14th European Signal Processing Conference (EUSIPCO 2006), Florence, Italy, September 4-8, 2006, copyright by EURASIP

PDA-BCJR ALGORITHM FOR FACTORIAL HIDDEN MARKOV MODELS WITH APPLICATION TO MIMO EQUALISATION

Robert J. Piechocki[†], *Christophe Andrieu*[∗], *Magnus Sandell*[♠] *and Joe McGeehan*[†♠]

[†] University of Bristol, Centre For Communications Research
Woodland Road, MVB, BS8 1UB, Bristol, UK.
E-mail: r.j.piechocki@bristol.ac.uk, Tel. +44 117 9545203, Fax. +44 117 9545206
[∗] School of Mathematics, University of Bristol, BS8 1TW, UK
[♠] Toshiba TRL Labs, Bristol, BS1 4ND, UK

ABSTRACT

In this paper we develop an efficient algorithm for inference in Factorial Hidden Markov Models (FHMM), which is particularly suitable for turbo equalisation in Multiple Input - Multiple Output (MIMO) systems. The proposed PDA-BCJR algorithm can be viewed as a generalisation of the PDA algorithm, which in its basic form handles single latent variables only. Our generalisation replaces each of the single latent variables with a HMM.

1. INTRODUCTION

High speed wireless communications systems require some form of equalisation to compensate for the dispersive nature of the wireless channel. The subject of (single transmit antenna) equalisation has been extensively researched and a plethora of technical papers published. Recently, so called MIMO wireless systems have gained an enormous interest. Whereas, the sub-optimal equalisation techniques with good performance are still sought after, the MIMO sub-optimal equalisation remains a real challenge. Nonetheless, two solutions are frequently cited as state of the art: sphere decoders [1][2] and the probabilistic data association (PDA) [3]. The PDA algorithm was first introduced in communications literature as a sub-optimal multi-user detector (MUD) in CDMA systems. Simultaneously to MIMO, turbo principle has reached engineering maturity, even thought some theoretical description remains to be provided. Turbo equalisation can offer significant gains, especially when equalising ill-posed channels. Turbo equalisation requires all subblocks in the system to exchange so-called soft information i.e. full sets of marginal posterior distributions, rather than point estimates (i.e. hard decisions). Both PDA and sphere decoders can operate in the soft mode. As aforementioned, the challenge remains how to efficiently decode wideband MIMO signals at a reasonable computational cost. The MIMO equalisation problem can be formulated as inference on a Hidden Markov Model, for which an exact and efficient algorithm is well known: forward-backward (aka BCJR). Unfortunately, the complexity of BCJR increases exponentially with the number of transmit antennas and the length of the channel measured in symbol intervals. Reduced complexity solutions based on heuristics [4] and Monte Carlo approximations [5] have been proposed in the literature to address this problem.

An interesting line of research was presented in [6], where a more appropriate Factorial Hidden Markov Model (FHMM) was used for MIMO equalisation. The authors used a Structured Variational Inference (SVI) approximation of [7] to compute marginal probabilities required in turbo equalisation. In this contribution we propose a different type of approximation for inference in FHMM based on the probabilistic data association (PDA) principle. Simulation results suggest improved results as compared to the SVI.

2. PROBLEM FORMULATION AND ESTIMATION TASKS

We consider a MIMO system with $N; (1 \leq n \leq N)$ transmit and $M; (1 \leq m \leq M)$ receive antennas. The system signals over a wideband channel, modelled as an equivalent multidimensional $(N \times M)$ FIR filter with L taps:

$$\mathbf{H} = \begin{bmatrix} \mathbf{h}_{1,1}^T & \cdots & \mathbf{h}_{1,N}^T \\ \vdots & \ddots & \vdots \\ \mathbf{h}_{M,1}^T & \cdots & \mathbf{h}_{M,N}^T \end{bmatrix} \tag{1}$$

Where:

$$\mathbf{h}_{m,n} = \left(h_{m,n}^{(\tau=0)}, h_{m,n}^{(\tau=1)}, \ldots, h_{m,n}^{(\tau=L-1)}\right)^T \tag{2}$$

We introduce an additional notation: $\mathbf{H}_{:,n}$ - is the n^{th} block column of $\mathbf{H}$ i.e:

$$\mathbf{H}_{:,n} \equiv \begin{bmatrix} \mathbf{h}_{1,n}^T \\ \vdots \\ \mathbf{h}_{M,n}^T \end{bmatrix}$$

At each time instant t, the n^{th} antenna signals a symbol $x_n^{(t)}$, that belongs to a digital modulation with Ω elements: $x_n^{(t)} \in \{a_1, \ldots, a_\Omega\}$. A collection of L consecutive symbols forms a vector: $\mathbf{x}_n^t = \left(x_n^{(t)}, x_n^{(t-1)}, \ldots, x_n^{(t-L+1)}\right)^T$. Stacking N such vectors produces another vector: $\mathbf{x}^{(t)} = \left(\left(\mathbf{x}_1^{(t)}\right)^T, \left(\mathbf{x}_2^{(t)}\right)^T, \ldots, \left(\mathbf{x}_N^{(t)}\right)^T\right)^T$. With this notation, the received signal at time t is:

$$\mathbf{y}^{(t)} = \mathbf{H}\mathbf{x}^{(t)} + \mathbf{n}^{(t)} \tag{3}$$

Where: $\mathbf{n}^{(t)}$ is a Gaussian noise i.e. $\mathbf{n}^{(t)} \sim \mathcal{N}\left(\mathbf{0}, \sigma_n^2 \mathbf{I}\right)$

We make an assumption that our system operates in a turbo configuration. This is conventional and not detailed here. It suffices to point out that in a turbo system the MIMO decoder has to supply a full set of marginal posterior probability mass functions (or an approximation to it):

$$\left\{ f\left(x_n^{(1)}\,|\mathbf{y}\right), f\left(x_n^{(2)}\,|\mathbf{y}\right), \ldots, f\left(x_n^{(T)}\,|\mathbf{y}\right)\right\} \tag{4}$$

The above problem can be elegantly described in terms of Probabilistic graphical models.

2.1 Probabilistic Graphical Models

Probabilistic graphical models (PGM) have recently emerged as a universal language describing problems in decision making and/or estimation in the presence of uncertainty. PGMs are used (and were often independently developed) in seemingly unrelated disciplines ranging from: Bio-informatics (Bio-statistics), Information retrieval, Speech processing and image processing Communications to Forensic Science and many more. It has to be emphasised that PGMs themselves are nothing more than pictorial representations of families of probability density functions, and as such, they do not offer any solutions. However, PGMs may provide an insight into existing models, motivation for new models and algorithms, and ease of transferring solutions from one discipline to another. There are three types of PGMs [8] that are used most often: Directed Acyclic Graphs (DAG), Undirected Graphs (UG) and Factor Graphs (FG). The diversity arises from a fact that some problems can be accurately described only by one type of graph. Other problems are well describes by more than one type of graph. Such example is the problem considered in this contribution. Reference [6] outlined the problem using FG, however here we opt for DAG. DAGs are useful in modelling causal relationships. In communications the causal relationship is appealingly natural: the received waveform arises as a response to the transmitted waveform and realisation of a medium (wireless channel in our case).

A constituent model that will build our overall FHMM is a factorial model depicted in figure 1. In DAGs the joint distribution factors in such a way that each factor (of total D) represents the conditional distribution of a "child" given its "parents": $f(x_1 \ldots, x_N) = \prod_{i=1}^{D} f(x_i\,|pa_i)$

In our case this becomes:

$$f(\mathbf{y}, x_1, \ldots, x_N, \mathbf{H}) = f(\mathbf{y}\,|x_1, \ldots, x_N\,, \mathbf{H})\, f(\mathbf{H}) \prod_{i=1}^{N} f(x_i) \tag{5}$$

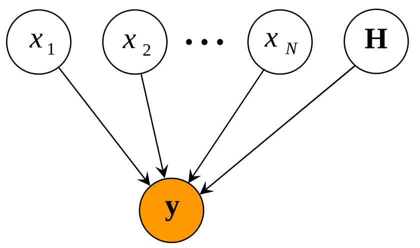

Figure 1: Directed Acyclic Graph (DAG) of a factorial model (MIMO and MUD/CDMA problems).

The task is to calculate a set of marginal posterior probabilities for all x_i, $i \in \{1, 2, \ldots, N\}$.

$$\begin{aligned} f(x_i\,|\mathbf{y}) &= \int_{\mathbf{H}} \sum_{x_{-i}} f(x_i, x_{-i}, \mathbf{H}\,|\mathbf{y})\, d\mathbf{H} = \\ &\int_{\mathbf{H}} \sum_{x_{-i}} f(x_i\,|x_{-i}, \mathbf{H}, \mathbf{y})\, f(x_{-i}\,|\mathbf{H}, \mathbf{y})\, f(\mathbf{H}\,|\mathbf{y})\, d\mathbf{H} \end{aligned} \tag{6}$$

Where "$-i$" stands for "all except the i^{th}". This also allows to treat the channel as a random variable. As we do not have access to a particular realisation of the channel $\mathbf{H}$ ($\mathbf{H}$ is latent), this variable is integrated out. Unfortunately, such integral is impossible to calculate in a typical scenario. The first solution (ubiquitous in practice) is to train the channel using known at the receiver data. This amounts to approximating the posterior distribution of the channels with one "spike" centered perhaps at an ML estimate: $f(\mathbf{H}\,|\mathbf{y}) = \delta\left(\mathbf{H} - \hat{\mathbf{H}}_{ML}\right)$

$$\begin{aligned} &\int_{\mathbf{H}} \sum_{x_{-i}} f(x_i\,|x_{-i}, \mathbf{H}, \mathbf{y})\, f(x_{-i}\,|\mathbf{H}, \mathbf{y})\, \delta\left(\mathbf{H} - \hat{\mathbf{H}}_{ML}\right) d\mathbf{H} = \\ &\sum_{x_{-i}} f\left(x_i \,\middle|\, x_{-i}, \hat{\mathbf{H}}_{ML}, \mathbf{y}\right) f\left(x_{-i} \,\middle|\, \hat{\mathbf{H}}_{ML}, \mathbf{y}\right) \end{aligned} \tag{7}$$

A better way to account for the channel uncertainty is via an EM algorithm as discussed in section 3.2.

Even though training simplifies the problem, another problem remains to be tackled. This (computational) problem is brought about by a fact that the joint *a posteriori* distribution is not separable: $f(x_1, x_2, \ldots, x_N\,|\mathbf{H}, \mathbf{y}) \neq \prod_{n=1}^{N} f(x_N\,|\mathbf{H}, \mathbf{y})$ since all the variables $\{x_1, x_2, \ldots, x_N\}$ co-operate in setting the mean of the observation $\mathbf{y}$ i.e. $E\{\mathbf{y}\} = h(x_1, x_2, \ldots, x_N)$, which is a root cause of numerical complexity for this type of problems.

A second component needed for our model is a HMM model of figure 2. This model is equivalently described by eq. (8).

$$\begin{aligned} &f\left(\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(T)}, s^{(1)}, \ldots, s^{(T)}, \mathbf{H}\right) = \\ &f(\mathbf{H}) \prod_{t=1}^{T} f\left(\mathbf{y}^{(t)} \,\middle|\, s^{(t)}, \mathbf{H}\right) f\left(s^{(t)} \,\middle|\, s^{(t-1)}\right) \end{aligned} \tag{8}$$

where the states $s^{(t)}$ are defined as $s^{(t)} \equiv \left\{x^{(t)}, x^{(t-1)}, \ldots, x^{(t-L+1)}\right\}$.

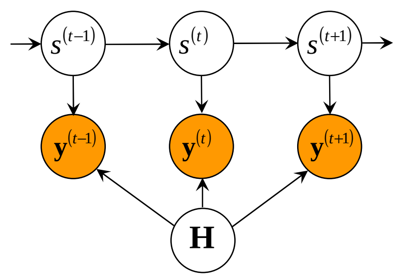

Figure 2: DAG of a Hidden Markov Model.

Both single antenna and MIMO equalisation can be represented by a HMM. In a case of known channels (pictorially: if the channel variable $\mathbf{H}$ was shaded) the task of marginalisation is achieved by a forward-backward algorithm (more specifically a variant known as BCJR).

Our overall FHMM model arises by replacing the single random variables $\{x_n\}$ in figure 1 with HMM in figure 2 - as depicted in figure 3.

$$f\left(\mathbf{y}^{(1)},\ldots,\mathbf{y}^{(T)},s_1^{(1)},\ldots,s_1^{(T)},\ldots,s_N^{(1)},\ldots,s_N^{(T)},\mathbf{H}\right) = f(\mathbf{H})\prod_{t=1}^{T}\left[f\left(\mathbf{y}^{(t)}\left|s_1^{(t)},\ldots s_N^{(t)},\mathbf{H}\right.\right)\prod_{n=1}^{N}f\left(s_n^{(t)}\left|s_n^{(t-1)}\right.\right)\right] \quad (9)$$

In a MIMO system the symbols sent from all N transmit antennas are a priori independent (which is a reasonable assumption when interleaves are used). However, the a posteriori distribution is again not separable for the very same reason, which again leads to exuberant complexities.

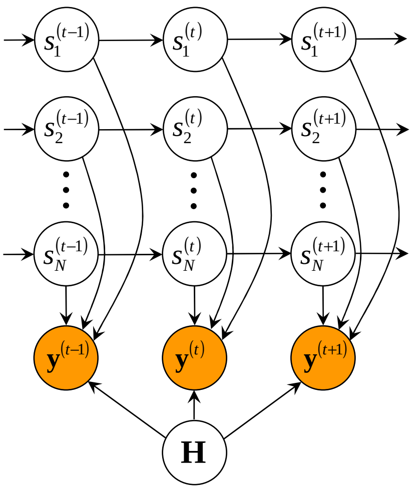

Figure 3: DAG of a Factorial Hidden Markov Model.

3. THE PDA-BCJR ALGORITHM

In this section we detail the PDA-BCJR algorithm. We also describe modifications to the BCJR algorithm that are required for the use in PDA-BCJR algorithm.

In each iteration the algorithm updates a set of marginal posterior distributions

$$\left\{f\left(x_n^{(1)}|\mathbf{y}\right),f\left(x_n^{(2)}|\mathbf{y}\right),\ldots,f\left(x_n^{(T)}|\mathbf{y}\right)\right\}$$

via a modified BCJR algorithm. This set pertains to all $T;(1\le t\le T)$ symbols sent from the n^{th} transmit antenna. In doing so, the modified BCJR algorithm exchanges two sets of moments: input $(\mu_{n,in},\mathbf{\Sigma}_{n,in})$ and output $(\mu_{n,out},\mathbf{\Sigma}_{n,out})$. The input moments in each iteration are calculated by summing all but the n^{th} output moments, as detailed in table 1.

In standard BCJR algorithm as used in MIMO (soft in - soft out) equalisation, the posterior transition probability is factored:

$$f(s',s|\mathbf{y}) \propto \underbrace{f\left(s',\mathbf{y}^{(1:t-1)}\right)}_{\alpha^{(t)}}\underbrace{f\left(s|s'\right)f\left(\mathbf{y}^{(t)}|s',s\right)}_{\gamma^{(t)}}\underbrace{f\left(\mathbf{y}^{(t+1:T)}|s\right)}_{\beta^{(t)}} \quad (10)$$

and calculated recursively for numerical efficiency as:

$$\alpha^{(t)}(s) = \sum_{s'}\alpha^{(t-1)}(s')\gamma^{(t)}(s',s) \quad (11)$$

$$\beta^{(t-1)}(s') = \sum_{s'}\beta^{(t)}(s)\gamma^{(t)}(s',s) \quad (12)$$

the middle factor γ is calculated as:

$$\gamma^{(t)}(s',s) \propto \exp\left(\frac{-1}{2\sigma^2}\left|\mathbf{y}_t-\hat{\mathbf{H}}\mathbf{x}_{i,j}\right|^2\right)\prod_{n=1}^{N}f\left(x_n^{(t)}\right) \quad (13)$$

In the modified BCJR (for the use in PDA-BCJR), the middle factor (to take into account non zero mean and non diagonal covariance) is calculated as:

$$\gamma^{*(t)}(s',s) \propto \exp\left(-\left(\mathbf{z}_n^{(t)}\right)^H\left(\mathbf{\Sigma}_{n,in}^{(t)}\right)^{-1}\left(\mathbf{z}_n^{(t)}\right)\right)f(x_n) \quad (14)$$

Where:

$$\mathbf{z}_n^{(t)} = \mathbf{y}^{(t)} - \hat{\mathbf{H}}_{(:,n)}\mathbf{x}_n - \mu_{n,in}^{(t)}$$

The α's and the β's are calculated as in the standard BCJR algorithm eq (11) and (12). Additionally, the modified BCJR algorithm calculates "output moments" (i.e. the mean and the covariance of the received signal in an absence of the noise and other antennae signals) at each time instant:

$$\mu_{n,out}^{(t)} = E\left\{\hat{\mathbf{H}}_{:,n}\mathbf{x}_n\right\} = \sum_{j=1}^{S}\sum_{k=1}^{K}\hat{\mathbf{H}}_{:,n}\mathbf{x}_n f(s',s|\mathbf{y}) \quad (15)$$

$$\begin{aligned}\mathbf{\Sigma}_{n,out}^{(t)} &= E\left\{\left(\hat{\mathbf{H}}_{:,n}\mathbf{x}_n-\mu_{n,out}^{(t)}\right)\left(\hat{\mathbf{H}}_{:,n}\mathbf{x}_n-\mu_{n,out}^{(t)}\right)^H\right\}\\ &= \sum_{j=1}^{S}\sum_{k=1}^{K}\left(\hat{\mathbf{H}}_{:,n}\mathbf{x}_n-\mu_{n,out}^{(t)}\right)\left(\hat{\mathbf{H}}_{:,n}\mathbf{x}_n-\mu_{n,out}^{(t)}\right)^H f(s',s|\mathbf{y})\end{aligned} \quad (16)$$

To initialise the algorithm two options are possible: 1) $f(s',s|\mathbf{y}) = \frac{1}{SK}$, which indeed is the only option in practice on the first turbo iteration; 2) $f(s',s|\mathbf{y}) = I(s',s)\frac{f\left(x_n^{(t)}\right)}{S}$, which becomes an option after the first turbo iteration when the "prior" distributions $f\left(x_n^{(t)}\right)$ become available. $I(s',s)$ is an indicator function that is 1 if x_n is compatible with state transition and 0 otherwise. For efficient implementation of the modified BCJR algorithm, one realises that both the branch metrics $\hat{\mathbf{H}}_{:,n}\mathbf{x}_n$ and the posterior probabilities $f(s',s|\mathbf{y}) \propto \alpha^{(t)}\beta^{(t)}\gamma^{(t)}$ are pre-calculated as a part of standard procedures of BCJR algorithm.

3.1 Connections with Variational approximation

It is interesting to relate the proposed PDA-BCJR algorithm to the structured variational approximation (SVI) of [7] [6]. The SVI approximation associates a fictitious observation with each Markov Chain, effectively decupling the chains. Our PDA-BCJR algorithm does the same, however it calculates a second moment in addition to the first calculated by VI. To see this and for simplicity of presentation we come back to the factorial model of figure 1 (relations in FHMM are analogous).

In a factorial model a set of variational updates is given by:

$$f_{VI}(x_i = a_\omega|\mathbf{y}) \overset{appox}{\propto} \exp\left(\frac{-1}{\sigma^2}\psi^T\mathbf{h}_i a_\omega - \frac{1}{2\sigma^2}|\mathbf{h}_i a_\omega|^2\right) \quad (17)$$

Table 1: The PDA-BCJR algorithm summary

"'Outer recursions"'

for $I = 1 : I_{PDAmax}$

"'PDA recursions"'

for $n = 1 : N$

$$\mu_{n,in} = \sum_{j=1, j\neq n}^{N} \mu_{j,out}; \quad \mathbf{\Sigma}_{n,in} = \sum_{j=1, j\neq n}^{N} \mathbf{\Sigma}_{j,out}$$

$$\left[f\left(x_n^{(t),t=1:T} | \mathbf{y}\right), \mu_{n,out}, \mathbf{\Sigma}_{n,out} \right] = BCJR^{*}\left(\mathbf{y}, \hat{\mathbf{H}}_{:,n}, f\left(x_n^{(t),t=1:T}\right), \mu_{n,in}, \mathbf{\Sigma}_{n,in}\right)$$

end, end

where:

$$\psi = \mathbf{y} - \sum_{j=1, j\neq i}^{N} \mathbf{h}_j \left\langle x_j \right\rangle_{x_j|\mathbf{y}}$$

or equivalently:

$$f_{VI}\left(x_i = a_\omega | \mathbf{y}\right) \overset{appox}{\propto} \exp\left(\frac{-1}{2\sigma^2} |\psi - \mathbf{h}_i a_\omega|^2\right) \tag{18}$$

The PDA algorithm updates this distribution via:

$$f_{PDA}\left(x_i = a_\omega | \mathbf{y}\right) \overset{appox}{\propto} \exp\left(\left(\psi - \mathbf{h}_i a_\omega\right)^T \mathbf{M}^{-1} \left(\psi - \mathbf{h}_i a_\omega\right)\right) \tag{19}$$

where:

$$\mathbf{M} = \frac{\sigma^2}{2}\mathbf{I} + \sum_{j=1, j\neq i}^{N} \mathbf{h}_j^T \mathbf{h}_j Var\left\{x_j\right\} \tag{20}$$

The relation is:

$$\mathbf{M}_{PDA} = \mathbf{M}_{VI} + \sum_{j=1, j\neq i}^{N} \mathbf{h}_j^T \mathbf{h}_j Var\left\{x_j\right\} \tag{21}$$

Hence, variational approximation neglects to update the covariance when compared to PDA. Here we mean a "naive variational approximation" a.k.a mean field approximation.

3.2 EM algorithms for Semi-blind equalisation

In the previous section we have presented the PDA-BCJR algorithm with an assumption that the channels were known (or estimated). As aforementioned a better approach to account for the channel uncertainly is an iterative procedure EM - Expectation Maximisation. The EM algorithm in our case can be traced directly to the Baum-Welch procedure known in speech processing problems [9]. We choose set up the problem so that the hidden (latent) data are the states (sets of transmitted symbols). To provide the algorithm with a reasonable initial estimate, a few known symbols are transmitted at the beginning of each frame. In standard iterative procedure E step is performed first. In our case this is done via PDA-BCJR algorithm. M step is identical to that of [7] and [6].

4. SIMULATIONS RESULTS

A MIMO system has been simulated with 3Tx and 3Rx antennas. The data bust of 768 bits has been encoded by a single RCC (7,5) encoder and interleaved prior to transmission. The BPSK signalling has been used and the channel taps are modelled as $h_{m,n}^{\tau} \sim \mathcal{N}\left(0, L^{-\frac{1}{2}}\right)$ with $L = 3$ (uniform power delay profile).

Three turbo MIMO detectors have been investigated. The decoders differ in soft in - soft out equaliser block. The first is a full complexity BCJR. This is an optimal decoder for this case, and it operates over HMM (neglecting FHMM structure). The two other decoders are reduced complexity versions. The first reduced complexity decoder is based on structured variational approximation [7] and [6]. The second is our PDA-BCJR decoder.

Figure 4 depicts the BER performance of the developed reduced complexity PDA-BCJR over FHMM, and the performance of the optimal BCJR over HMM. After 5 iterations the performance is essentially indifferent in both cases. SVI (not depicted) also performs very well in this case.

Figure 5 shows performances of semi-blind versions (short training with EM) of the three decoders. M step was performed on the outputs of the channel decoder i.e. EM was incorporated into the "turbo loop". First 18 symbols were known to the receiver to initialise EM. In this case PDA-BCJR clearly outperforms structured variational inference decoder.

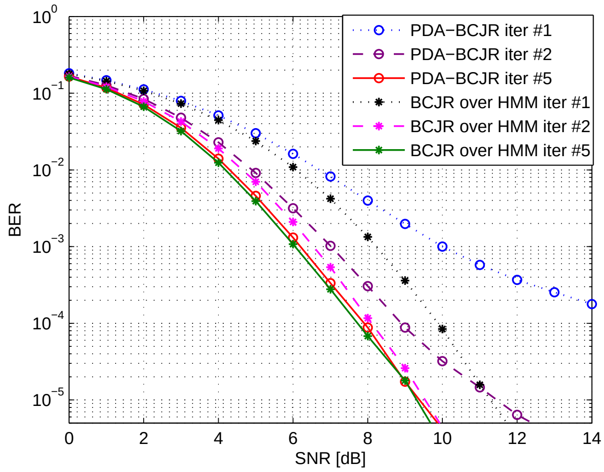

Figure 4: Performance comparison of PDA-BCJR over FHMM and BCJR over HMM.

5. CONCLUSIONS

We have developed a generalisation to a popular PDA algorithm. The algorithm iterates on entire hidden Markov models, and as such is suitable for MIMO equalisation problems.

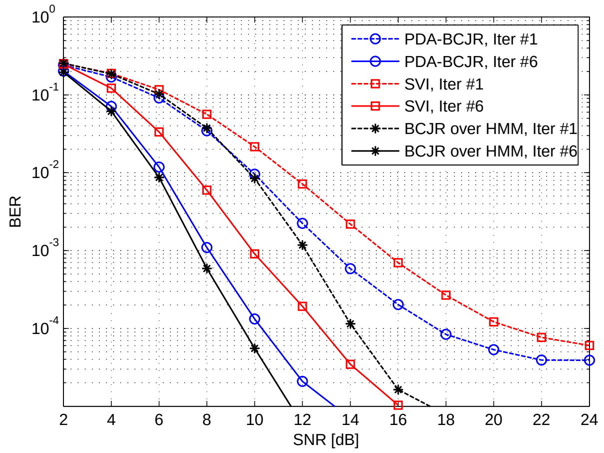

Figure 5: Performance comparison in a semi-blind setting (all via EM): PDA-BCJR over FHMM, SVI over FHMM and BCJR over HMM.

For degenerate Markov chains $f\left(s^{(t)}\middle|s^{(t-1)}\right) = f\left(s^{(t)}\right)$ (i.e. no time dispersion), the developed PDA-BCJR becomes a standard PDA [3]. The PDA-BCJR and the FHMM can serve as a framework for the development of further approximate algorithms by introducing the next level of approximation within the BCJR step.

Acknowledgments

The authors would like to thank C. Vithanage for helpful comments and Toshiba TRL Bristol UK for supporting this work.

REFERENCES

[1] B.M. Hochwald and S. ten Brink, ”*Achieving near-capacity on a multiple-antenna channel*”, IEEE Transactions on Communications, vol. 51, No. 5, 2003, pp. 389 - 399.

[2] M.O. Damen and K. Abed-Meraim and J.C. Belfiore, ”*Generalised sphere decoder for asymmetrical space-time communication architecture*”, IEE Electronics Letters, Vol. 36, 2000, pp. 166-167.

[3] J. Luo and K. Pattipati and P. Willett and F. Hasegawa, ”*Near-Optimal Multiuser Detection in Synchronous CDMA using Probabilistic Data Association*”, IEEE Communications Letters, Vol. 5, No. 9, 2001, pp. 361-363.

[4] P. Thiennviboon and G. Ferrari and K.M Chugg, ”*Generalized trellis-based reduced-state soft-input/soft-output algorithms*”, IEEE International Conference on Communications, ICC 2002, Vol. 3, 2002, pp.1667 - 1671.

[5] Christophe Andrieu and Robert J. Piechocki and Joe McGeehan, ”*Particle Smoothing Techniques with Turbo Principle for MIMO Systems*”, IEEE Signal Processing Advances in Wireless Communications (SPAWC-2003), 2003.

[6] M. Nissila and S. Pasupathy,”*Reduced-complexity Turbo Receivers for Single and Multi-Antenna Systems via Variational Inference in Factor Graphs*”, In Proc. IEEE ICC conference, Paris, 2000.

[7] Z. Ghahramani and M. Jordan”, ”*Factorial Hidden Markov Models*”, Machine Learning, Kluwer Academic Publishers, Vol. 29, No. 3, 1997, pp. 245-275.

[8] Michael I. Jordan, ”*Graphical models*”, Statistical Science - Special Issue on Bayesian Statistics, Vol. 19, 2004, pp. 140-155.

[9] L. Rabiner, ”*A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition*”, Proceedings of the IEEE, Vol. 77, No.2, 1989, pp. 257-286.