ITERATIVE SYNCHRONIZATION : EM ALGORITHM VERSUS NEWTON-RAPHSON METHOD

C. Herzet, X. Wautelet, V. Ramon and L. Vandendorpe

Communications Laboratory, Université catholique de Louvain, Pl. du Levant 2, B1348 Louvain-la-Neuve, Belgium

ABSTRACT

This paper deals with iterative maximum-likelihood synchronization of a scalar parameter. An efficient implementation of the Newton-Raphson (NR) maximum-search method is proposed. Considering the latter implementation, the NR approach is shown to be an attractive alternative to synchronization methods based on the expectation-maximization (EM) algorithm. Simulation results for the case of phase-offset synchronization show that NR method usually increases the speed of convergence of the synchronization algorithm.

1. INTRODUCTION

Since the discovery of the turbo principle by Berrou and Glavieux, code-aided (CA) synchronization has become a central problem. Conventional non-data-aided (NDA) synchronizers, which operate properly at medium-to-high signal-to-noise ratios (SNR), exhibit poor performance at the very low operating point of turbo receivers. Unfortunately, maximum-likelihood (ML) CA synchronization is an intrinsicly complex problem. Therefore, iterative synchronization algorithms have been proposed to solve the CA ML problem, see e.g. [1], [2], [3], [4]. In [1], [2], the authors propose to solve the ML problem via the the well-known expectation-maximization (EM) algorithm [5]. In [3], [4], the synchronization problem is placed into the context of factor-graph representation and the sum-product algorithm [6]. This approach leads to iterative synchronization algorithms in which an intermediate maximum-likelihood problem has to be solved at each iteration. In this paper, we show that ML synchronization problems may be solved efficiently using the Newton-Raphson maximization method. In particular, we show that simple analytical expressions may be found for the first and second derivatives of the likelihood function.

2. ML ESTIMATION IN THE PRESENCE OF A NUISANCE VECTOR

Let $\mathbf{r}$ denote a vector of observations and let b indicate a deterministic scalar parameter to be estimated from the observations $\mathbf{r}$. Assume that $\mathbf{r}$ also depends on a random discrete-valued nuisance vector $\mathbf{a}$ independent of b and with a priori probability $p(\mathbf{a})$. The problem addressed in this paper is to find the ML estimate $\hat{b}_{ML}$ of b that is to say the solution of

$$\hat{b}_{ML} = \arg\max_{\tilde{b}} \Big\{ \ln p(\mathbf{r}|b=\tilde{b}) \Big\}, \tag{1}$$

where

$$p(\mathbf{r}|b) = \sum_{\mathbf{a}} p(\mathbf{r}|\mathbf{a},b)\, p(\mathbf{a}), \tag{2}$$

and $\tilde{b}$ is a trial value of b. Most of the time there is unfortunately no analytical solution to such a problem. In this case one has to use iterative numerical methods to find the solution of (1). In the sequel we compare two well-known methods used to iteratively search the maximum of a function: the expectation-maximization (EM) algorithm and the Newton-Raphson (NR) method.

3. THE EXPECTATION-MAXIMIZATION ALGORITHM

Like many other methods, the EM algorithm [5] is a method for finding the zeros of a function, namely $\frac{\partial \ln p(\mathbf{r}|b)}{\partial b}$. Using the formalism of the EM algorithm, let us set $\mathbf{r}$ as the *incomplete* data set and $\mathbf{z} \triangleq [\mathbf{r}, \mathbf{a}]$ as the *complete* data set. The EM algorithm states that the sequence $\hat{b}^{(n)}$ defined as

$$\mathcal{Q}(b, \hat{b}^{(n-1)}) = \int_{\mathbf{z}} p(\mathbf{z}|\mathbf{r}, \hat{b}^{(n-1)}) \; \ln p(\mathbf{z}|b)\; d\mathbf{z} \tag{3}$$

$$\hat{b}^{(n)} = \arg\max_{\tilde{b}} \Big\{ \mathcal{Q}(b=\tilde{b}, \hat{b}^{(n-1)}) \Big\} \tag{4}$$

converges towards either a saddle point or a maximum of (2). Local minima are avoided. In practice, the final convergence point of the EM algorithm depends on the initialization value. The EM algorithm exhibits a linear speed of convergence. Its rate of convergence is a function of the quantity of *missing information* in the incomplete data set [5].

In the particular case of parameter b estimation in the presence of independent nuisance vector $\mathbf{a}$, the $\mathcal{Q}$-function has the following structure (see [1]):

$$\begin{aligned}\mathcal{Q}(\tilde{b},\hat{b}^{(n-1)}) &= \sum_{\mathbf{a}} p(\mathbf{a}|\mathbf{r},b=\hat{b}^{(n-1)})\ \ln p(\mathbf{r}|\mathbf{a},b=\tilde{b}) \\ &= \mathbb{E}_{\mathbf{a}|\mathbf{r},b=\hat{b}^{(n-1)}}\Big[\ln p(\mathbf{r}|\mathbf{a},b=\tilde{b})\Big]. \end{aligned} \tag{5}$$

The solution of (1) can therefore be found iteratively by only using posterior probabilities $p(\mathbf{a}|\mathbf{r},b)$ and the log-likelihood function $\ln p(\mathbf{r}|\mathbf{a},b)$.

4. THE NEWTON-RAPHSON METHOD

Another well-known method for finding the zeros of a function is the Newton-Raphson (NR) method. The sequence of estimates computed by the NR method may be written as

$$\hat{b}^{(n)} = \hat{b}^{(n-1)} - \left(\frac{\partial^2}{\partial b^2}\ln p(\mathbf{r}|b)\right)^{-1}\left(\frac{\partial}{\partial b}\ln p(\mathbf{r}|b)\right)_{\big|b=\hat{b}^{(n-1)}}. \tag{6}$$

The NR method may converge to saddle points and maxima but also to minima. As the EM algorithm, its ultimate point of convergence depends on the initial estimate.

The speed of convergence of the NR approach is quadratic [7]. This very-fast speed of convergence is often regarded as the major strength of the NR method. Apart from its possible convergence to minima, the evaluation of the first and second derivatives of $\ln p(\mathbf{r}|b)$ is often considered as one of the major drawback of the method. In [8], an efficient way to compute the first derivative is proposed, i.e.

$$\left(\frac{\partial}{\partial b}\ln p(\mathbf{r}|b)\right)_{\big|b=\tilde{b}} = \left(\mathbb{E}_{\mathbf{a}|\mathbf{r},b}\left[\frac{\partial}{\partial b}\ln p(\mathbf{r}|\mathbf{a},b)\right]\right)_{\big|b=\tilde{b}}. \tag{7}$$

In this paper, we derive an expression (see the appendix for the details) which enables to efficiently compute the second derivative of $\ln p(\mathbf{r}|b)$, namely

$$\begin{aligned}\left(\frac{\partial^2}{\partial b^2}\ln p(\mathbf{r}|b)\right)_{\big|b=\tilde{b}} &= \left(\mathbb{E}_{\mathbf{a}|\mathbf{r},b}\left[\frac{\partial^2}{\partial b^2}\ln p(\mathbf{r}|\mathbf{a},b)\right]\right)_{\big|b=\tilde{b}} \\ &+ \left(\mathbb{E}_{\mathbf{a}|\mathbf{r},b}\left[\left(\frac{\partial}{\partial b}\ln p(\mathbf{r}|\mathbf{a},b)\right)^2\right]\right)_{\big|b=\tilde{b}} \\ &- \left(\mathbb{E}_{\mathbf{a}|\mathbf{r},b}\left[\frac{\partial}{\partial b}\ln p(\mathbf{r}|\mathbf{a},b)\right]\right)^2_{\big|b=\tilde{b}}. \end{aligned} \tag{8}$$

As the EM algorithm, the NR method may therefore be implemented by using posterior probabilities $p(\mathbf{a}|\mathbf{r},b)$ and the log-likelihood function $\ln p(\mathbf{r}|\mathbf{a},b)$.

5. ITERATIVE SYNCHRONIZATION

In this section we apply the general frameworks of the previous sections to the particular case of the synchronization of a digital data-modulated bandpass signal. In this context, the nuisance parameter vector $\mathbf{a}$ contains the values of the K unknown transmitted data symbols $(a_0, a_1, ..., a_{K-1}) \in \mathcal{A}^K$ where $\mathcal{A}$ is the constellation alphabet. We assume that the parameter b to estimate can be either the symbol timing τ, the carrier frequency offset ν or the phase offset θ. For the sake of simplicity we will consider an AWGN channel in the sequel. The baseband received samples $r(lT_s)$ can then be written as

$$r(lT_s) = \sum_{k=0}^{K-1} a_k\, p(lT_s - kT - \tau)\, e^{j(2\pi\nu lT_s+\theta)} + w(lT_s), \tag{9}$$

where T_s is the sampling period, T is the symbol period, $p(t)$ is a unit-energy square-root raised-cosine pulse and $w(lT_s)$ is a complex-valued AWGN with variance σ_w^2.

Neglecting terms independent of b, the log-likelihood function $\ln p(\mathbf{r}|\mathbf{a},b)$ present in (5), (7) and (8) can be written as

$$\ln p(\mathbf{r}|\mathbf{a},b) = -\frac{2}{\sigma_w^2}\,\mathrm{Re}\Big\{\sum_{k=0}^{K-1} a_k^*\, y_k(\nu,\tau)\, e^{-j\theta}\Big\}, \tag{10}$$

where

$$y_k(\nu,\tau) \triangleq T_s \sum_l r(lT_s)\, e^{-j(2\pi\nu lT_s)}\, p(lT_s - kT - \tau).$$

In the remainder of this section we particularize the EM and NR equations to (10).

5.1. EM Synchronization

Let us define for each transmitted symbol a_k

$$\eta_k(\hat{b}^{(n-1)}) \triangleq \sum_{a\,\in\mathcal{A}} a\ p(a_k = a|\mathbf{r},\hat{b}^{(n-1)}). \tag{11}$$

Using this definition and replacing $\ln p(\mathbf{r}|\mathbf{a},b)$ by (10) in (5), we get

$$\mathcal{Q}(b,\hat{b}^{(n-1)}) = -\frac{2}{\sigma_w^2}\,\mathrm{Re}\Big\{\sum_{k=0}^{K-1}\eta_k^*(\hat{b}^{(n-1)})\, y_k(\nu,\tau)\, e^{-j\theta}\Big\}. \tag{12}$$

If we consider the case $b=\theta$, the maximization of (12) has an analytical solution:

$$\hat{\theta}^{(n)} = \arg\Big\{\sum_{k=0}^{K-1}\eta_k^*(\hat{\theta}^{(n-1)})\, y_k(\nu,\tau)\Big\}. \tag{13}$$

On the contrary, there are no analytical solutions for the cases $b = \tau$ or $b = \nu$. In such cases, we still have to resort to numerical maximum-search method to solve the intermediate maximization problem (4), see e.g. [2]. We also notice from (11) that the evaluation of (12) only requires the knowledge of the posterior marginals $p(a_k|\mathbf{r}, \hat{b}^{(n-1)})$. The computation of these probabilities usually dominates the complexity of the implementation of an EM synchronizer. In the case of a convolutionally-coded transmission, this requires for example a decoding operation via a BCJR algorithm.

5.2. NR Synchronization

For the sake of conciseness, we only consider the case $b = \theta$. The cases $b = \tau$ or $b = \nu$ are similar and their derivation is straightforward. Particularizing (7) and (8) to (10) with $b = \theta$, we have

$$\frac{\partial \ln p(\mathbf{r}|\theta)}{\partial \theta} = \frac{2}{\sigma_w^2} \,\mathrm{Im}\Big\{\sum_k \eta_k^*(\theta)\, y_k(\nu,\tau)\, e^{-j\theta}\Big\},$$

$$\begin{aligned}\frac{\partial^2 \ln p(\mathbf{r}|\theta)}{\partial \theta^2} = &-\frac{2}{\sigma_w^2} \,\mathrm{Re}\Big\{\sum_k \eta_k^*(\theta)\, y_k(\nu,\tau)\, e^{-j\theta}\Big\}\\ &+\frac{2}{\sigma_w^4} \,\mathrm{Re}\Big\{\sum_{k,l} \rho_{k,l}(\theta)\, y_k(\nu,\tau)\, y_l^*(\nu,\tau)\Big\}\\ &-\frac{2}{\sigma_w^4} \,\mathrm{Re}\Big\{\sum_{k,l} \varepsilon_{k,l}^*(\theta)\, y_k(\nu,\tau)\, y_l(\nu,\tau)\, e^{-j2\theta}\Big\},\end{aligned}$$

where

$$\rho_{k,l}(\theta) \triangleq \sum_{a',a\in\mathcal{A}} a^*\, a'\, p(a_k = a, a_l = a'|\mathbf{r},\theta) - \eta_k^*(\theta)\, \eta_l(\theta),$$

$$\varepsilon_{k,l}(\theta) \triangleq \sum_{a,a'\in\mathcal{A}} a\, a'\; p(a_k = a, a_l = a'|\mathbf{r},\theta) - \eta_k(\theta)\, \eta_l(\theta).$$

From the latter definitions, we note that the implementation of the NR maximization method requires the knowledge of joint probabilities $p(a_k, a_l|\mathbf{r}, \theta)$. These probabilities are not available in practical receivers. We will therefore assume that $p(a_k, a_l|\mathbf{r}, \theta)$ may be well-approximated as

$$p(a_k, a_l|\mathbf{r},\theta) \simeq p(a_k|\mathbf{r},\theta)\, p(a_l|\mathbf{r},\theta), \qquad (14)$$

i.e. by the product of its marginals. Doing this approximation, $\varepsilon_{k,l} = 0$ and $\rho_{k,l} = 0\ \forall\, k \neq l$. The expression of the second derivative therefore simplifies as

$$\begin{aligned}\frac{\partial^2 \ln p(\mathbf{r}|\theta)}{\partial \theta^2} = &-\frac{2}{\sigma_w^2} \,\mathrm{Re}\Big\{\sum_k \eta_k^*(\theta)\, y_k(\nu,\tau)\, e^{-j\theta}\Big\}\\ &+\frac{2}{\sigma_w^4} \,\mathrm{Re}\Big\{\sum_k \rho_{k,k}(\theta)\, |y_k(\nu,\tau)|^2\Big\}\\ &-\frac{2}{\sigma_w^4} \,\mathrm{Re}\Big\{\sum_k \varepsilon_{k,k}^*(\theta)\; y_k^2(\nu,\tau)\, e^{-j2\theta}\Big\}. \qquad (15)\end{aligned}$$

Doing approximation (14), the NR implementation complexity is usually dominated by the computation of marginals $p(a_k|\mathbf{r}, \theta)$. In such a case, the complexity *order* of one (approximated) NR iteration is similar to the EM case.

Note that (14) is no longer an approximation when $p(\mathbf{a}) = \prod_k p(a_k)$. It is for example the case for NDA synchronization. It is also the case in the context of factor-graph-based synchronization proposed in [3] and [4]. Indeed, in the latter case, it is shown in [4] that the "intermediate" likelihood function which has to be maximized at each SP iteration assumes $p(\mathbf{a}) = \prod_k p_{ext}(a_k)$ where $p_{ext}(a_k)$ is the extrinsic probability about symbol a_k. We consider the latter case in the next section.

6. NUMERICAL RESULTS

We consider carrier phase estimation for a 8-PSK transmission. Other synchronization parameters are assumed to be known and $K = 1000$. The E_s/N_0-ratio is set to 4.77dB. We asssume perfect phase ambiguity resolution so that $|\hat{\theta} - \theta| \leq \pi/8$. In the factor-graph approach [3]-[4], we have $p_{ext}(a_k) = \prod_l p_{ext}(c_{k,l})$ where $c_{k,l}$ denotes the l^{th} bit making up the k^{th} transmitted symbols. Probabilities $p_{ext}(c_{k,l})$ are assumed to derive from extrinsic log-likelihood ratios (LLRs) $L_{ext}(c_{k,l})$. We randomly generate extrinsic LLRs according to a Gaussian distribution $\mathcal{N}(\sigma^2/2, \sigma^2)$. The latter distribution is parameterized by mutual information $I(c_{k,l}, L_{ext}(c_{k,l}))$ which univoquely defines parameter σ^2. Fig. 1 shows the mean square error (MSE) as a function of the number of iterations for the EM and NR methods. We prevent the convergence of the NR method to local minima by correcting the estimate as $\hat{\theta} = \hat{\theta} - \pi/8$ when the second derivative is positive. We consider $I(c_{k,l}, L_{ext}(c_{k,l}))$ equal to 0, 0.2, 0.8 and 1. $I(c_{k,l}, L_{ext}(c_{k,l})) = 0$ corresponds to the NDA case whereas $I(c_{k,l}, L_{ext}(c_{k,l})) = 1$ corresponds to a perfect knowledge of the transmitted symbols. We note that the EM-algorithm convergence is strongly affected by the amount of information available about the transmitted symbols. On the contrary the NR convergence is always quite fast irrespective of the available symbol information. As a consequence, we note that, in the considered case, the NR method should be prefered when the amount of information about the transmitted symbols is low. On the contrary, the EM algorithm seems to be the best solution when $I(c_{k,l}, L_{ext}(c_{k,l}))$ is large since it exhibits about the same speed of convergence as the NR approach while slightly less complex to implement.

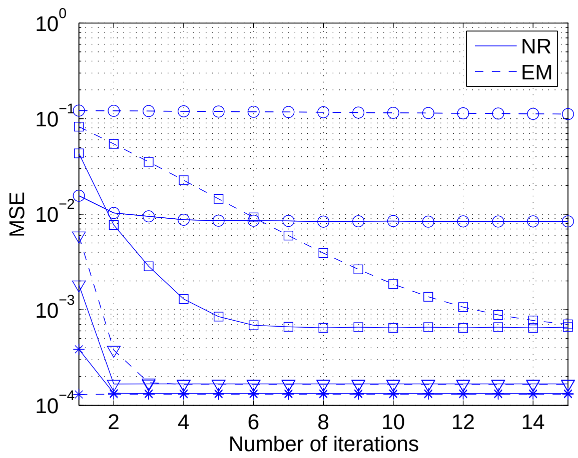

Fig. 1. Phase-estimation mean square error for $I(c_{k,l}, L_{ext}(c_{k,l}))$ equal to 0 (circle), 0.2 (square), 0.8 (triangle) and 1 (star).

7. CONCLUSION

We propose an efficient implementation of the NR method in the context of synchronization. On the one hand we show that the proposed NR implementation has, up to an approximation, a complexity *order* per iteration similar to the EM algorithm. On the other hand, we illustrate that the NR method converges in some cases much faster than the EM algorithm. As a conclusion, we emphasize that, thanks to the proposed implemention, the NR method may be an attractive candidate for iterative resolution of ML problems.

8. APPENDIX

In this appendix, we show that equality (8) holds. From (7) we have

$$\frac{\partial \ln p(\mathbf{r}|b)}{\partial b} = \mathbb{E}_{\mathbf{a}|\mathbf{r},b}\left[\frac{\partial \ln p(\mathbf{r}|\mathbf{a},b)}{\partial b}\right], = \sum_{\mathbf{a}} p(\mathbf{a}|\mathbf{r},b)\,\frac{\partial \ln p(\mathbf{r}|\mathbf{a},b)}{\partial b}. \quad (16)$$

Applying the derivative operator on (16), we get

$$\frac{\partial^2 \ln p(\mathbf{r}|\mathbf{b})}{\partial b^2} = \sum_{\mathbf{a}} p(\mathbf{a}|\mathbf{r},b)\,\frac{\partial^2 \ln p(\mathbf{r}|\mathbf{a},b)}{\partial b^2} + \sum_{\mathbf{a}} \frac{\partial p(\mathbf{a}|\mathbf{r},b)}{\partial b}\,\frac{\partial \ln p(\mathbf{r}|\mathbf{a},b)}{\partial b} \quad (17)$$

Note that $\frac{\partial p(\mathbf{a}|\mathbf{r},b)}{\partial b}$ may also be expressed as

$$\frac{\partial p(\mathbf{a}|\mathbf{r},b)}{\partial b} = p(\mathbf{a}|\mathbf{r},b)\,\frac{\partial \ln p(\mathbf{a}|\mathbf{r},b)}{\partial b}.$$

Taking then into account that $p(\mathbf{a}|\mathbf{r},b) = \frac{p(\mathbf{r}|\mathbf{a},b)p(\mathbf{a})}{p(\mathbf{r}|b)}$, we have

$$\frac{\partial p(\mathbf{a}|\mathbf{r},b)}{\partial b} = p(\mathbf{a}|\mathbf{r},b)\left(\frac{\partial \ln p(\mathbf{r}|\mathbf{a},b)}{\partial b} - \frac{\partial \ln p(\mathbf{r}|b)}{\partial b}\right). \quad (18)$$

Plugging (18) into (17), we get

$$\frac{\partial^2 \ln p(\mathbf{r}|b)}{\partial b^2} = \sum_{\mathbf{a}} p(\mathbf{a}|\mathbf{r},b)\,\frac{\partial^2 \ln p(\mathbf{r}|\mathbf{a},b)}{\partial b^2} + \sum_{\mathbf{a}} p(\mathbf{a}|\mathbf{r},b)\left(\frac{\partial \ln p(\mathbf{r}|\mathbf{a},b)}{\partial b}\right)^2 - \left(\sum_{\mathbf{a}} p(\mathbf{a}|\mathbf{r},b)\,\frac{\partial \ln p(\mathbf{r}|\mathbf{a},b)}{\partial b}\right)^2, \quad (19)$$

where the last term in (19) follows from (16).

9. REFERENCES

[1] N. Noels, C. Herzet, A. Dejonghe, V. Lottici, H. Steendam, M. Moeneclaey, M Luise and L. Vandendorpe, "Turbo-synchronization: an EM algorithm approach," in *Proc. IEEE ICC*, Anchorage, May 2003.

[2] C. Herzet, V. Ramon, L. Vandendorpe and M. Moeneclaey, "EM algorithm-based timing synchronization in turbo receivers," in *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP'*, Hong Kong, April 2003.

[3] J. Dauwels and H.-A. Loeliger, "Phase Estimation by Message Passing," in *IEEE International Conference on Communications, ICC'*, Paris, France, June 2004.

[4] C. Herzet, V. Ramon, and L. Vandendorpe, "Turbo-synchronization: a combined sum-product and expectation-maximization algorithm approach," New-York, 2005, IEEE Workshop on Signal Processing Advances in Wireless Communications, SPAWC'.

[5] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum-likelihood from incomplete data via the em algorithm," *J. Roy. Stat. Soc.*, vol. 39, no. 1, pp. pp. 1–38, Jan. 1977.

[6] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. on Inform. Theory*, vol. 47, pp. pp. 498–519, Feb. 2001.

[7] D.P. Bertsekas, *Nonlinear Programming*, Athena Scientific, USA, 2003.

[8] N. Noels, H. Steendam and M. Moeneclaey, ""Carrier and Clock recovery in (turbo) coded systems: Cramer-Rao Bound and Synchronizer Performance"," accepted for publication in *EURASIP Journal on Applied Signal processing*, available at http://telin.rug.ac.be/ nnoels/articles.html.