Expectation maximization for joint decoding and phase estimation

Justin Dauwels, Sascha Korl, and Hans-Andrea Loeliger

Abstract—We apply the expectation maximization (EM) algorithm to the problem of joint decoding and phase estimation. We investigate a model in which (1) the phase is constant; (2) the phase performs a Gaussian random walk. In earlier work, we presented a "local" view of EM, in which EM is interpreted as message passing on factor graphs. We demonstrate how phase estimation algorithms can be derived mechanically within this setting. By the example of the random walk phase model, we illustrate how EM can be extended to parameter vectors (or processes) with nontrivial priors. For this model, the M-step can not be carried out analytically; we show how this problem can be handled solved by steepest descent, which we also interpret as summary propagation on the factor graph. We compare the EM based phase estimators to other message passing phase estimators proposed earlier. We assess the performance of the phase estimators by comparing their mean square estimation error to the (modified) posterior Cramér-Rao bound derived in earlier work.

1 Introduction

As in [1], we consider channels of the form

$$Y_k = X_k e^{j\Theta_k} + N_k, \tag{1}$$

where X_k is a (complex) channel input symbol at time $k \in \{1, 2, \ldots, L\}$, Y_k is the corresponding received symbol, Θ_k is the unknown phase, and N_k is complex white Gaussian noise with (known) variance $2\sigma_N^2$, i.e., σ_N^2 per dimension. We assume that the channel input symbols X_k are M-PSK symbols and are protected by a low-density parity check (LDPC) code.

We will consider two different models for the evolution of the phase Θ_k:

Constant Phase: $\Theta_k = \Theta$, an unknown constant.

Random Walk:

$$\Theta_k = (\Theta_{k-1} + W_k) \bmod 2\pi, \tag{2}$$

where W_k is white Gaussian noise with known variance σ_W^2.

In the present paper, we will apply the expectation maximization [2] algorithm to derive phase estimation algorithms. In earlier work [3] (see also [4]), we have shown that EM can be viewed as message passing in factor graphs. From this perspective, EM can be implemented *mechanically* by computing *local* message update rules. We will illustrate this procedure by deriving code-aided phase estimators for both phase models. Within the setting of summary propagation on factor graphs, the EM algorithm can straightforwardly (and rigorously) be extended to parameter vectors (or processes) with nontrivial priors [3] [4]. This should become clear from our exposition on the random walk phase model. Most often (but see [8] [9]), the so-called *recursive* EM algorithm equipped with some "forgetting" mechanism is used to deal with time-varying parameters [5]–[7]; this approach is rather ad-hoc and less appropriate in the context of blockwise estimation (a.k.a "batch" estimation).

The algorithm we present for the constant phase model is not new; it has been proposed earlier by Noels et al. [10] [11]. We show here how it may be derived as message passing on factor graphs (see also [12]). The EM based algorithm we propose for the random walk phase model is to our knowledge novel. It should be straightforward to extend the algorithm to more sophisticated phase models.

This paper is structured as follows. In the next section, we discuss the factor graphs we will use. In Section 3, we first briefly review the EM algorithm in the context of factor graphs. We then derive EM based estimation algorithms for both phase models. In Section 4, we present simulation results. We offer some concluding remarks in Section 5.

2 Factor graphs

We start by reviewing the factor graphs described in [1], since we will also use them in the present

The authors are with the Dept. of Information Technology and Electrical Engineering, ETH, CH-8092 Zürich, Switzerland. Email: {dauwels, korl, loeliger}@isi.ee.ethz.ch.

This project was supported in part by the Swiss National Science Foundation grant 200021-101955.

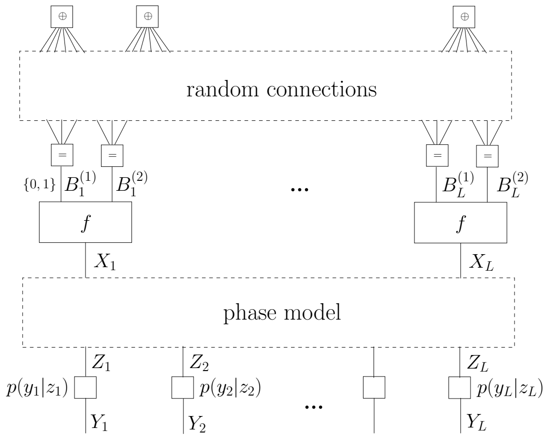

Figure 1: FFG of LDPC code and the channel model.

paper. The factor graph of Fig. 1 represents the factorization of the joint probability function of all variables in the system described in Section 1. The upper part of the graph is the indicator function of the LDPC code, with parity check nodes in the top row that are "randomly" connected to equality constraint nodes ("bit nodes"). The function f is the deterministic mapping $f : \left(B_k^{(1)}, \ldots, B_k^{(\log_2 M)}\right) \mapsto X_k$ of the bits $B_k^{(1)}, \ldots, B_k^{(\log_2 M)}$ to the symbol X_k. The nodes labelled "f" ("bit mapper nodes") correspond to the factors

$$\delta_f\left(b_k^{(1)}, \ldots, b_k^{(\log_2 M)}, x_k\right) \tag{3}$$

$$\triangleq \begin{cases} 1, & \text{if } f\left(b_k^{(1)}, \ldots, b_k^{(\log_2 M)}\right) = x_k; \\ 0, & \text{otherwise.} \end{cases} \tag{4}$$

The bottom row of the graph represents the factors $p(y_k|z_k) \triangleq (2\pi\sigma_N^2)^{-1}\, e^{-||y_k - z_k||^2/2\sigma_N^2}$.

The "phase model" in Figure 1 is detailed in Figures 2 and 3. In these figures, S_k is defined as $S_k \triangleq e^{j\Theta_k}$ and Z_k is defined as $Z_k \triangleq X_k S_k$. The top row of nodes ("multiply nodes") in Figures 2 and 3 represents the factors $\delta(z_k - x_k s_k)$. The function g is the deterministic mapping $g : \Theta_k \mapsto S_k$ of the phase Θ_k to S_k; the nodes labelled "g" in Figures 2 and 3 represent the factors $\delta(s_k - e^{j\theta_k})$. The equality constraint node in Figure 2 imposes the constraint $\Theta_k = \Theta$, for all k. In Fig. 3, the nodes labeled $p(\theta_k|\theta_{k-1})$ ("phase noise nodes") represent the factors $p(\theta_k|\theta_{k-1}) \triangleq (2\pi\sigma_W^2)^{-1/2}\ \sum_{n\in\mathbb{Z}} e^{-((\theta_k-\theta_{k-1})+n2\pi)^2/2\sigma_W^2}$.

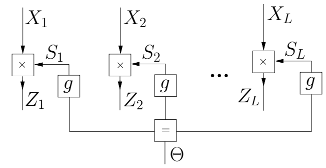

Figure 2: Factor graph of the constant-phase model.

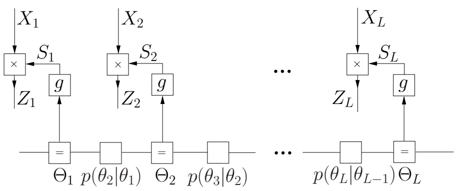

Figure 3: Factor graph of the random-walk phase model.

3 Algorithms

We start from the exposition in [3], where it is shown how the EM algorithm can be viewed as message passing on factor graphs (see also [4]). Consider the factorization

$$f(x, \theta) \triangleq f_A(\theta) f_B(x, \theta), \tag{5}$$

which is represented by the factor graph of Fig. 4.

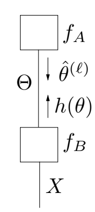

Figure 4: Factor graph of (5).

In this setup, EM amounts to iterative computation of the following messages [3]:

Upwards message $h(\theta)$:

$$h(\theta) = \frac{\int_x f_B(x, \hat{\theta}^{(\ell)}) \log f_B(x, \theta)}{\int_x f_B(x, \hat{\theta}^{(\ell)})} \tag{6}$$

$$= \mathrm{E}_{p_B}[\log f_B(x, \theta)], \tag{7}$$

where E_{p_B} denotes the expectation with respect to the probability distribution

$$p_B(x|\hat{\theta}^{(\ell)}) \triangleq \frac{f_B(x, \hat{\theta}^{(\ell)})}{\int_{x'} f_B(x', \hat{\theta}^{(\ell)})}. \quad (8)$$

Downwards message $\hat{\theta}^{(\ell+1)}$:

$$\hat{\theta}^{(\ell+1)} = \underset{\theta}{\operatorname{argmax}}(\log f_A(\theta) + h(\theta)). \quad (9)$$

The computations (7) and (9) may be simplified when f_A and f_B have "nice" factorizations; this is the case for both phase models. In the constant phase model, the maximization step (9) can be carried out analytically; in the random walk phase model, the maximization step is intractable and we will apply steepest descent. First we treat the constant phase model, then the random walk phase model.

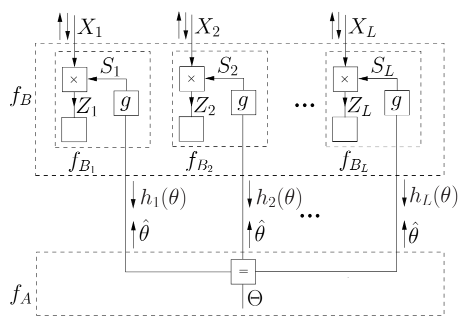

Figure 5: EM in the constant-phase model.

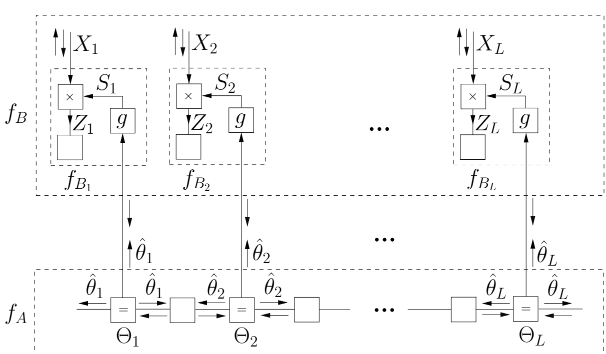

Figure 6: Steepest descent/EM in the random-walk phase model.

3.1 Constant phase model

We start by choosing the "boxes" f_A and f_B appropriately. Since we wish to estimate the phase Θ, we choose the Θ edges as cut set of edges that separates the boxes f_A and f_B (see Fig. 2). The box f_A contains all nodes that correspond to the *prior* of Θ, i.e., all nodes that are solely connected to Θ edges. In this case, f_A exclusively contains the equality constraint node Θ, as illustrated in Fig. 5; all other nodes belong to the box f_B. Note that the box f_B has a "nice" structure, i.e., f_B has a nontrivial factorization. We now apply the message update rules (7) and (9) to the factor graph of Fig. 5. The message $h(\theta)$ equals (see also Eq. (18)–(20) in [3])

$$h(\theta) = \sum_{k=1}^{L} \sum_{x_k} p_B(x_k|y, \hat{\theta}^{(\ell)}) \log f_B(x_k, y_k, \theta) \quad (10)$$

$$\triangleq \sum_{k=1}^{L} h_k(\theta), \quad (11)$$

where the function $f_B(x_k, y_k, \theta)$ is given by

$$f_B(x_k, y_k, \theta) \triangleq \int_{z_k} \oint_{s_k} \delta(z_k - x_k s_k)\delta(s_k - e^{j\theta}) \cdot (2\pi\sigma_N^2)^{-1}\, e^{-||y_k - z_k||^2/2\sigma_N^2} dz_k ds_k \quad (12)$$

$$= (2\pi\sigma_N^2)^{-1}\, e^{-||y_k - x_k e^{j\theta}||^2/2\sigma_N^2}. \quad (13)$$

The small dashed boxes in Fig. 5 represent the functions $f_B(x_k, y_k, \theta)$. The marginalization (12)–(13) corresponds to "closing" the boxes f_B (see [13]). The distribution $p_B(x_k|y, \hat{\theta}^{(\ell)})$ is defined as

$$p_B(x_k|y, \hat{\theta}^{(\ell)}) \triangleq \gamma_k f_B(x_k, y_k, \hat{\theta}^{(\ell)}) \mu_{X_k \to \boxtimes}(x_k), \quad (14)$$

where γ_k is a normalization factor, i.e.,

$$\gamma_k \triangleq \sum_{x_k} p_B(x_k|y, \hat{\theta}^{(\ell)}). \quad (15)$$

Note that $f_B(x_k, y_k, \hat{\theta}^{(\ell)})$ is nothing else than the message $\mu_{\boxtimes \to X_k}(x_k)$, which leaves the multiply node along the edge X_k (at the ℓ-th iteration). Therefore

$$p_B(x_k|y, \hat{\theta}^{(\ell)}) \triangleq \gamma_k \mu_{X_k \to \boxtimes}(x_k) \mu_{\boxtimes \to X_k}(x_k). \quad (16)$$

In the following, we will use the short hand notation $p_B(x_k)$ for $p_B(x_k|y, \hat{\theta}^{(\ell)})$. The phase Θ has a uniform prior, i.e., $f_A(\theta) \triangleq 1$ for all θ. The message $\hat{\theta}^{(\ell+1)}$ equals

$$\hat{\theta}^{(\ell+1)} = \operatorname*{argmax}_{\theta} h(\theta) \tag{17}$$

$$= \operatorname*{argmax}_{\theta} \sum_{k=1}^{L} \sum_{x_k} p_B(x_k) \log f_B(x_k, y_k, \theta) \tag{18}$$

$$= \arg \sum_{k=1}^{L} \sum_{x_k} p_B(x_k) \left[y_k x_k^* \right] \tag{19}$$

$$= \arg \sum_{k=1}^{L} \left[y_k \left(\tilde{\mathrm{E}}[x_k] \right)^* \right], \tag{20}$$

where the expectation $\tilde{\mathrm{E}}[x_k]$ is defined as

$$\tilde{\mathrm{E}}[x_k] \triangleq \sum_{x_k} p_B(x_k)\, x_k. \tag{21}$$

In this particular case, a closed-form expression for the messages $\hat{\theta}$ can thus be found. The expression (20) was proposed earlier in [11]. Note that the initial estimate $\theta^{(0)}$ can be generated by a classical phase estimator such as the M-law. This estimator "summarizes" over the M-PSK input symbols X_k by applying the M-th power to the received symbols; it does not use any soft information from the decoder.

3.2 Random walk phase model

We again start by determining the boxes f_A and f_B; the box f_B is chosen as in the previous subsection. The box f_A is now more interesting, as illustrated in Fig. 6: it contains the nodes $p(\theta_k|\theta_{k-1})$ besides the equality constraint nodes Θ_k. Both f_A and f_B have in this case a nontrivial structure. We now apply the message update rules (7) and (9) to the factor graph of Fig. 6. The message $h(\theta)$, summarizing the box f_B, is again given by (10), since the box f_B remained unchanged. The function $\log f_A(\theta)$ equals

$$\log f_A(\theta) = \sum_{k=2}^{L} \log p(\theta_k|\theta_{k-1}), \tag{22}$$

where

$$p(\theta_k|\theta_{k-1}) \triangleq (2\pi\sigma_W^2)^{-1/2} \sum_{n\in\mathbb{Z}} e^{-((\theta_k-\theta_{k-1})+n2\pi)^2/2\sigma_W^2}. \tag{23}$$

The message $\hat{\theta}$ is computed as

$$\hat{\theta} = \operatorname*{argmax}_{\theta} (\log f_A(\theta) + h(\theta)) \tag{24}$$

$$= \operatorname*{argmax}_{\theta} \Bigg[\sum_{k=2}^{L} \log p(\theta_k|\theta_{k-1}) + \tag{25}$$

$$\sum_{k=1}^{L} \sum_{x_k} p_B(x_k) \log f_B(x_k, y_k, \theta_k) \Bigg]. \tag{26}$$

The maximization (26) can not be carried out analytically. One may use classical optimization algorithms as for example steepest descent to solve this non-linear optimization problem. In steepest descent, one iterates the update rule

$$\hat{\theta}^{(\ell+1)} = \hat{\theta}^{(\ell)} + \lambda\, \nabla_\theta (\log f_A(\theta) + h(\theta)) \big|_{\theta=\hat{\theta}^{(\ell)}} \tag{27}$$

until a fixed point is reached or the available time is over. The parameter λ is a real positive number referred to as "step size" or "learning rate".
In earlier work [16], we have shown how applying the rule (27) can be viewed as summary propagation on factor graphs. In this particular case, the procedure is as follows (see Fig. 6):

1. The equality constraint nodes Θ_k broadcast the estimates $\hat{\theta}_k^{(\ell)}$.
2. The messages $\mu_{\boxtimes \to X_k}(x_k)$ are updated. Next, the LDPC decoder is iterated a suitable number of times, which leads to new messages $\mu_{X_k \to \boxtimes}(x_k)$.
3. The nodes f_{B_k} reply with $\frac{dh_k(\theta_k)}{d\theta_k}\Big|_{\hat{\theta}_k^{(\ell)}}$.
4. The nodes $p(\theta_k|\theta_{k-1})$ and $p(\theta_{k+1}|\theta_k)$ reply with $\frac{\partial \log p(\theta_k|\theta_{k-1})}{\partial\theta_k}\Big|_{\hat{\theta}^{(\ell)}}$ and $\frac{\partial \log p(\theta_{k+1}|\theta_k)}{\partial\theta_k}\Big|_{\hat{\theta}^{(\ell)}}$ respectively.
5. The new estimate $\hat{\theta}^{(\ell+1)}$ is computed:

$$\hat{\theta}_k^{(\ell+1)} = \hat{\theta}_k^{(\ell)} + \lambda \left(\frac{\partial \log p(\theta_k|\theta_{k-1})}{\partial\theta_k}\Big|_{\hat{\theta}_k^{(\ell)}} + \frac{\partial \log p(\theta_{k+1}|\theta_k)}{\partial\theta_k}\Big|_{\hat{\theta}_k^{(\ell)}} + \frac{dh_k(\theta_k)}{d\theta_k}\Big|_{\hat{\theta}_k^{(\ell)}} \right). \tag{28}$$

6. Iterate 1–5.

As usual, several update schedules are possible. For example, it is not necessary to perform Step 2 at

each iteration.
The (partial) derivatives required in Step 4 are computed as

$$\frac{\partial \log p(\theta_k|\theta_{k-1})}{\partial \theta_k} = -\frac{\sum_{n\in\mathbb{Z}}(\theta_k-\theta_{k-1}+n2\pi)e^{-(\theta_k-\theta_{k-1}+n2\pi)^2/2\sigma_W^2}}{\sigma_W^2\sum_{n\in\mathbb{Z}}e^{-(\theta_k-\theta_{k-1}+n2\pi)^2/2\sigma_W^2}}, \tag{29}$$

and

$$\frac{\partial \log p(\theta_k|\theta_{k-1})}{\partial \theta_{k-1}} = -\frac{\partial \log p(\theta_k|\theta_{k-1})}{\partial \theta_k}. \tag{30}$$

If σ_W is small, i.e., $\sigma_W \ll \pi$, then the evaluation of the RHS of (29) leads to numerical problems: for large values of the difference $\theta_k - \theta_{k-1}$, both the numerator and denominator in the RHS of (29) are zero. This issue can be circumvented by the approximation:

$$\frac{\partial \log p(\theta_k|\theta_{k-1})}{\partial \theta_k} \approx -\frac{\theta_k-\theta_{k-1}}{\sigma_W^2}, \tag{31}$$

which is a good approximation as long as $\theta_k - \theta_{k-1} \ll \pi$. In our simulations, we use the approximation (31), since the latter condition is satisfied. The derivative $dh_k(\theta_k)/d\theta_k$ required in Step 4 is given by

$$\frac{dh_k(\theta_k)}{d\theta_k} = \sum_{x_k} p_B(x_k)\frac{\partial \log f(x_k,y_k,\theta_k)}{\partial \theta_k}, \tag{32}$$

where

$$\frac{\partial \log f(x_k,y_k,\theta_k)}{\partial \theta_k} = -\frac{1}{\sigma_N^2}\big[\sin\theta_k \mathrm{Re}(x_k y_k^*) + \cos\theta_k \mathrm{Im}(x_k y_k^*)\big], \tag{33}$$

and a^* stands for the complex conjugate of a. Consequently,

$$\frac{dh_k(\theta_k)}{d\theta_k} = -\frac{1}{\sigma_N^2}\big[\sin\theta_k \mathrm{Re}(\tilde{\mathrm{E}}[x_k]y_k^*) \tag{34}$$

$$+ \cos\theta_k \mathrm{Im}(\tilde{\mathrm{E}}[x_k]y_k^*)\big]. \tag{35}$$

The initial phase $\theta^{(0)}$ may simply be constant, i.e., $\theta_k^{(0)} = \hat{\theta}$ for all k, where $\hat{\theta}$ may be computed by the M-law. Note that the above algorithm (Step 1–6) is very similar to the steepest descent algorithm presented in [1]. There, steepest descent is applied to sum-product messages, whereas it is applied to EM messages here. If one replaces h_k by $\log \mu_{g\to\Theta_k}$ in the above procedure (more precisely in Step 3 and 5), one obtains the gradient based sum-product algorithm of [1].

4 Results and discussion

We assessed the performance of the presented algorithms through simulations, using the same setup as in [1]; the symbols are protected by a fixed rate 1/2 LDPC code of length 100 that was randomly generated and was not optimized for the channel at hand. The factor graph of the code does not contain cycles of length four. The degree of all bit nodes equals 3; the degrees of the check nodes are distributed as follows: 1, 14, 69 and 16 check nodes have degree 4, 5, 6 and 7 respectively. The symbol constellation was Gray-encoded 4-PSK. We iterated three times between the LDPC decoder and the phase estimator, each time with hundred iterations inside the LDPC decoder. We did not iterate between the LDPC decoder and the mapper. We optimized the (constant) step size parameter λ. In the simulations, we assume the phase ambiguity has been resolved (see [12]). In Figure 7, the proposed phase synchronizers are compared to some synchronizers of [1] (see also [15]) in terms of mean squared (phase) estimation error (MSEE) as a function of the SNR; the figure shows that EM-based algorithms have a similar performance as the gradient based methods of [1]. The particle based algorithms have the lowest MSEE, but are the most complex. At high SNR, i.e., SNR > 3dB, all algorithms except the M-law achieve the modified Cramér-Rao bound computed in [16]. Also the frame error rates (FER) of the proposed phase synchronizers (not shown here) are comparable to the FER of the gradient based methods of [1].
We have implemented several schemes to generate the initial estimate $\theta^{(0)}$ in the random walk phase model; they all led to about the same performance (results not shown here). We tried out several schedules for updating the messages in the random walk phase model (both in the EM based as the sum-product based approach), i.e., alternations of forward and backward sweeps, "parallel" updates and others; again, we did not observe any effect on the performance. We noticed that the optimal learning rate λ (see (28)) was *identical* for each value of σ_W and σ_N. A popular technique to cope with time-varying channel parameters is to introduce "forgetting factors" [5]–[7]. The latter need to be optimized for *each* specific value of the channel parameters, since the dynamics of the channel is solely handled by those factors; in the algorithms we proposed, the dynamics is encoded in the prior f_A: the estimates are not only updated based on the

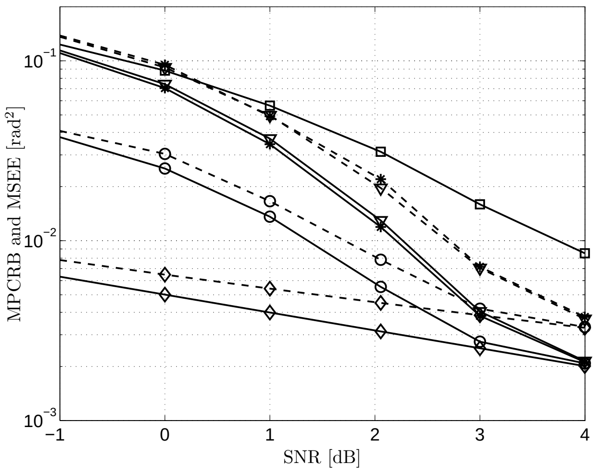

Figure 7: MPCRB and MSEE for $\sigma_W^2 = 0$ and $\sigma_W^2 = 10^{-4}$ rad^2; shown are (1) the MPCRBs of [16] (diamonds); (2) the MSEE of the particle based (circles), EM based (stars), and the sum-product steepest descent based (triangles) synchronizers; (3) the MSEE of the M-law (squares), only for $\sigma_W = 0$.

local observations, but also based on f_A (see (28)), which explains why the optimal λ does not vary with σ_W and σ_N.

5 Conclusions

We have derived phase estimators for two phase models by straightforwardly applying the EM message update rules on a factor graph. In the case of the random walk phase model, the maximization step was infeasible; we carried out the maximization by steepest descent, which we also viewed as summary propagation on factor graphs. The performance of the resulting phase estimators is comparable to gradient based algorithms proposed earlier; they achieve the (modified) Cramér-Rao bound at high SNR.

The goal of this paper was to illustrate that (1) EM based algorithms can be derived mechanically from a factor graph; (2) EM can straightforwardly be extended to parameters with non-trivial priors, or in other words, that there is no need to introduce forgetting factors; (3) when the M-step can not be performed analytically, it can be carried out by steepest descent, which also can be interpreted as summary propagation on a factor graph; (4) EM based algorithms can achieve high performance in spite of their low complexity, as already observed by numerous researchers in the past.

References

[1] J. Dauwels and H. -A. Loeliger, "Phase estimation by message passing," *IEEE International Conference on Communications*, Paris, France, June 20–24, 2004, pp. 523–527.

[2] A. P. Dempster, N. M. Laird, and D. B. Rubin "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society,* B 39, pp. 1–38, 1977.

[3] J. Dauwels, S. Korl, and H.-A. Loeliger, "Expectation maximization on factor graphs", submitted to the *IEEE Int. Symp. Information Theory 2005.* Available online from `http://www.dauwels.com/files/EM_FG.pdf`.

[4] A. W. Eckford and S. Pasupathy, "Iterative multiuser detection with graphical modeling" *IEEE International Conference on Personal Wireless Communications,* Hyderabad, India, 2000.

[5] D. M. Titterington, "Recursive parameter estimation using incomplete data," R. Roy. Stat. Soc., vol. 46(B), pp. 256–267, 1984.

[6] E. Weinstein, M. Feder, A. V. Oppenheim, "Sequential algorithms for parameter estimation based on the Kullback-Leibler information measure," *IEEE Transactions on Acoustics, Speech, and Signal Processing,* vol. 38, Sept. 1990, pp. 1652-1654.

[7] V. Krishnamurthy and J.B. Moore, "On-line estimation of hidden Markov model parameters based on the Kullback-Leibler information measure," *IEEE Transactions on Signal Processing,* vol. 41, Aug. 1993, pp. 2557–2573.

[8] H. Zamiri-Jafarian and S. Pasupathy, "EM-based recursive estimation of channel parameters," *IEEE Transactions on Communications,* vol. 47, Sept. 1999, pp. 1297–1302.

[9] L. Frenkel and M. Feder, "Recursive estimate-maximize (EM) algorithms for time varying parameters with applications to multiple target tracking," *IEEE International Conference on Acoustics, Speech, and Signal Processing,* vol. 3, 9–12 May, 1995, pp. 2068–2071.

[10] N. Noels, C. Herzet, A. Dejonghe, V. Lottici, H. Steendam, M. Moeneclaey, M. Luise and L. Vandendorpe, "Turbo synchronization : an EM algorithm interpretation," *International Conference on Communications 2003, ICC'03,* Anchorage, Alaska, May 11–15, 2003, pp. 2933–2937

[11] H. Steendam, N. Noels, M. Moeneclaey, "Iterative carrier phase synchronization for low-density parity-check coded systems", *International Conference on Communications 2003*, Anchorage, Alaska, May 11–15, 2003, pp. 3120–3124.

[12] J. Dauwels, H. Wymeersch, H.-A. Loeliger and M. Moeneclaey, "Phase Estimation and Phase Ambiguity Resolution by Message Passing," *Proc. 11th International Conference on Telecommunications and Networking,* pp. 150–155, Fortaleza, Brazil, August 1–6, 2004.

[13] H.-A. Loeliger, "An introduction to factor graphs," *IEEE Signal Proc. Mag.,* Jan. 2004, pp. 28–41.

[14] J. Dauwels, S. Korl, and H.-A. Loeliger, "Steepest descent on factor graphs", submitted to the *IEEE ITSOC Information Theory Workshop 2005 on Coding and Complexity.* Available online from `http://www.dauwels.com/files/steepest_descent.pdf`.

[15] G. Colavolpe and G. Caire, "Iterative decoding in the presence of strong phase noise," *IEEE Journal on Selected Areas in Communications,* Differential and Noncoherent Wireless Communications, to appear.

[16] J. Dauwels, "Posterior Cramér-Rao bounds for estimation on trees," submitted to *IEEE Int. Symp. Information Theory, 2005.* Available online from `http://www.dauwels.com/files/CRB_long.pdf`.