ON DATA AND PARAMETER ESTIMATION USING THE VARIATIONAL BAYESIAN EM-ALGORITHM FOR BLOCK-FADING FREQUENCY-SELECTIVE MIMO CHANNELS

Lars P. B. Christensen, Jan Larsen

Informatics and Mathematical Modelling, Technical University of Denmark
Email: {lc, jl}@imm.dtu.dk

ABSTRACT

A general Variational Bayesian framework for iterative data and parameter estimation for coherent detection is introduced as a generalization of the EM-algorithm. Explicit solutions are given for MIMO channel estimation with Gaussian prior and noise covariance estimation with inverse-Wishart prior. Simulation of a GSM-like system provides empirical proof that the VBEM-algorithm is able to provide better performance than the EM-algorithm. However, if the posterior distribution is highly peaked, the VBEM-algorithm approaches the EM-algorithm and the gain disappears. The potential gain is therefore greatest in systems with a small amount of observations compared to the number of parameters to be estimated.

1. INTRODUCTION

The focus of this paper is on improved iterative data and parameter estimation for coherent detection in block-fading frequency-selective MIMO channels. Much work has been done within this field and many variants of the EM-algorithm have been applied to communication systems, see for example [1, 2, 3]. However, previous estimators have all provided point-estimates of the parameters, not distributions as offered by the full Bayesian approach. On the other hand, Bayesian estimators average over the distribution of the unknown variables or parameters to provide improved inference about the system. Previously, a so-called Bayesian EM (BEM)-algorithm was introduced for communication systems [2, 3]. However, the BEM-algorithm provides a Maximum A Posteriori (MAP) point-estimate and is therefore not a true Bayesian estimator.

The contribution of this paper is to introduce the Variational Bayesian EM (VBEM)-algorithm, already used extensively in the machine-learning community, to the communications society. Explicitly, the contribution is to formulate an iterative data, channel and noise covariance estimator based on the VBEM-algorithm. By simulations it is shown, that the performance of a communication system can be improved over that based on the EM-algorithm when there is significant uncertainty in the parameter estimates.

2. SYSTEM MODEL

We will consider the uncoded linear $M \times N$ MIMO system

$$\mathbf{y}_i = \mathbf{H}\mathbf{x}_i + \mathbf{n}_i \tag{1}$$

where $\mathbf{H} \in \mathbb{C}^{M \times N}$ is the channel matrix and $\mathbf{x}_i \in \Omega^{N \times 1}$ is the vector of transmitted symbols at time index i, each belonging to the complex-valued alphabet Ω. The received signal vector $\mathbf{y}_i \in \mathbb{C}^{M \times 1}$ holds the observations at time i and the additive noise $\mathbf{n}_i \in \mathbb{C}^{M \times 1}$ is assumed to be circular zero-mean Gaussian with covariance $\mathbf{\Sigma} \triangleq E\left[\mathbf{n}_i\mathbf{n}_i^H\right]$ and $E\left[\mathbf{n}_i\mathbf{n}_i^T\right] = \mathbf{0}$. The generalization of the estimation framework to Gauss-Markov noise is straightforward [4].

The frequency-selective channel is assumed to have a temporal length of L symbols. Let N_t and N_r denote the number of transmitters and receivers respectively leading to $N = LN_t$ and $M = N_r$. For channel estimation, it is desirable to rewrite the channel matrix into a vector notation as

$$\mathbf{y}_i = \mathbf{X}_i\mathbf{h} + \mathbf{n}_i \tag{2}$$

with $\mathbf{h} \triangleq vec\left(\mathbf{H}\right)$ where $vec\left(\cdot\right)$ is the column stacking operator. The k'th row of the symbol matrix $\mathbf{X}_i \in \mathbb{C}^{N_r \times LN_tN_r}$ is found by upsampling $\mathbf{x}_i^T$ by N_r and shifting it right by $k-1$ positions producing a Toeplitz structure. The two representations are equivalent and we can use the best suited depending on conditions.

Assuming data is sent in frames of N_f symbols per transmitter, the collection of all transmitted symbols and observations is given by

$$\mathbf{x} = \{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_{N_f}\}, \quad \mathbf{y} = \{\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_{N_e}\} \tag{3}$$

where $N_e = N_f + L - 1$ due to the convolutive multipath channel.

3. MAXIMUM LIKELIHOOD ESTIMATION

In this section, a quick outline of Maximum Likelihood (ML) estimation using the EM-algorithm is presented as the VBEM-algorithm is a generalization of the EM-algorithm. The framework is in a general form and is carried over to the formulation of the VBEM-algorithm. For further details, see [5, 6].

The first author would like to thank Nokia Denmark for funding

1-4244-0469-X/06/$20.00 ©2006 IEEE

The idea behind the EM-algorithm is to consider the observations $\mathbf{y}$ being incomplete data as the underlying hidden variables $\mathbf{x}$ are unknown. This problem is overcome by considering the hidden variables as being random variables and averaging over their distribution. By this philosophy we can write the complete-data log-likelihood of the parameter set $\boldsymbol{\theta}$

$$E: \quad Q\left(\boldsymbol{\theta}, \boldsymbol{\theta}^{(j-1)}\right) \triangleq \langle ln\left[p\left(\mathbf{y}, \mathbf{x} \mid \boldsymbol{\theta}\right)\right]\rangle_{p\left(\mathbf{x}|\mathbf{y}, \boldsymbol{\theta}^{(j-1)}\right)} \tag{4}$$

where $\boldsymbol{\theta}^{(j-1)}$ is the parameter set from the previous iteration and $\langle\cdot\rangle_{p(\cdot)}$ indicates averaging w.r.t. the distribution in the subscript. Carrying out the above averaging is often termed the E-step. Next, in the so-called M-step we maximize w.r.t. $\boldsymbol{\theta}$, i.e.

$$M: \quad \boldsymbol{\theta}^{(j)} \triangleq \arg\max_{\boldsymbol{\theta}} \quad Q\left(\boldsymbol{\theta}, \boldsymbol{\theta}^{(j-1)}\right) \tag{5}$$

We now have an iterative algorithm, which can be shown to converge to a local maximum in $p\left(\mathbf{y}, \boldsymbol{\theta}\right)$. However, the fact that the algorithm "only" converges to a local maximum makes initialization of the algorithm crucial, as it may otherwise converge to an incorrect maximum.

In terms of the system model from Section 2, the observations are the received samples $\mathbf{y}$, the hidden variables correspond to the transmitted symbols $\mathbf{x}$ and the parameter set is $\boldsymbol{\theta} = \{\mathbf{h}, \boldsymbol{\Sigma}\}$. In the E-step, the posterior distribution of the transmitted symbols $p\left(\mathbf{x} \mid \mathbf{y}, \mathbf{h}^{(j)}, \boldsymbol{\Sigma}^{(j)}\right)$ is found by the well-known BCJR algorithm using forward-backward recursions, see e.g. [4]. The M-step finds the joint ML channel and covariance estimate, but this produces non-linear systems of equations that in the general case appear to have no closed-form solution. A solution is to find the individual ML estimates and possibly iterate between them in the M-step. The individual solutions are easily found to be the Weighted Least-Squares estimator and the sample covariance for the channel and covariance estimate respectively, both averaged over the posterior of the symbols. This common result is not reproduced here, but is given by $\mathbf{h}_{MAP}^{(j)}$ in (13) and $\mathbf{S}^{(j)}$ in (14) by replacing the parameter distribution with a delta-function in the ML point estimate, i.e. $q_{\boldsymbol{\theta}}\left(\boldsymbol{\theta}\right) = \delta\left(\boldsymbol{\theta} - \boldsymbol{\theta}_{ML}\right)$ and $\boldsymbol{\Sigma}_1^{-1} = \mathbf{0}$.

4. BAYESIAN ESTIMATION

In a truly Bayesian framework, all unknown variables and parameters are treated as random variables with some distribution that can be integrated out. The marginal likelihood of the model is therefore found by integrating out the uncertainty as

$$p\left(\mathbf{y}\right) = \int p\left(\mathbf{y}, \mathbf{x}, \boldsymbol{\theta}\right) d\mathbf{x} d\boldsymbol{\theta} \tag{6}$$

However, for interesting models the integration is likely to be intractable as it involves multi-dimensional integrals over complicated expressions. Instead, we lower-bound the marginal log-likelihood by Jensen's inequality as

$$\begin{aligned} ln\left[p\left(\mathbf{y}\right)\right] &= ln\left[\int q\left(\mathbf{x}, \boldsymbol{\theta}\right) \frac{p\left(\mathbf{y}, \mathbf{x}, \boldsymbol{\theta}\right)}{q\left(\mathbf{x}, \boldsymbol{\theta}\right)} d\mathbf{x} d\boldsymbol{\theta}\right] \\ &\geq \int q\left(\mathbf{x}, \boldsymbol{\theta}\right) ln\left[\frac{p\left(\mathbf{y}, \mathbf{x}, \boldsymbol{\theta}\right)}{q\left(\mathbf{x}, \boldsymbol{\theta}\right)}\right] d\mathbf{x} d\boldsymbol{\theta} \end{aligned} \tag{7}$$

where $q\left(\mathbf{x}, \boldsymbol{\theta}\right)$ is a free distribution used to approximate the posterior $p\left(\mathbf{y}, \mathbf{x}, \boldsymbol{\theta}\right)$. Maximizing the lower-bound w.r.t. the free distribution $q\left(\mathbf{x}, \boldsymbol{\theta}\right)$ yields the exact posterior, which was what we started out with, and is therefore of no interest. Constraining the free distribution to factorize between the hidden variables and the parameters by requiring

$$q\left(\mathbf{x}, \boldsymbol{\theta}\right) = q_{\mathbf{x}}\left(\mathbf{x}\right) q_{\boldsymbol{\theta}}\left(\boldsymbol{\theta}\right) \tag{8}$$

provides the intriguing solution that we can optimize the free distributions individually and iterate between them to maximize the lower-bound. This is done by the alternating between the VBE-step and the VBM-step given by

$$\begin{aligned} VBE: \quad & q_{\mathbf{x}}^{(j)}\left(\mathbf{x}\right) \propto e^{\langle ln\left[p\left(\mathbf{y}, \mathbf{x}|\boldsymbol{\theta}\right)\right]\rangle_{q_{\boldsymbol{\theta}}^{(j-1)}\left(\boldsymbol{\theta}\right)}} \\ VBM: \quad & q_{\boldsymbol{\theta}}^{(j)}\left(\boldsymbol{\theta}\right) \propto p\left(\boldsymbol{\theta}\right) e^{\langle ln\left[p\left(\mathbf{y}, \mathbf{x}|\boldsymbol{\theta}\right)\right]\rangle_{q_{\mathbf{x}}^{(j)}\left(\mathbf{x}\right)}} \end{aligned} \tag{9}$$

where $p\left(\boldsymbol{\theta}\right)$ is a parameter prior. Due to the factorization, global convergence can not be guaranteed, but it can be shown to converge to a local maximum in $p\left(\mathbf{y}\right)$. From (9) we see that the VBEM-algorithm is similar to the EM-algorithm, but the distinction between hidden variables and unknown parameters has vanished as the VBE- and VBM-steps are both averaging over posterior distributions. For more details on Bayesian estimation and the VBEM-algorithm, see [6, 7].

Returning to the system model of Section 2 we now have

$$q\left(\mathbf{x}, \mathbf{h}, \boldsymbol{\Sigma}\right) = q_{\mathbf{x}}\left(\mathbf{x}\right) q_{\mathbf{h}}\left(\mathbf{h}\right) q_{\boldsymbol{\Sigma}}\left(\boldsymbol{\Sigma}\right) \tag{10}$$

where the free distribution is further assumed to factorize between the channel and noise covariance posterior. This approach is equivalent to the individual maximization described in Section 3 for the M-step. The above facorization can be seen to yield the updates

$$\begin{aligned} & q_{\mathbf{x}}^{(j)}\left(\mathbf{x}\right) \propto e^{\langle ln\left[p\left(\mathbf{y}, \mathbf{x}|\mathbf{h}, \boldsymbol{\Sigma}\right)\right]\rangle_{q_{\mathbf{h}}^{(j-1)}\left(\mathbf{h}\right) q_{\boldsymbol{\Sigma}}^{(j-1)}\left(\boldsymbol{\Sigma}\right)}} \\ & q_{\mathbf{h}}^{(j)}\left(\mathbf{h}\right) \propto p\left(\mathbf{h}\right) e^{\langle ln\left[p\left(\mathbf{y}, \mathbf{x}|\mathbf{h}, \boldsymbol{\Sigma}\right)\right]\rangle_{q_{\mathbf{x}}^{(j)}\left(\mathbf{x}\right) q_{\boldsymbol{\Sigma}}^{(j-1)}\left(\boldsymbol{\Sigma}\right)}} \\ & q_{\boldsymbol{\Sigma}}^{(j)}\left(\boldsymbol{\Sigma}\right) \propto p\left(\boldsymbol{\Sigma}\right) e^{\langle ln\left[p\left(\mathbf{y}, \mathbf{x}|\mathbf{h}, \boldsymbol{\Sigma}\right)\right]\rangle_{q_{\mathbf{x}}^{(j)}\left(\mathbf{x}\right) q_{\mathbf{h}}^{(j)}\left(\mathbf{h}\right)}} \end{aligned} \tag{11}$$

To simplify the updates, the parameter priors should be conjugate meaning that the posterior is of the same type as the prior. For the channel estimate, the conjugate prior is $\mathbf{h} \sim \mathcal{CN}\left(\mathbf{h}_1, \boldsymbol{\Sigma}_1\right)$ and for the covariance, it is the inverse-Wishart distribution [8].

For the channel estimate, using (2)-(3) and the fact that the noise and prior is Gaussian, we get

$$\begin{aligned}&-ln\left[q_{\mathbf{h}}^{(j)}(\mathbf{h})\right]+Z_1=(\mathbf{h}-\mathbf{h}_1)^H\boldsymbol{\Sigma}_1^{-1}(\mathbf{h}-\mathbf{h}_1)\\&\quad+\sum_{i=1}^{N_e}\left\langle(\mathbf{y}_i-\mathbf{X}_i\mathbf{h})^H\left\langle\boldsymbol{\Sigma}^{-1}\right\rangle_{q_{\boldsymbol{\Sigma}}^{(j-1)}(\boldsymbol{\Sigma})}(\mathbf{y}_i-\mathbf{X}_i\mathbf{h})\right\rangle_{q_{\mathbf{x}}^{(j)}(\mathbf{x})}\end{aligned}\tag{12}$$

with Z indicating a normalization constant. Due to the choice of a conjugate prior, the posterior is Gaussian and given by $q_{\mathbf{h}}^{(j)}(\mathbf{h})\sim\mathcal{CN}\left(\mathbf{h}_{MAP}^{(j)},\boldsymbol{\Sigma}_{\mathbf{h}}^{(j)}\right)$ with covariance and mean

$$\begin{aligned}&\boldsymbol{\Sigma}_{\mathbf{h}}^{(j)}=\left(\sum_{i=1}^{N_e}\left\langle\mathbf{X}_i^H\left\langle\boldsymbol{\Sigma}^{-1}\right\rangle_{q_{\boldsymbol{\Sigma}}^{(j-1)}(\boldsymbol{\Sigma})}\mathbf{X}_i\right\rangle_{q_{\mathbf{x}}^{(j)}(\mathbf{x})}+\boldsymbol{\Sigma}_1^{-1}\right)^{-1}\\&\mathbf{h}_{MAP}^{(j)}=\\&\quad\boldsymbol{\Sigma}_{\mathbf{h}}^{(j)}\left(\sum_{i=1}^{N_e}\left\langle\mathbf{X}_i^H\right\rangle_{q_{\mathbf{x}}^{(j)}(\mathbf{x})}\left\langle\boldsymbol{\Sigma}^{-1}\right\rangle_{q_{\boldsymbol{\Sigma}}^{(j-1)}(\boldsymbol{\Sigma})}\mathbf{y}_i+\boldsymbol{\Sigma}_1^{-1}\mathbf{h}_1\right)\end{aligned}\tag{13}$$

The distribution of the noise covariance is

$$\begin{aligned}&-ln\left[q_{\boldsymbol{\Sigma}}^{(j)}(\boldsymbol{\Sigma})\right]+ln\left[p(\boldsymbol{\Sigma})\right]-N_e ln\left|\boldsymbol{\Sigma}\right|+Z_2\\&=\sum_{i=1}^{N_e}\left\langle(\mathbf{y}_i-\mathbf{X}_i\mathbf{h})^H\boldsymbol{\Sigma}^{-1}(\mathbf{y}_i-\mathbf{X}_i\mathbf{h})\right\rangle_{q_{\mathbf{x}}^{(j)}(\mathbf{x})q_{\mathbf{h}}^{(j)}(\mathbf{h})}\\&=tr\left\{\boldsymbol{\Sigma}^{-1}\sum_{i=1}^{N_e}\left\langle(\mathbf{y}_i-\mathbf{X}_i\mathbf{h})(\mathbf{y}_i-\mathbf{X}_i\mathbf{h})^H\right\rangle_{q_{\mathbf{x}}^{(j)}(\mathbf{x})q_{\mathbf{h}}^{(j)}(\mathbf{h})}\right\}\\&=tr\left\{\boldsymbol{\Sigma}^{-1}\mathbf{S}^{(j)}\right\}\end{aligned}\tag{14}$$

where $\mathbf{S}^{(j)}$ is the sample covariance averaged over the posteriors. It can be shown [8], that for the inverse-Wishart prior $\boldsymbol{\Sigma}\sim\mathcal{CW}^{-1}(\nu,\boldsymbol{\Sigma}_2)$, we get

$$\left\langle\boldsymbol{\Sigma}^{-1}\right\rangle_{q_{\boldsymbol{\Sigma}}^{(j)}(\boldsymbol{\Sigma})}=(N_e+\nu)\left(\mathbf{S}^{(j)}+\boldsymbol{\Sigma}_2\right)^{-1}\tag{15}$$

which for the noninformative prior $\boldsymbol{\Sigma}\sim\mathcal{CW}^{-1}(0,\mathbf{0})$ is equivalent to the ML covariance estimate. The conjugate priors can therefore be interpreted as inserting virtual observations into the estimation.

The VBE-step is similar to the traditional BCJR algorithm, only now we average over the posterior distribution of the parameters. The required state transition probabilities $\gamma(\mathbf{y}_i\mid\mathbf{X}_i,\boldsymbol{\theta})$ are therefore of the form

$$\begin{aligned}&-ln\left[\gamma(\mathbf{y}_i\mid\mathbf{X}_i,\boldsymbol{\theta})\right]+Z_3\\&=\left\langle(\mathbf{y}_i-\mathbf{X}_i\mathbf{h})^H\left\langle\boldsymbol{\Sigma}^{-1}\right\rangle_{q_{\boldsymbol{\Sigma}}^{(j-1)}(\boldsymbol{\Sigma})}(\mathbf{y}_i-\mathbf{X}_i\mathbf{h})\right\rangle_{q_{\mathbf{h}}^{(j-1)}(\mathbf{h})}\\&=Z_4-2Re\left\{\mathbf{y}_i^H\left\langle\boldsymbol{\Sigma}^{-1}\right\rangle_{q_{\boldsymbol{\Sigma}}^{(j-1)}(\boldsymbol{\Sigma})}\mathbf{X}_i\left\langle\mathbf{h}\right\rangle_{q_{\mathbf{h}}^{(j-1)}(\mathbf{h})}\right\}\\&\quad+tr\left\{\left\langle\mathbf{h}\mathbf{h}^H\right\rangle_{q_{\mathbf{h}}^{(j-1)}(\mathbf{h})}\mathbf{X}_i^H\left\langle\boldsymbol{\Sigma}^{-1}\right\rangle_{q_{\boldsymbol{\Sigma}}^{(j-1)}(\boldsymbol{\Sigma})}\mathbf{X}_i\right\}\end{aligned}\tag{16}$$

As the posterior distribution of the channel estimate is Gaussian, we have

$$\begin{aligned}\left\langle\mathbf{h}\right\rangle_{q_{\mathbf{h}}^{(j)}(\mathbf{h})}&=\mathbf{h}_{MAP}^{(j)}\\\boldsymbol{\Sigma}_m^{(j)}\triangleq\left\langle\mathbf{h}\mathbf{h}^H\right\rangle_{q_{\mathbf{h}}^{(j)}(\mathbf{h})}&=\mathbf{h}_{MAP}^{(j)}\left(\mathbf{h}_{MAP}^{(j)}\right)^H+\boldsymbol{\Sigma}_{\mathbf{h}}^{(j)}\end{aligned}\tag{17}$$

Inserting this into (16), we get

$$\begin{aligned}&-ln\left[\gamma(\mathbf{y}_i\mid\mathbf{X}_i,\boldsymbol{\theta})\right]+Z_3-Z_4\\&\quad=-2Re\left\{\mathbf{y}_i^H\left\langle\boldsymbol{\Sigma}^{-1}\right\rangle_{q_{\boldsymbol{\Sigma}}^{(j-1)}(\boldsymbol{\Sigma})}\mathbf{X}_i\mathbf{h}_{MAP}^{(j-1)}\right\}\\&\quad\quad+tr\left\{\boldsymbol{\Sigma}_m^{(j-1)}\mathbf{X}_i^H\left\langle\boldsymbol{\Sigma}^{-1}\right\rangle_{q_{\boldsymbol{\Sigma}}^{(j-1)}(\boldsymbol{\Sigma})}\mathbf{X}_i\right\}\end{aligned}\tag{18}$$

The exchange of soft-information between the data and parameter estimators is now complete with the complexity being similar to that of the equivalent EM-algorithm per iteration.

5. NUMERICAL EXAMPLE AND DISCUSSION

In order to indicate the advantage of the VBEM-algorithm and keep things as simple as possible, a single-antenna noise-limited GSM-like system is considered. The GSM system has a burst structure with $N_f=142+6$ transmitted symbols, including the 6 so-called tailbits, and has $N_{tr}=26$ known training symbols placed in the middle. The noise is assumed to be Additive White Gaussian Noise (AWGN) and the noise covariance estimation therefore reduces to a scalar variance estimation. The used channel model is the GSM Typical Urban (TU) multipath channel profile [9] with a speed of 0 km/h and using ideal frequency hopping. This ensures that the channel stays constant over a burst and that a new channel is drawn from the distribution for every burst, i.e. making it block-fading. The overall length of the transmission pulse-shaping and channel model is $L=7$. To make a fair comparison with the EM-algorithm and not go into a discussion on the correctness of various choices of priors, only noninformative priors are used for the VBEM-algorithm, i.e. $\boldsymbol{\Sigma}_1^{-1}=\mathbf{0}$ and $\boldsymbol{\Sigma}\sim\mathcal{CW}^{-1}(0,\mathbf{0})$.

A difference between the considered system and a GSM system is, that the considered modulation is linearized in order to eliminate the non-linearities introduced by the GMSK modulation used in GSM. The resulting linear modulation is simply a BPSK modulation with a rotation of $\pi/2$ per symbol.

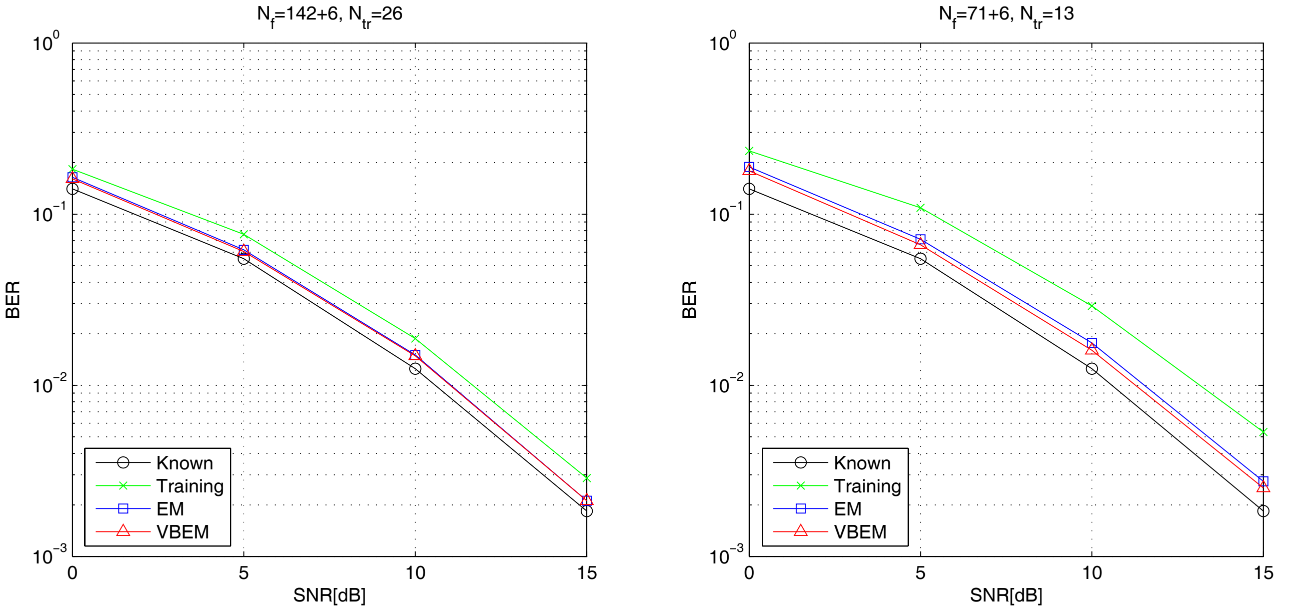

Fig. 1. Simulation of a GSM-like system using a TU0iFH channel profile, $N_t = N_r = 1$, $L = 7$.

On the left of Fig. 1, the Bit Error Rate (BER) of the above mentioned system is plotted. The results termed "Known" and "Training" are respectively the BER using the correct parameters and using only the training symbols for estimation. It can be seen that the BER of the EM and VBEM estimators are pretty much the same, although the VBEM estimator is actually better. The reason for this result is, that the number of observations is large compared to the number of parameters to be estimated. This makes the posterior distribution highly peaked around the ML solution effectively making the VBEM-algorithm fall back to the EM-algorithm.

However, changing the ratio between the number of estimated parameters and the number of observations affects the posterior distribution. On the right of Fig. 1, the length of the GSM burst has been reduced to half its original size leading to a less peaked posterior. The result is that the EM-algorithm now performs worse than the VBEM-algorithm, as the latter incorporates knowledge about the uncertainty in the parameters. The VBEM-algorithm is therefore beneficial when "few" observations are present or when "a lot" of parameters have to be estimated. This little example illustrates the advantage of the VBEM-algorithm for systems employing short packet structures and/or MIMO systems with many parameters to be estimated from a limited number of observations.

6. REFERENCES

[1] C. N. Georghiades and J. C. Han, "Sequence Estimation in the Presence of Random Parameters via the EM algorithm," *IEEE Trans. Commun.*, vol. 45, pp. 300–308, Mar. 1997.

[2] E. Chiavaccini and G. M. Vitetta, "MAP Symbol Estimation on Frequency-Flat Rayleigh Fading Channels Via a Bayesian EM Algorithm," *IEEE Trans. Commun.*, vol. 49, pp. 1869–1872, Nov. 2001.

[3] M. Nissil and S. Pasupathy, "Adaptive Bayesian and EM-Based Detectors for Frequency-Selective Fading Channels," *IEEE Trans. Commun.*, vol. 51, pp. 1325–1336, Aug. 2003.

[4] L. P. B. Christensen, "Minimum Symbol Error Rate Detection in Single-input Multiple-output Channels with Markov Noise," in *IEEE SPAWC Workshop*, 2005, pp. 236–240.

[5] N. M. Laird A. P. Dempster and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Royal Stat. Soc.*, vol. 39, pp. 1–38, 1977.

[6] M. Beal, *Variational Algorithms For Approximate Bayesian Inference*, Ph.D. thesis, 2003.

[7] M. Beal and Z. Ghahramani, "The Variational Bayesian EM Algorithm for Incomplete Data: with Application to Scoring Graphical Model Structures," *Bayesian Stat., Oxford University Press*, vol. 7, pp. 453–464, 2003.

[8] C. P. Robert, *The Bayesian Choice*, Springer, 1994.

[9] 3GPP TS 45.005, *3GPP TSG GERAN; Radio transmission and reception (Release 5)*.