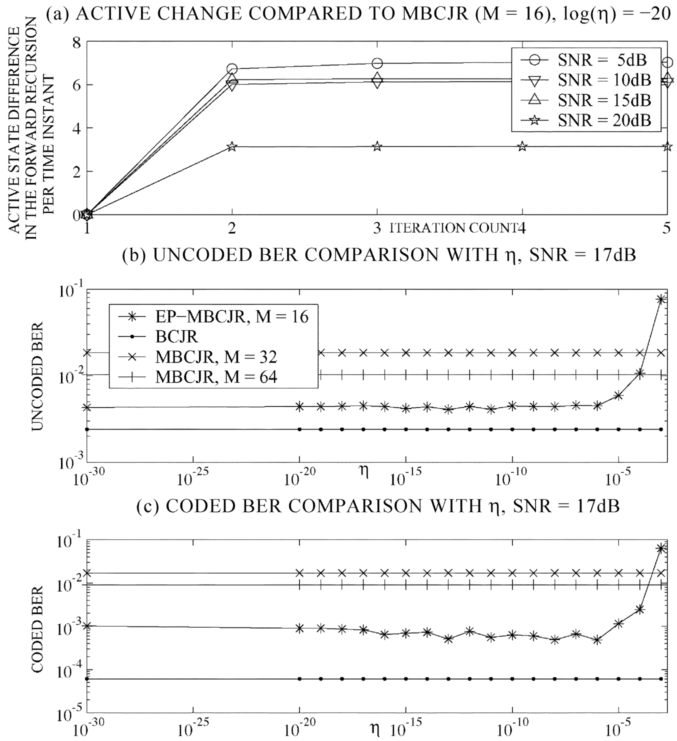

Fig. 3. 8PSK transmission into a five-tap uniform delay profile channel with serial equalization and channel decoding. EP-MBCJR scheme is with $\mathrm{M} = 16$. (a) Average active states difference in the forward recursion. (b) Effect of the threshold η on the uncoded BER. (c) Effect on the coded BER.

It is straightforward to consider a corresponding extension to the T-BCJR algorithm presented in [3] and also to consider the application of these ideas to the sum product algorithm itself operating on discrete state variables of a general graphical model.

Appendix
Proof of Minimization of Kullback–Leibler Divergence $D(q\|\tilde{p})$, by the Selection of the Largest Probabilities of $\tilde{p}(\mathbf{s}_t|y_{1:T})$

Suppose $\mathbf{s}_t$ has a configuration space S. Let $\tilde{p}(\mathbf{s}_t|y_{1:T})$ be a distribution with a support Ξ, and let Ω denote a set of M states of S that have the largest $\tilde{p}(\mathbf{s}_t|y_{1:T})$. Let us denote the cardinalities of the sets S and Ω by $|S|$ and $|\Omega|$, respectively. We have the constraint $|\Omega| \le M \le |S|$.

Now, let $q(\mathbf{s}_t)$ be a probability distribution such that

$$q(\mathbf{s}_t) = \begin{cases} \frac{1}{Z}\tilde{p}(\mathbf{s}_t|y_{1:T}), & \mathbf{s}_t \in \Omega \cap \Xi \\ 0, & \mathbf{s}_t \notin \Omega \cap \Xi. \end{cases}$$

The normalization constant, $Z = \sum_{\mathbf{s}_t \in \Omega} \tilde{p}(\mathbf{s}_t|y_{1:T})$.

Let $\tilde{q}(\mathbf{s}_t)$ be any other probability distribution on $\mathbf{s}_t$ that has a support J with a cardinality $|J| \le |\Omega|$. The Kullback–Leibler divergence between $q(\mathbf{s}_t)$ and $\tilde{p}(\mathbf{s}_t|y_{1:T})$ is

$$\begin{aligned} D(q\|\tilde{p}) &= \sum_{\mathbf{s}_t \in \Omega} q(\mathbf{s}_t) \log \frac{q(\mathbf{s}_t)}{\tilde{p}(\mathbf{s}_t|y_{1:T})} \\ &= -\log\left(\sum_{\mathbf{s}_t \in \Omega} \tilde{p}(\mathbf{s}_t|y_{1:T})\right) \end{aligned}$$

and the divergence between $\tilde{q}(\mathbf{s}_t)$ and $\tilde{p}(\mathbf{s}_t|y_{1:T})$ is

$$\begin{aligned} D(\tilde{q}\|\tilde{p}) &= \sum_{\mathbf{s}_t \in J} \tilde{q}(\mathbf{s}_t) \log \frac{\tilde{q}(\mathbf{s}_t)}{\tilde{p}(\mathbf{s}_t|y_{1:T})} \\ &= -E_{\tilde{q}}\left(\log \frac{\tilde{p}(\mathbf{s}_t|y_{1:T})}{\tilde{q}(\mathbf{s}_t)}\right). \end{aligned}$$

Using Jensen's inequality

$$\begin{aligned} E_{\tilde{q}}\left(\log \frac{\tilde{p}(\mathbf{s}_t|y_{1:T})}{\tilde{q}(\mathbf{s}_t)}\right) &\le \log\left(E_{\tilde{q}}\left(\frac{\tilde{p}(\mathbf{s}_t|y_{1:T})}{\tilde{q}(\mathbf{s}_t)}\right)\right) \\ &= \log\left(\sum_{\mathbf{s}_t \in J} \tilde{p}(\mathbf{s}_t|y_{1:T})\right). \end{aligned}$$

Hence, $D(\tilde{q}\|\tilde{p}) - D(q\|\tilde{p}) \ge$

$$\log\left(\sum_{\mathbf{s}_t \in \Omega} \tilde{p}(\mathbf{s}_t|y_{1:T})\right) - \log\left(\sum_{\mathbf{s}_t \in J} \tilde{p}(\mathbf{s}_t|y_{1:T})\right) \ge 0.$$

The final inequality is since the M largest components of $\tilde{p}(\mathbf{s}_t|y_{1:T})$ reside in Ω. Thus, $q(\mathbf{s}_t)$ is the approximating distribution to $\tilde{p}(\mathbf{s}_t|y_{1:T})$ with a maximum support of M elements, which minimizes the Kullback–Leibler divergence $D(q\|\tilde{p})$.

Acknowledgment

The authors would like to thank Dr. J. P. Coon and the anonymous reviewers for the valuable comments regarding the manuscript.

References

[1] F. R. Kschischang, B. J. Frey, and H. A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 498–519, Feb. 2001.

[2] L. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Trans. Inf. Theory*, vol. IT-20, no. 3, pp. 284–287, Mar. 1974.

[3] V. Franz and J. B. Anderson, "Concatenated decoding with a reduced-search BCJR algorithm," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 2, pp. 186–195, Feb. 1998.

[4] T. P. Minka, "A family of algorithms for approximate Bayesian inference," Ph.D. dissertation, Massachusetts Inst.Technol., Cambridge, MA, Jan. 2001.

[5] Y. Qi and T. Minka, "Expectation propagation for signal detection in flat-fading channels," in *Proc. IEEE Int. Symp. Information Theory*, 2003, p. 358.

[6] T. Heskes and O. Zoeter, "Expectation propagation for approximate inference in dynamic Bayesian networks," in *Proc. UAI*, A. Darwiche and N. Friedman, Eds., 2002, pp. 216–233.

[7] T. Minka. (2004) Power EP. [Online]. Available: http://research.microsoft.com/~minka/papers/minka-power-ep.pdf.

[8] C. Fragouli, N. Al-Dhahir, S. N. Diggavi, and W. Turin, "Prefiltered space-time M-BCJR equalizer for frequency-selective channels," *IEEE Trans. Commun.*, vol. 50, no. 5, pp. 742–753, May 2002.