EM-Based Recursive Estimation of Channel Parameters

Hossein Zamiri-Jafarian, *Member, IEEE,* and Subbarayan Pasupathy, *Fellow, IEEE*

***Abstract*— Recursive (online) expectation–maximization (EM) algorithm along with stochastic approximation is employed in this paper to estimate unknown time-invariant/variant parameters. The impulse response of a linear system (channel) is modeled as an unknown deterministic vector/process and as a Gaussian vector/process with unknown stochastic characteristics. Using these models which are embedded in white or colored Gaussian noise, different types of recursive least squares (RLS), Kalman filtering and smoothing and combined RLS and Kalman-type algorithms are derived directly from the recursive EM algorithm. The estimation of unknown parameters also generates new recursive algorithms for situations, such as additive colored noise modeled by an autoregressive process. The recursive EM algorithm is shown as a powerful tool which unifies the derivations of many adaptive estimation methods.**

***Index Terms*— Adaptive estimation, colored noise, estimation and maximization algorithm, estimation theory, Kalman filtering, maximum-likelihood estimation, recursive estimation.**

I. Introduction

MAXIMUM-LIKELIHOOD (ML) criterion serves as a benchmark in estimation when the unknown parameters are deterministic. However in many cases the received data does not provide complete information necessary for such maximization. The expectation–maximization (EM) algorithm [1], [2] provides an iterative solution in such situations.

Applications of the EM algorithm to parameter estimation are considered in [3]–[9]. In particular, the online estimation of parameters based on the Kullback–Leibler information measure and using stochastic approximation is considered in [7] and [8]. Several applications of the EM algorithm to receiver design are also presented in [10]–[12]. Other recursive algorithms such as recursive ML and prediction error methods are presented in [13]. EM algorithm is a batch-oriented approach which processes the entire received data. In order to eliminate the delay in decision-making, reduce storage and increase the computational efficiency in real-time applications, it is desirable and often necessary to process the received data in a recursive manner.

The recursive (online) EM algorithm developed in this paper extends and modifies the algorithm in [7] by using iteration in each recursion and by considering time-variant unknown parameters. The proposed algorithm leads, for special cases, to some new RLS/Kalman-type algorithms for colored Gaussian noise. Although achieving ML estimation is not always guaranteed and we do not provide a proof of the convergence of the recursive EM algorithm, the recursive estimator increases the likelihood monotonically.

Paper approved by E. Eleftheriou, the Editor for Equalizationa and Coding of the IEEE Communications Society. Manuscript received December 18, 1997; revised November 17, 1998 and March 1, 1999.

The authors are with the Department of Electrical and Computer Engineering, University of Toronto, Toronto, Ont., Canada M5S 3G4.

Publisher Item Identifier S 0090-6778(99)07451-6.

II. ML Estimation Via the Recursive EM

Let us consider $\boldsymbol{\theta}_k$ as a column vector of deterministic channel parameters up to time k to be estimated from the data vector observed up to time $k, \mathbf{y}_k = [y(k), \cdots, y(0)]^{\mathcal{T}}$, where $X^{\mathcal{T}}$ denotes the transpose of X. ML estimation of $\boldsymbol{\theta}_k$ is given by

$$\begin{aligned}\hat{\boldsymbol{\theta}}_{k|k} &= \arg \max_{\boldsymbol{\theta}_k} \{p(\mathbf{y}_k|\boldsymbol{\theta}_k)\} \\ &\equiv \arg \max_{\boldsymbol{\theta}_k} \{\log\, p(\mathbf{y}_k|\boldsymbol{\theta}_k)\} \end{aligned} \tag{1}$$

where $\hat{\boldsymbol{\theta}}_{k|l}$ is ML estimation of $\boldsymbol{\theta}_k$ based on received signal $\mathbf{y}_l$. When $\mathbf{y}_k$ is incomplete data, the maximization of $\boldsymbol{L}_k(\boldsymbol{\theta}_k) = \log\, p(\mathbf{y}_k|\boldsymbol{\theta}_k)$ is not tractable. Denoting $\mathbf{y}_k$ as incomplete data $\mathcal{I}_k = \mathbf{y}_k$ and $\mathcal{D}_k$ as the desired additional information needed at time k to complete $\mathcal{I}_k$ and following the regular (offline) EM algorithm [1], the two steps of the recursive EM algorithm at time k are as follows.

1-E step:

$$Q_k\left(\boldsymbol{\theta}_k|\tilde{\boldsymbol{\theta}}_{k|k}^{(m-1)}\right) = E\left[\log\, p(\mathcal{C}_k|\boldsymbol{\theta}_k)|\mathcal{I}_k, \tilde{\boldsymbol{\theta}}_{k|k}^{(m-1)}\right] \tag{2}$$

2-M step:

$$\tilde{\boldsymbol{\theta}}_{k|k}^{(m)} = \arg \max_{\boldsymbol{\theta}_k} \left\{Q_k\left(\boldsymbol{\theta}_k|\tilde{\boldsymbol{\theta}}_{k|k}^{(m-1)}\right)\right\} \tag{3}$$

where $\mathcal{C}_k = \{\mathcal{I}_k, \mathcal{D}_k\}$ is the complete data at time k and $\tilde{\boldsymbol{\theta}}_{k|l}^{(m)}$ is the estimation of $\boldsymbol{\theta}_k$ at the mth iteration based on the signal received up to time $l, \mathbf{y}_l$, for $m > 0$. When $m = 0$, $\tilde{\boldsymbol{\theta}}_{k|l}^{(m)}$ is the initial value of estimate $\boldsymbol{\theta}_k$ based on the received signal $\mathbf{y}_{l-1}$. The steps of the algorithm at time k are repeated until at $\acute{m}$th iteration $\tilde{\boldsymbol{\theta}}_{k|k}^{(\acute{m})} = \tilde{\boldsymbol{\theta}}_{k|k}^{(\acute{m}-1)}$. When $\boldsymbol{\varphi}_l$ is the unknown parameter vector just for time l, where $y(l)$ is independent of other parameters by knowing $\boldsymbol{\varphi}_l$, we have $\boldsymbol{\theta}_k = [\boldsymbol{\varphi}_k^{\mathcal{T}}\, \boldsymbol{\varphi}_{k-1}^{\mathcal{T}}, \cdots, \boldsymbol{\varphi}_0^{\mathcal{T}}]^{\mathcal{T}}$ and time-update vector $\tilde{\boldsymbol{\theta}}_{k+1|k+1}^{(0)}$ for the next recursion of the procedures (2) and (3) is given by

$$\tilde{\boldsymbol{\theta}}_{k+1|k} \doteq \tilde{\boldsymbol{\theta}}_{k+1|k+1}^{(0)} = \begin{bmatrix} \tilde{\boldsymbol{\varphi}}_{k+1|k+1}^{(0)} \\ \tilde{\boldsymbol{\theta}}_{k|k}^{(\acute{m})} \end{bmatrix} \tag{4}$$

where $\tilde{\boldsymbol{\varphi}}_{k+1|k} \doteq \tilde{\boldsymbol{\varphi}}_{k+1|k+1}^{(0)}$ and $\tilde{\boldsymbol{\theta}}_{k+1|k} \doteq \tilde{\boldsymbol{\theta}}_{k+1|k+1}^{(0)}$ are the estimates of $\boldsymbol{\varphi}_{k+1}$ and $\boldsymbol{\theta}_{k+1}$, respectively, based on the entire received signal up to time k. In general, $\tilde{\boldsymbol{\varphi}}_{k+1|k}$ can be a function of $\tilde{\boldsymbol{\theta}}_{k|k}^{(\acute{m})}$ and is obtained by using the dynamic evolution of the $\boldsymbol{\varphi}_{k+1}$ process. From (4) it is clear that $\tilde{\boldsymbol{\theta}}_{k|k}^{(\acute{m})}$ is used as an initial value for next recursion or, in other words, $\tilde{\boldsymbol{\theta}}_{k|k+1}^{(0)} = \tilde{\boldsymbol{\theta}}_{k|k}^{(\acute{m})}$.

Since for each iteration $Q_k(\tilde{\boldsymbol{\theta}}_{k|k}^{(m)}|\tilde{\boldsymbol{\theta}}_{k|k}^{(m-1)}) \geq Q_k(\tilde{\boldsymbol{\theta}}_{k|k}^{(m-1)}|\tilde{\boldsymbol{\theta}}_{k|k}^{(m-1)})$, one can show that [1]

$$\boldsymbol{L}_k\left(\tilde{\boldsymbol{\theta}}_{k|k}^{(m)}\right) \geq \boldsymbol{L}_k\left(\tilde{\boldsymbol{\theta}}_{k|k}^{(m-1)}\right). \quad (5)$$

Meanwhile, from (4) at time $k+1$, we have $\boldsymbol{L}_{k+1}(\tilde{\boldsymbol{\theta}}_{k+1|k+1}^{(0)}) = \boldsymbol{L}_k(\tilde{\boldsymbol{\theta}}_{k|k}^{(\acute{m})}) + L_{k+1}(\tilde{\boldsymbol{\varphi}}_{k+1|k+1}^{(0)})$, where $L_{k+1}(\tilde{\boldsymbol{\varphi}}_{k+1|k+1}^{(0)}) = \log p(y(k+1)|\ \tilde{\boldsymbol{\varphi}}_{k+1}^{(0)}, \mathbf{y}_k)$. From (5), $\boldsymbol{L}_{k+1}(\tilde{\boldsymbol{\theta}}_{k+1|k+1}^{(m)}) \geq \boldsymbol{L}_{k+1}(\tilde{\boldsymbol{\theta}}_{k+1|k+1}^{(0)})$, thus we have

$$\boldsymbol{L}_{k+1}(\tilde{\boldsymbol{\theta}}_{k+1|k+1}^{(m)}) \geq \boldsymbol{L}_k(\tilde{\boldsymbol{\theta}}_{k|k}^{(\acute{m})}) + L_{k+1}(\tilde{\boldsymbol{\varphi}}_{k+1|k+1}^{(0)}). \quad (6)$$

Therefore, as (5) and (6) show, the log-likelihood function is increased monotonically both in each iteration and recursion. The estimation procedure based on the recursive EM algorithm is more attractive when its maximization step can be done analytically in a recursive manner.

Following Titterington's approach of stochastic approximation of $Q_k(\cdot|\cdot)$ based on three elements of its Taylor series [14], one can show that

$$\tilde{\boldsymbol{\theta}}_{k|k}^{(m)} \simeq \tilde{\boldsymbol{\theta}}_{k|k}^{(m-1)} - \left(\left.\frac{\partial^2 Q_k(\boldsymbol{\theta}_k|\tilde{\boldsymbol{\theta}}_{k|k}^{(m-1)})}{\partial^2\boldsymbol{\theta}_k}\right|_{\boldsymbol{\theta}_k=\tilde{\boldsymbol{\theta}}_{k|k}^{(m-1)}}\right)^{-1} \cdot \left(\left.\frac{\partial Q_k(\boldsymbol{\theta}_k|\tilde{\boldsymbol{\theta}}_{k|k}^{(m-1)})}{\partial\boldsymbol{\theta}_k^*}\right|_{\boldsymbol{\theta}_k=\tilde{\boldsymbol{\theta}}_{k|k}^{(m-1)}}\right) \quad (7)$$

where $(\partial^2 Q_k(\cdot|\cdot)/\partial^2\boldsymbol{\theta}_k) \doteq (\partial^2 Q_k(\cdot|\cdot)/\partial\boldsymbol{\theta}_k^*\partial\boldsymbol{\theta}_k^{\mathcal{T}})$, $\boldsymbol{\theta}_k^*$ denotes complex conjugate of $\boldsymbol{\theta}_k$. When the third and higher derivatives of $Q_k(\cdot|\cdot)$ are zero, as is usually true for the Gaussian noise case in a linear system, the recursive formula (7) is exact.

In the following sections, we assume that only one iteration is used at each recursion. This idea is similar to the generalized EM algorithm [1] which aims at just increasing the value of $Q_k(\cdot|\cdot)$ instead of trying to obtain its maximum. However, when only one iteration can achieve the maximum of $Q_k(\cdot|\cdot)$ at time k (a situation which can be true for some cases), the method achieves ML estimation or a local maximum point when the likelihood function has many local maxima. In order to avoid complicated notations, in the following sections we use $\tilde{\boldsymbol{\theta}}_{k|k-1}$ and $\tilde{\boldsymbol{\theta}}_{k|k}$ instead of $\tilde{\boldsymbol{\theta}}_{k|k}^{(0)}$ and $\tilde{\boldsymbol{\theta}}_{k|k}^{(1)}$, respectively, in applying (2), (3), and (7).

III. Channel Estimation

Data transmission through a linear noisy channel can be described as $y(k) = \Sigma_{l=0}^{L}\ s(k-l)h(l,k) + z(k) = \boldsymbol{s}(k)\boldsymbol{h}(k) + z(k)$, where $s(k)$ is the transmitted signal, $h(l,k)$ is the impulse response of the linear channel,[1] $z(k)$ is additive noise generally modeled as a complex, circularly symmetric, white/colored Gaussian random process, $\boldsymbol{s}(k) = [s(k), \cdots, s(k-L)]$, $\boldsymbol{h}(k) = [h(0,k), \cdots, h(L,k)]^{\mathcal{T}}$ and $y(k)$ is the received signal. To detect the transmitted data, the receiver needs to know channel parameters such as the channel-impulse response (CIR) for deterministic channels or the stochastic characteristics of the CIR for stochastic channels. The communication channel usually suffering from intersymbol interference (ISI) and multipath fading is modeled as a discrete finite memory system whose impulse response length is limited to $L+1$, $h(l,k) = 0$ for $0 > l > L$.

In this section, we focus on estimating the channel parameters ($h(l,k)$ or its statistical parameters) for different models of CIR based on the recursive EM algorithm. We assume that the receiver knows $s(k)$. The knowledge about $s(k)$ can be achieved in the detection algorithm based on the estimation of $s(k)$ in a decision feedback equalization method or by using the different hypotheses of $s(k)$ in a maximum-likelihood sequence detection (MLSD) method or when the communication system is in the training mode [15].

A. Unknown Deterministic CIR

In this model CIR is considered as an unknown vector of deterministic parameters which may be time-invariant or time-variant. We assume the linear dynamic change in time-variant CIR as $\mathbf{h}(k) = \boldsymbol{F}(k)\mathbf{h}(k-1)$, where $\mathbf{h}_k = \mathbf{h}(k) = [\boldsymbol{h}(k)^{\mathcal{T}}, \cdots, \boldsymbol{h}(k-N+1)^{\mathcal{T}}]^{\mathcal{T}}$ and $\boldsymbol{F}(k)$ is an $N(L+1) \times N(L+1)$ matrix. $z(k)$ is a stationary colored Gaussian noise modeled by an $(M-1)$th order autoregressive process with positive definite covariance matrix $\Sigma_z = \operatorname{cov}(\boldsymbol{z}_k) = E[(\boldsymbol{z}_k - \boldsymbol{\mu}_z)(\boldsymbol{z}_k - \boldsymbol{\mu}_z)^{\mathcal{H}}]$, where $\boldsymbol{z}_k = [z(k), \cdots, z(k-M+1)]^{\mathcal{T}}$, $\boldsymbol{\mu}_z = E[\boldsymbol{z}_k]$. From the above we have

$$\begin{aligned}\boldsymbol{y}_k &= \begin{bmatrix} y(k) \\ y(k-1) \\ \vdots \\ y(k-M+1) \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{s}(k) & & & \\ & \mathbf{s}(k-1) & & 0 \\ & 0 & \ddots & \\ & & & \mathbf{s}(k-M+1) \end{bmatrix} \begin{bmatrix} \mathbf{h}_k \\ \mathbf{h}_{k-1} \\ \vdots \\ \mathbf{h}_{k-M+1} \end{bmatrix} + \boldsymbol{z}_k \\ &= S(k)\mathbf{h}_k^{\diamond} + \boldsymbol{z}_k \end{aligned} \quad (8)$$

where $\mathbf{s}(k) = [\boldsymbol{s}(k), \mathbf{0}]$ and $\mathbf{0}$ is an $(N-1)(L+1)$ zero row vector. The unknown parameter vector up to time k is $\overline{\mathbf{h}}_k = [\mathbf{h}_k^{\mathcal{T}}, \cdots, \mathbf{h}_0^{\mathcal{T}}]^{\mathcal{T}}$ in this model. Also, the complete and incomplete data are defined as $\mathcal{C}_k = \mathcal{I}_k = \mathbf{y}_k = [y(k), \cdots, y(0)]^{\mathcal{T}}$. Thus, $Q_k(\overline{\mathbf{h}}_k|\tilde{\overline{\mathbf{h}}}_{k|k-1})$ is given as (9) at the bottom of the next page, where $\acute{\boldsymbol{z}}_k = [z(k-1), \cdots, z(k-M+1)]^{\mathcal{T}}$ and $\Sigma_{\acute{z}} = \operatorname{cov}(\acute{\boldsymbol{z}}_k)$. Σ_z^{-1} and $\Sigma_{\acute{z}}^{-1}$ are symmetric positive definite matrices, and by using Cholesky decomposition they can be factorized into a product of two triangular matrices which are complex-conjugate transposes of each other. After some manipulations we get

$$\begin{aligned} Q_k(\overline{\mathbf{h}}_k|\tilde{\overline{\mathbf{h}}}_{k|k-1}) &= Q_{k-1}(\overline{\mathbf{h}}_{k-1}|\tilde{\overline{\mathbf{h}}}_{k-1|k-2}) - \{\log(\pi) - \log(\gamma_0\gamma_0^*) \\ &\quad + (\boldsymbol{y}_k - S(k)\mathbf{h}_k^{\diamond} - \boldsymbol{\mu}_z)^{\mathcal{H}}\boldsymbol{\gamma}\boldsymbol{\gamma}^{\mathcal{H}}(\boldsymbol{y}_k - S(k)\mathbf{h}_k^{\diamond} - \boldsymbol{\mu}_z)\} \end{aligned} \quad (10)$$

where $\boldsymbol{\gamma} = [\gamma_0, \gamma_1, \cdots, \gamma_{M-1}]^{\mathcal{T}}$ is the first column of the

[1] $h(l,k)$ is the time-variant channel response at time k due to an impulse applied at time $k-l$.

lower triangular matrix of the decomposition of Σ_z^{-1}. Due to the maximization step at time $k-1$, the first derivative of $Q_k(\overline{\mathbf{h}}_k|\tilde{\overline{\mathbf{h}}}_{k|k-1})$ with respect to $\mathbf{h}_l^{\diamond^*}$ at point $\overline{\mathbf{h}}_k = \tilde{\overline{\mathbf{h}}}_{k|k-1}$ is given in (11), shown at the bottom of the page.

Since the first derivative of $Q_k(\cdot|\cdot)$ with respect to $\mathbf{h}_l^{\diamond^*}$ is zero for $l \neq k$, only the estimate $\mathbf{h}_k^{\diamond}$ is needed at time k. The second derivative of $Q_k(\cdot|\cdot)$ at point $\overline{\mathbf{h}}_k = \tilde{\overline{\mathbf{h}}}_{k|k-1}$ is shown in (12) at the bottom of the page, where $\mathbf{F}_k$ is a diagonal matrix and $\{\boldsymbol{F}(k), \boldsymbol{F}(k-1), \cdots, \boldsymbol{F}(k-M+1)\}$ are its diagonal elements. By defining $\mathbf{P}_{k|k} = (-(\partial^2 Q_k(\overline{\mathbf{h}}_k|\tilde{\overline{\mathbf{h}}}_{k|k-1})/\partial^2\mathbf{h}_k^{\diamond})|_{\overline{\mathbf{h}}_k=\tilde{\overline{\mathbf{h}}}_{k|k-1}})^{-1}$ and using $(A+BC)^{-1} = A^{-1} - A^{-1}B(I+CA^{-1}B)^{-1}CA^{-1}$, we have

$$\mathbf{P}_{k|k} = \mathbf{P}_{k|k-1} - \mathbf{P}_{k|k-1}S(k)^{\mathcal{H}}\boldsymbol{\gamma}(1+\boldsymbol{\gamma}^{\mathcal{H}}S(k) \cdot\mathbf{P}_{k|k-1}S(k)^{\mathcal{H}}\boldsymbol{\gamma})^{-1}\boldsymbol{\gamma}^{\mathcal{H}}S(k)\mathbf{P}_{k|k-1} \tag{13}$$

where from (12) $\mathbf{P}_{k|k-1} = \mathbf{F}_k\mathbf{P}_{k-1|k-1}\mathbf{F}_k^{\mathcal{H}}$. Therefore from (7) the recursive estimation of $\mathbf{h}_k^{\diamond}$ at time k becomes

$$\tilde{\mathbf{h}}_{k|k}^{\diamond} = \tilde{\mathbf{h}}_{k|k-1}^{\diamond} + \mathbf{P}_{k|k-1}S(k)^{\mathcal{H}}\boldsymbol{\gamma}(1+\boldsymbol{\gamma}^{\mathcal{H}}S(k) \cdot\mathbf{P}_{k|k-1}S(k)^{\mathcal{H}}\boldsymbol{\gamma})^{-1}\boldsymbol{\gamma}^{\mathcal{H}}(\boldsymbol{y}_k - S(k)\tilde{\mathbf{h}}_{k|k-1}^{\diamond} - \boldsymbol{\mu}_z) \tag{14}$$

where $\tilde{\mathbf{h}}_{k|k-1}^{\diamond} = \mathbf{F}_k\tilde{\mathbf{h}}_{k-1|k-1}^{\diamond}$. As can be seen the recursive formula (14) is similar to RLS/Kalman-type algorithm. When $z(k)$ is colored noise, the recursive formula in RLS algorithm needs to invert a nondiagonal matrix and the unknown parameters may not be estimated sequentially in time [16, p. 248]. In developing the recursive formula (14), however, the inverse of a matrix is not necessary. This result can be interpreted as using the whitening filter along with a RLS/Kalman algorithm where $\boldsymbol{\gamma}^{\mathcal{H}}$ is the coefficients of the whitening filter. When $z(k)$ is a zero mean white Gaussian noise with variance N_0, we have $\boldsymbol{\gamma} = [(N_0)^{-(1/2)}, \overbrace{0, \cdots, 0}^{M-1}]^{\mathcal{T}}$ and it is easy to show that

$$\begin{aligned}\tilde{\mathbf{h}}_{k|k} &= \tilde{\mathbf{h}}_{k|k-1} + P_{k|k-1}\mathbf{s}(k)^{\mathcal{H}}(1+\mathbf{s}(k)(k)P_{k|k-1}\mathbf{s}(k)^{\mathcal{H}})^{-1} \\ &\quad \cdot (y(k) - \mathbf{s}(k)\tilde{\mathbf{h}}_{k|k-1}) \\ P_{k|k} &= P_{k|k-1} - P_{k|k-1}\mathbf{s}(k)^{\mathcal{H}}(1+\mathbf{s}(k)P_{k|k-1}\mathbf{s}(k)^{\mathcal{H}})^{-1} \\ &\quad \cdot \mathbf{s}(k)P_{k|k-1}\end{aligned} \tag{15}$$

where $P_{k|k} = (-N_0(\partial^2 Q_k(\overline{\mathbf{h}}_k|\tilde{\overline{\mathbf{h}}}_{k|k-1})/\partial^2\mathbf{h}_k)|_{\overline{\mathbf{h}}_k=\tilde{\overline{\mathbf{h}}}_{k|k-1}})^{-1}$ and $P_{k|k-1} = \boldsymbol{F}(k)P_{k-1|k-1}\boldsymbol{F}(k)^{\mathcal{H}}$. Meanwhile, selecting $\boldsymbol{F}(k) = \lambda^{-(1/2)}I$, where $0 < \lambda \leq 1$ and I is an $N(L+1) \times N(L+1)$ identity matrix, and defining unknown parameters $\boldsymbol{\varphi} = \boldsymbol{h} = [h(0), \cdots, h(L)]^{\mathcal{T}}$, the time-variant model leads to a modified RLS algorithm with a forgetting factor λ. When $\boldsymbol{F}(k) = I$, the time-variant impulse response becomes time-invariant and the estimation of $\boldsymbol{h}$ leads to the well-known RLS algorithm. Meanwhile, due to Gaussian assumption for $z(k)$ the recursive estimating formula is exact. In addition, if $\mathbf{h}(k) = \boldsymbol{F}(k)\mathbf{h}(k-1) + G(k)W(k)$ where $W(k)$ is a zero-mean random process, $\mathbf{h}(k)$ becomes a random process (see Section III-B-2). However, by assuming a deterministic model for $\mathbf{h}(k)$, all of the obtained results in this section remain the same.

$$\begin{aligned}Q_k(\overline{\mathbf{h}}_k|\tilde{\overline{\mathbf{h}}}_{k|k-1}) &= E[\log\, p(\mathbf{y}_k|\overline{\mathbf{h}}_k)|\mathbf{y}_k, \tilde{\overline{\mathbf{h}}}_{k|k-1}] \\ &= \log\, p(\mathbf{y}_{k-1}|\overline{\mathbf{h}}_{k-1}) + \log\, p(y(k)|\overline{\mathbf{h}}_k, \mathbf{y}_{k-1}) \\ &= Q_{k-1}(\overline{\mathbf{h}}_{k-1}|\tilde{\overline{\mathbf{h}}}_{k-1|k-2}) + \log\left\{\frac{p(\boldsymbol{y}_k|\mathbf{h}_k^{\diamond})}{p(y(k-1), \cdots, y(k-M+1)|\mathbf{h}_k^{\diamond})}\right\} \\ &= Q_{k-1}(\overline{\mathbf{h}}_{k-1}|\tilde{\overline{\mathbf{h}}}_{k-1|k-2}) - \Bigg\{\log(\pi) + \log(\det(\Sigma_z)) - \log(\det(\Sigma_{\acute{z}})) + (\boldsymbol{y}_k - S(k)\mathbf{h}_k^{\diamond} - \boldsymbol{\mu}_z)^{\mathcal{H}} \\ &\qquad \cdot \left(\Sigma_z^{-1} - \begin{bmatrix} 0 & \vdots & \mathbf{0} \\ \cdots & \cdots & \cdots \\ \mathbf{0} & \vdots & \Sigma_{\acute{z}}^{-1}\end{bmatrix}\right) \cdot (\boldsymbol{y}_k - S(k)\mathbf{h}_k^{\diamond} - \boldsymbol{\mu}_z)\Bigg\}\end{aligned} \tag{9}$$

$$\left.\frac{\partial Q_k(\overline{\mathbf{h}}_k|\tilde{\overline{\mathbf{h}}}_{k|k-1})}{\partial \mathbf{h}_l^{\diamond^*}}\right|_{\overline{\mathbf{h}}_k=\tilde{\overline{\mathbf{h}}}_{k|k-1}} = \begin{cases} \left.\dfrac{\partial Q_{k-1}(\overline{\mathbf{h}}_{k-1}|\tilde{\overline{\mathbf{h}}}_{k-1|k-2})}{\partial \mathbf{h}_l^{\diamond^*}}\right|_{\overline{\mathbf{h}}_{k-1}=\tilde{\overline{\mathbf{h}}}_{k-1|k-1}} = 0, & 0 \leq l \leq k-1 \\ S(k)^{\mathcal{H}}\boldsymbol{\gamma\gamma}^{\mathcal{H}}(\boldsymbol{y}_k - S(k)\tilde{\mathbf{h}}_{k|k-1}^{\diamond} - \boldsymbol{\mu}_z), & l = k \end{cases} \tag{11}$$

$$\begin{aligned}\left.\frac{\partial^2 Q_k(\overline{\mathbf{h}}_k|\tilde{\overline{\mathbf{h}}}_{k|k-1})}{\partial^2\mathbf{h}_k^{\diamond}}\right|_{\overline{\mathbf{h}}_k=\tilde{\overline{\mathbf{h}}}_{k|k-1}} &= \left(\frac{\partial \mathbf{h}_{k-1}^{\diamond}}{\partial \mathbf{h}_k^{\diamond}}\right)^* \left(\left.\frac{\partial^2 Q_{k-1}(\overline{\mathbf{h}}_{k-1}|\tilde{\overline{\mathbf{h}}}_{k-1|k-2})}{\partial^2\mathbf{h}_{k-1}^{\diamond}}\right|_{\overline{\mathbf{h}}_{k-1}=\tilde{\overline{\mathbf{h}}}_{k-1|k-1}}\right)\left(\frac{\partial \mathbf{h}_{k-1}^{\diamond}}{\partial \mathbf{h}_k^{\diamond}}\right)^{\mathcal{T}} - S(k)^{\mathcal{H}}\boldsymbol{\gamma\gamma}^{\mathcal{H}}S(k) \\ &= \mathbf{F}_k^{-\mathcal{H}}\left(\left.\frac{\partial^2 Q_{k-1}(\overline{\mathbf{h}}_{k-1}|\tilde{\overline{\mathbf{h}}}_{k-1|k-2})}{\partial^2\mathbf{h}_{k-1}^{\diamond}}\right|_{\overline{\mathbf{h}}_{k-1}=\tilde{\overline{\mathbf{h}}}_{k-1|k-1}}\right)\mathbf{F}_k^{-1} - S(k)^{\mathcal{H}}\boldsymbol{\gamma\gamma}^{\mathcal{H}}S(k)\end{aligned} \tag{12}$$

B. Gaussian Random CIR

The channel in a mobile communication system is generally modeled as a linear system whose impulse response is a random vector or random process. One such common model is the Rayleigh multipath fading channel. In this model, the CIR is a complex Gaussian random vector/process whose amplitude is Rayleigh distributed. Also, without loss of generality we assume that $z(k)$ is a zero-mean white Gaussian noise with autocorrelation $R_z(k) = N_0\delta(k)$; for colored noise one can follow the same procedure developed in Section III-A. In the following, we consider the estimation of the Gaussian CIR parameters using the recursive EM algorithm.

1) Gaussian Random Vector: The received signal is obtained from $y(k) = \boldsymbol{s}(k)\boldsymbol{h} + z(k)$, where $\boldsymbol{h}$ is the Gaussian random vector. The maximum *a posteriori* (MAP) estimation of the CIR at time k is $\hat{\boldsymbol{h}} = \arg\ \max_{\boldsymbol{h}}\{\log\ p(\boldsymbol{h}|\boldsymbol{y}_k)\} = E[\boldsymbol{h}|\boldsymbol{y}_k] = \boldsymbol{\mu}_{|k}$. Therefore, the unknown vector of deterministic parameters at time k is $\boldsymbol{\mu}_{|k}$, the conditional mean of $\boldsymbol{h}$. The complete and incomplete data are defined as $\mathcal{C}_k = \{\mathbf{y}_k, \boldsymbol{h}\}$ and $\mathcal{I}_k = \mathbf{y}_k$, respectively, at time k. Then (2) becomes

$$\begin{aligned} Q_k(\boldsymbol{\mu}_{|k}|\tilde{\boldsymbol{\mu}}_{|k-1}) &= E[\log\ p(\mathbf{y}_k, \boldsymbol{h}|\boldsymbol{\mu}_{|k})|\mathbf{y}_k, \tilde{\boldsymbol{\mu}}_{|k-1}] \\ &= E[\log\ p(\mathbf{y}_{k-1}, \boldsymbol{h}|\boldsymbol{\mu}_{|k}) \\ &\quad + \log\ p(y(k)|\boldsymbol{\mu}_{|k}, \mathbf{y}_{k-1}, \boldsymbol{h})|\mathbf{y}_k, \tilde{\boldsymbol{\mu}}_{|k-1}] \\ &= Q_{k-1}(\boldsymbol{\mu}_{|k}|\tilde{\boldsymbol{\mu}}_{|k-1}) - \{\log(\pi N_0) + E[(y(k) \\ &\quad - \boldsymbol{s}(k)\boldsymbol{h})^{\mathcal{H}} N_0^{-1}(y(k) - \boldsymbol{s}(k)\boldsymbol{h})|\mathbf{y}_k, \tilde{\boldsymbol{\mu}}_{|k-1}]\}. \end{aligned} \tag{16}$$

From the definition of $Q_{k-1}(\boldsymbol{\mu}_{|k}|\tilde{\boldsymbol{\mu}}_{|k-1}) = E[\log\ p(\mathbf{y}_{k-1}, \boldsymbol{h}|\boldsymbol{\mu}_{|k})|\mathbf{y}_k, \tilde{\boldsymbol{\mu}}_{|k-1}]$, we have

$$\begin{aligned} \left.\frac{\partial Q_{k-1}(\boldsymbol{\mu}_{|k}|\tilde{\boldsymbol{\mu}}_{|k-1})}{\partial \boldsymbol{\mu}^*_{|k}}\right|_{\boldsymbol{\mu}_{|k}=\tilde{\boldsymbol{\mu}}_{|k-1}} &= \left.\frac{\partial Q_{k-1}(\boldsymbol{\mu}_{|k-1}|\tilde{\boldsymbol{\mu}}_{|k-2})}{\partial \boldsymbol{\mu}^*_{|k-1}}\right|_{\boldsymbol{\mu}_{|k-1}=\tilde{\boldsymbol{\mu}}_{|k-1}} \\ &= 0. \end{aligned} \tag{17}$$

Therefore, by replacing $E[\boldsymbol{h}^{\mathcal{H}} A\boldsymbol{h}|\mathbf{y}_k] = E[(\boldsymbol{h} - \boldsymbol{\mu}_{|k})^{\mathcal{H}} A(\boldsymbol{h} - \boldsymbol{\mu}_{|k})|\mathbf{y}_k] + \boldsymbol{\mu}_{|k}{}^{\mathcal{H}} A\boldsymbol{\mu}_{|k}$, the first derivative of $Q_k(\boldsymbol{\mu}_{|k}|\tilde{\boldsymbol{\mu}}_{|k-1})$ at point $\boldsymbol{\mu}_{|k} = \tilde{\boldsymbol{\mu}}_{|k-1}$ becomes

$$\left.\frac{\partial Q_k(\boldsymbol{\mu}_{|k}|\tilde{\boldsymbol{\mu}}_{|k-1})}{\partial \boldsymbol{\mu}^*_{|k}}\right|_{\boldsymbol{\mu}_{|k}=\tilde{\boldsymbol{\mu}}_{|k-1}} = \boldsymbol{s}(k)^{\mathcal{H}} N_0^{-1}(y(k) - \boldsymbol{s}(k)\tilde{\boldsymbol{\mu}}_{|k-1}). \tag{18}$$

Meanwhile $Q_k(\boldsymbol{\mu}_{|k}|\tilde{\boldsymbol{\mu}}_{|k-1})$ and $Q_{k-1}(\boldsymbol{\mu}_{|k}|\tilde{\boldsymbol{\mu}}_{|k-1})$ can be expanded as

$$\begin{aligned} Q_k(\boldsymbol{\mu}_{|k}|\tilde{\boldsymbol{\mu}}_{|k-1}) &= -\{\log(\pi)^{L+1} + \log(\det(\Sigma_{|k})) \\ &\quad + E[(\boldsymbol{h} - \boldsymbol{\mu}_{|k})^{\mathcal{H}}\Sigma_{|k}^{-1}(\boldsymbol{h} - \boldsymbol{\mu}_{|k})|\mathbf{y}_k, \tilde{\boldsymbol{\mu}}_{|k-1}]\} \\ &\quad + \log\ p(\mathbf{y}_k|\boldsymbol{\mu}_{|k}) \end{aligned} \tag{19}$$

$$\begin{aligned} Q_{k-1}(\mu_{|k}|\tilde{\boldsymbol{\mu}}_{|k-1}) &= -\{\log(\pi)^{L+1} + \log(\det(\Sigma_{|k-1})) \\ &\quad + E[(\boldsymbol{h} - \boldsymbol{\mu}_{|k-1})^{\mathcal{H}}\Sigma_{|k-1}^{-1}(\boldsymbol{h} - \boldsymbol{\mu}_{|k-1})|\mathbf{y}_k, \tilde{\boldsymbol{\mu}}_{|k-1}]\} \\ &\quad + \log\ p(\mathbf{y}_{k-1}|\boldsymbol{\mu}_{|k}) \end{aligned} \tag{20}$$

where $\Sigma_{|l} = \operatorname{cov}(\boldsymbol{h}|\mathbf{y}_l)$. From (16) and (20), the second derivative of $Q_k(\boldsymbol{\mu}_{|k}|\tilde{\boldsymbol{\mu}}_{|k-1})$ at point $\boldsymbol{\mu}_{|k} = \tilde{\boldsymbol{\mu}}_{|k-1}$ becomes

$$\left.\frac{\partial^2 Q_k(\boldsymbol{\mu}_{|k}|\tilde{\boldsymbol{\mu}}_{|k-1})}{\partial^2 \boldsymbol{\mu}_{|k}}\right|_{\boldsymbol{\mu}_{|k}=\tilde{\boldsymbol{\mu}}_{|k-1}} = -\Sigma_{|k-1}^{-1} - \boldsymbol{s}(k)^{\mathcal{H}} N_0^{-1}\boldsymbol{s}(k). \tag{21}$$

As shown in (21), the estimation of $\Sigma_{|k-1}$ is necessary to estimate $\boldsymbol{\mu}_{|k}$. By choosing the initial value of the covariance matrix as an estimate value, from (16), (19), and (20), it is easy to show that the estimation of $\Sigma_{|k}$ at time k is

$$\begin{aligned} \tilde{\Sigma}_{|k} &= (\tilde{\Sigma}_{|k-1}^{-1} + \boldsymbol{s}(k)^{\mathcal{H}} N_0^{-1}\boldsymbol{s}(k))^{-1} \\ &= \tilde{\Sigma}_{|k-1} - \tilde{\Sigma}_{|k-1}\boldsymbol{s}(k)^{\mathcal{H}}(N_0 + \boldsymbol{s}(k)\tilde{\Sigma}_{|k-1}\boldsymbol{s}(k)^{\mathcal{H}})^{-1} \\ &\quad \cdot \boldsymbol{s}(k)\tilde{\Sigma}_{|k-1}. \end{aligned} \tag{22}$$

Hence, the recursive formula for estimating $\boldsymbol{\mu}_{|k}$ from (7) becomes

$$\begin{aligned} \tilde{\boldsymbol{\mu}}_{|k} = \tilde{\boldsymbol{\mu}}_{|k-1} &+ \tilde{\Sigma}_{|k-1}\boldsymbol{s}(k)^{\mathcal{H}}(N_0 + \boldsymbol{s}(k)\tilde{\Sigma}_{|k-1}\boldsymbol{s}(k)^{\mathcal{H}})^{-1} \\ &\cdot (y(k) - \boldsymbol{s}(k)\tilde{\boldsymbol{\mu}}_{|k-1}). \end{aligned} \tag{23}$$

As seen in (22) and (23), the recursive relation is the same as the stochastic RLS algorithm.

2) Gaussian Random Process: It is very common to model the CIR as a Gaussian random process in the mobile communication system with relatively fast fading rate. The dynamic changing of the CIR can be represented by $\mathbf{h}(k) = \boldsymbol{F}(k)\mathbf{h}(k-1) + G(k)W(k)$, where $W(k) = [w_0(k), \cdots, w_L(k)]^{\mathcal{T}}$. Also, $F(k)$ and $G(k)$ are $N(L+1)\times N(L+1)$ and $N(L+1)\times(L+1)$ matrices defined by

$$\begin{aligned} \boldsymbol{F}(k) &= \begin{bmatrix} F_1(k) & F_2(k) & \cdots & F_N(k) \\ I & 0 & \cdots & 0 \\ & \ddots & 0 & \\ & 0 & I & 0 \end{bmatrix} \\ G(k) &= \begin{bmatrix} \mathbf{g}(k) \\ 0 \\ \vdots \\ 0 \end{bmatrix} \end{aligned} \tag{24}$$

where I is $(L+1)\times(L+1)$ identity matrix and 0 is $(L+1)\times(L+1)$ zero matrix in (24). Meanwhile, $W(k)$ is a zero-mean complex white Gaussian random vector whose $(L+1)\times(L+1)$ autocorrelation matrix is $R_W(k) = I\delta(k)$ and it is independent of $z(k)$. Similar to the Gaussian vector case, it can be shown that in MAP estimation of time-variant CIR, the conditional mean of $\overline{\mathbf{h}}_k = [\mathbf{h}_k^{\mathcal{T}}, \mathbf{h}_{k-1}^{\mathcal{T}}, \cdots, \mathbf{h}_0^{\mathcal{T}}]^{\mathcal{T}}$ at time k, $\overline{\boldsymbol{\mu}}_{k|k} = E[\overline{\mathbf{h}}_k|\mathbf{y}_k]$, is necessary and should be considered as the unknown parameter vector. Let us define $\boldsymbol{\mu}_{k|k}$ and $\boldsymbol{\Sigma}_{k|k} = \operatorname{cov}(\overline{\mathbf{h}}_k|\mathbf{y}_k) = E[(\overline{\mathbf{h}}_k - \overline{\boldsymbol{\mu}}_{k|k})(\overline{\mathbf{h}}_k - \overline{\boldsymbol{\mu}}_{k|k})^{\mathcal{H}}|\mathbf{y}_k]$ based on their elements

$$\overline{\boldsymbol{\mu}}_{k|k} = \begin{bmatrix} \boldsymbol{\mu}_{k|k} \\ \boldsymbol{\mu}_{k-1|k} \\ \vdots \\ \boldsymbol{\mu}_{0|k} \end{bmatrix}$$

and

$$\boldsymbol{\Sigma}_{k|k} = \begin{bmatrix} \Sigma_{k,k|k} & \Sigma_{k,k-1|k} & \cdots & \Sigma_{k,0|k} \\ \Sigma_{k-1,k|k} & \Sigma_{k-1,k-1|k} & \cdots & \Sigma_{k-1,0|k} \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma_{0,k|k} & \Sigma_{0,k-1|k} & \cdots & \Sigma_{0,0|k} \end{bmatrix} \tag{25}$$

where $\boldsymbol{\mu}_{j|k} = E[\mathbf{h}_j|\mathbf{y}_k]$ and $\Sigma_{i,j|k} = E[(\mathbf{h}_i - \boldsymbol{\mu}_{i|k})(\mathbf{h}_j - \mu_{j|k})^{\mathcal{H}}|\mathbf{y}_k]$. The complete and incomplete data at time k are defined as $\mathcal{C}_k = \{\mathbf{y}_k, \overline{\mathbf{h}}_k\}$ and $\mathcal{I}_k = \mathbf{y}_k$, respectively. From (2), the E-step becomes

$$\begin{aligned} Q_k(\overline{\boldsymbol{\mu}}_{k|k}|\tilde{\overline{\boldsymbol{\mu}}}_{k|k-1}) &= E[\log p(\mathbf{y}_k, \overline{\mathbf{h}}_k|\overline{\boldsymbol{\mu}}_{k|k})|\boldsymbol{y}_k, \tilde{\overline{\boldsymbol{\mu}}}_{k|k-1}] \\ &= E[\log p(y(k)|\overline{\mathbf{h}}_k, \mathbf{y}_{k-1}, \overline{\boldsymbol{\mu}}_{k|k}) \\ &\quad + \log p(\overline{\mathbf{h}}_k|\mathbf{y}_{k-1}, \overline{\boldsymbol{\mu}}_{k|k}) \\ &\quad + \log p(\mathbf{y}_{k-1}|\overline{\boldsymbol{\mu}}_{k|k})|\mathbf{y}_k, \tilde{\overline{\boldsymbol{\mu}}}_{k|k-1}] \\ &= -\{\log(\pi N_0) + E[(y(k) - \mathbf{s}(k)\mathbf{h}_k)^{\mathcal{H}} N_0^{-1} \\ &\quad \cdot (y(k) - \mathbf{s}(k)\mathbf{h}_k)|\mathbf{y}_k, \tilde{\overline{\boldsymbol{\mu}}}_{k|k-1}] \\ &\quad + \log((\pi)^{N(k+1)(L+1)} \det(\boldsymbol{\Sigma}_{k|k-1})) \\ &\quad + E[(\overline{\mathbf{h}}_k - \overline{\boldsymbol{\mu}}_{k|k-1})^{\mathcal{H}} \\ &\quad \cdot \boldsymbol{\Sigma}_{k|k-1}^{-1}(\overline{\mathbf{h}}_k - \overline{\boldsymbol{\mu}}_{k|k-1})|\mathbf{y}_k, \tilde{\overline{\boldsymbol{\mu}}}_{k|k-1}]\} \\ &\quad + E[\log p(\mathbf{y}_{k-1}|\overline{\boldsymbol{\mu}}_{k|k})|\mathbf{y}_k, \tilde{\overline{\boldsymbol{\mu}}}_{k|k-1}]. \end{aligned} \tag{26}$$

Similar to Section III-B-1, by taking the first and the second derivatives of $Q_k(\overline{\boldsymbol{\mu}}_{k|k}|\tilde{\overline{\boldsymbol{\mu}}}_{k|k-1})$ with respect to $\overline{\boldsymbol{\mu}}_{k|k}$ at point $\overline{\boldsymbol{\mu}}_{k|k} = \tilde{\overline{\boldsymbol{\mu}}}_{k|k-1}$, the estimation of $\boldsymbol{\Sigma}_{k|k}$ and $\overline{\boldsymbol{\mu}}_{k|k}$ becomes

$$\begin{aligned} \tilde{\boldsymbol{\Sigma}}_{k|k} &= \tilde{\boldsymbol{\Sigma}}_{k|k-1} - \tilde{\boldsymbol{\Sigma}}_{k|k-1}\overline{\mathbf{s}}(k)^{\mathcal{H}}(N_0 + \overline{\mathbf{s}}(k)\tilde{\boldsymbol{\Sigma}}_{k|k-1}\overline{\mathbf{s}}(k)^{\mathcal{H}})^{-1} \\ &\quad \cdot \overline{\mathbf{s}}(k)\tilde{\boldsymbol{\Sigma}}_{k|k-1} \end{aligned} \tag{27}$$

$$\begin{aligned} \tilde{\overline{\boldsymbol{\mu}}}_{k|k} &= \tilde{\overline{\boldsymbol{\mu}}}_{k|k-1} + \tilde{\boldsymbol{\Sigma}}_{k|k-1}\overline{\mathbf{s}}(k)^{\mathcal{H}}(N_0 + \overline{\mathbf{s}}(k)\tilde{\boldsymbol{\Sigma}}_{k|k-1}\overline{\mathbf{s}}(k)^{\mathcal{H}})^{-1} \\ &\quad \cdot (y(k) - \mathbf{s}(k)\tilde{\boldsymbol{\mu}}_{k|k-1}) \end{aligned} \tag{28}$$

where $\overline{\mathbf{s}}(k) = [\mathbf{s}(k), \mathbf{0}]$ and $\mathbf{0}$ is the $kN(L+1)$ zero row vector. The relation between $\overline{\mathbf{h}}_k$ and $\overline{\mathbf{h}}_{k-1}$ is given by

$$\overline{\mathbf{h}}_k = \begin{bmatrix} \boldsymbol{F}(k)\mathbf{h}_{k-1} \\ \cdots \\ \overline{\mathbf{h}}_{k-1} \end{bmatrix} + \begin{bmatrix} G(k)W(k) \\ \cdots \\ 0 \end{bmatrix}. \tag{29}$$

Obtaining $\tilde{\overline{\boldsymbol{\mu}}}_{k|k-1}$ and $\tilde{\boldsymbol{\Sigma}}_{k|k-1}$ from $\tilde{\overline{\boldsymbol{\mu}}}_{k-1|k-1}$ and $\tilde{\boldsymbol{\Sigma}}_{k-1|k-1}$ are straight forward by using (29). Meanwhile, since the first $L+1$ elements of the vector $\overline{\mathbf{s}}(k)$ are nonzero, only the first $L+1$ columns of $\tilde{\boldsymbol{\Sigma}}_{k|k-1}$ are necessary to be calculated for obtaining $\tilde{\overline{\boldsymbol{\mu}}}_{k|k}$.

The recursive relation (28) estimates the entire unknown sequence of parameters in each recursion by processing over a new sample of the received signal. If only $\tilde{\boldsymbol{\mu}}_{k|k}$, part of $\tilde{\overline{\boldsymbol{\mu}}}_{k|k}$, is selected from (28), the recursive formula becomes Kalman filtering. In general, by defining the unknown parameter vector $\boldsymbol{\mu}_{j|k}$ and the complete data $\mathcal{C}_k = \{\mathbf{y}_k, \mathbf{h}_j\}$ at time k and following the same procedure used in this subsection, it can be shown that the recursive EM algorithm leads to predicting, filtering and smoothing algorithms for $j>k$, $j=k$, and $j<k$, respectively. The recursive formula (28) is the estimation procedure of smoothing for the entire unknown parameters up to time k based on the available information up to this time. By selecting $\mathcal{C}_k = \{\mathbf{y}_k, \overline{\mathbf{h}}_{k+j}\}$ and unknown parameter vector $\overline{\boldsymbol{\mu}}_{k+j|k}$ where $j>0$, it is straightforward to modify the estimation procedure to include the prediction problem as well. In a smoothing algorithm it is common to assume that the estimator knows the entire received sequence and two Kalman algorithms, forward and backward, are applied in the estimating procedure. However, the recursive estimating method developed here based on the recursive EM algorithm is more general. Not only are the entire unknown processes re-estimated at each time, but also the boundary of the available data is changing with time.

C. Hybrid of Unknown Parameters and Gaussian Process

Sometimes the unknown set of parameters is a combination of the constant and sequential parameters. In this situation the estimation procedure needs both the methods developed in the Sections III-A and III-B. In general, the complete data may also be different for the two types of unknown parameters and the estimating algorithm can contain two combined recursive EM algorithms. Although there is interaction between the unknown parameters at different times, estimating parameters at time k depends on the estimation of the other parameters at time $k-1$. Hence, not only the unknown parameters, but also the complete data may be partitioned to run the recursive EM algorithms. Let us focus on more details of this situation in channel estimation.

The procedure for estimating the Gaussian random process CIR is based on knowing $\boldsymbol{F}(k)$ and $G(k)G(k)^{\mathcal{H}}$ matrices. These parameters are generally unknown and should also be estimated in the receiver. Assuming time-invariant $\boldsymbol{F}$ and G matrices, there are two unknown parameter sets at time k, $\boldsymbol{\theta}_{1k} = \overline{\boldsymbol{\mu}}_{k|k}$ and $\boldsymbol{\varphi}_2$, the elements of $\boldsymbol{F}$ and $GG^{\mathcal{H}}$. The complete data for estimating $\boldsymbol{\theta}_{1k}$ and $\boldsymbol{\varphi}_2$ is not the same. While the complete data for estimating $\boldsymbol{\theta}_{1k}$ is $\mathcal{C}_{1k} = \{\mathbf{y}_k, \overline{\mathbf{h}}_k\}$, the complete data for estimating $\boldsymbol{\varphi}_2$ at time k is $\mathcal{C}_{2k} = \overline{\mathbf{h}}_k$; however the incomplete data for both is $\mathcal{I}_{1k} = \mathcal{I}_{2k} = \mathbf{y}_k$. Since less informative complete data improves asymptotic convergence rate [17], entire set of unknown parameters $\{\boldsymbol{\theta}_{1k}, \boldsymbol{\varphi}_2\}$ is estimated with two separate recursive EM algorithms. The first recursive formula for estimating $\boldsymbol{\theta}_{1k}$ is the same as the Gaussian random process case where $Q_{1k}(\boldsymbol{\theta}_{1k}|\tilde{\boldsymbol{\theta}}_{1k|k-1}, \tilde{\boldsymbol{\varphi}}_{2|k-1})$ is defined as

$$\begin{aligned} &Q_{1k}(\boldsymbol{\theta}_{1k}|\tilde{\boldsymbol{\theta}}_{1k|k-1}, \tilde{\boldsymbol{\varphi}}_{2|k-1}) \\ &\quad = E[\log\, p(\mathbf{y}_k, \overline{\mathbf{h}}_k|\boldsymbol{\theta}_{1k}, \tilde{\boldsymbol{\varphi}}_{2|k-1})|\mathbf{y}_k, \tilde{\boldsymbol{\theta}}_{1k|k-1}, \tilde{\boldsymbol{\varphi}}_{2|k-1}] \end{aligned} \tag{30}$$

and its derivatives are taken with respect to $\boldsymbol{\theta}_{1k}^*$ at point $\boldsymbol{\theta}_{1k} = \tilde{\boldsymbol{\theta}}_{1k|k-1}$ and $\boldsymbol{\varphi}_2 = \tilde{\boldsymbol{\varphi}}_{2|k-1}$. The estimation procedure of $\boldsymbol{\theta}_{1k}$ is the same as the procedure described in Gaussian random process in Section III-B-2 using the estimates of $\boldsymbol{F}$ and $GG^{\mathcal{H}}$ at time $k-1$ instead of their real values. It is more convenient to estimate $\boldsymbol{\varphi}_2$ based on an ARMA model of $h_l(k) = h(l,k)$ instead of the state-space model. It is clear from (24), only F_i for $i = 1, \cdots, N$ and diagonal elements of $\mathbf{g}\mathbf{g}^{\mathcal{H}}$ should be estimated ($R_W(k) = I\delta(k)$). When $\mathrm{diag}(\mathbf{g}) = \{\mathbf{g}^0, \mathbf{g}^1, \cdots, \mathbf{g}^L\}$, the ARMA model of $h_l(k)$ is

$$h_l(k) = \mathbf{f}^l \mathbf{h}_{k-1} + \mathbf{g}^l w_l(k), \qquad 0 \le l \le L \tag{31}$$

where $\mathbf{f}^l$ is the lth row of $\boldsymbol{F}$ matrix. By defining $\boldsymbol{\varphi}_2 = [\mathbf{f}^L, \cdots, \mathbf{f}^0, R_{g^L}, \cdots, R_{g^0}]$ where $R_{g^l} = \mathbf{g}^l \mathbf{g}^{l^{\mathcal{H}}}$ for $l = 0, \cdots, L$, the E-step of estimating $\boldsymbol{\varphi}_2$ at time k is given by

$$\begin{aligned} &Q_{2k}(\boldsymbol{\varphi}_2 | \tilde{\boldsymbol{\theta}}_{1k|k}, \tilde{\boldsymbol{\varphi}}_{2|k-1}) \\ &\quad = E[\log\, p(\overline{\mathbf{h}}_k | \tilde{\boldsymbol{\theta}}_{1k|k}, \boldsymbol{\varphi}_2) | \mathbf{y}_k, \tilde{\boldsymbol{\theta}}_{1k|k-1}, \tilde{\boldsymbol{\varphi}}_{2|k-1}] \\ &\quad = \sum_{l=0}^{L} \left\{ \sum_{j=1}^{k} -\{\log(\pi R_{g^l}) + E[(h_l(j) - \mathbf{f}^l \mathbf{h}_{j-1})^{\mathcal{H}} \right. \\ &\qquad \cdot R_{g^l}^{-1}(h_l(j) - \mathbf{f}^l \mathbf{h}_{j-1}) | \mathbf{y}_k, \tilde{\boldsymbol{\theta}}_{1k|k-1}, \tilde{\boldsymbol{\varphi}}_{2|k-1}]\} \\ &\qquad \left. + E[\log\, p(h_l(0) | \tilde{\boldsymbol{\theta}}_{1k|k}, \boldsymbol{\varphi}_2) | \mathbf{y}_k, \tilde{\boldsymbol{\theta}}_{1k|k-1}, \tilde{\boldsymbol{\varphi}}_{2|k-1}] \right\}. \end{aligned} \tag{32}$$

By taking the first and the second derivatives of $Q_{2k}(\cdot|\cdot)$ with respect to $\mathbf{f}^{l^*}$ and $R_{g^l}^{-1}$ at point $\boldsymbol{\theta}_{1k} = \tilde{\boldsymbol{\theta}}_{1k|k}$ and $\boldsymbol{\varphi}_{2k} = \tilde{\boldsymbol{\varphi}}_{2|k-1}$ and also assuming $\tilde{\Sigma}_{j-1,j-1|j} \ll \tilde{\boldsymbol{\mu}}_{j-1|j} \tilde{\boldsymbol{\mu}}_{j-1|j}^{\mathcal{H}}$, $\tilde{\Sigma}_{j-1|j} \ll \tilde{\boldsymbol{\mu}}_{j-1|j} \tilde{\boldsymbol{\mu}}_{j-1|j}^{\mathcal{H}}$ and $\tilde{\Sigma}_{j|j} \ll \tilde{\boldsymbol{\mu}}_{j|j} \tilde{\boldsymbol{\mu}}_{j|j}^{\mathcal{H}}$ for $0 \le j \le k$ and doing some manipulations, it can be shown that [18]

$$\begin{aligned} \tilde{\mathbf{f}}_{|k}^l \simeq{}& \tilde{\mathbf{f}}_{|k-1}^l + \tilde{\boldsymbol{\mu}}_{k-1|k}^{\mathcal{H}} P_{k-1|k}^{\diamond} (1 + \tilde{\boldsymbol{\mu}}_{k-1|k}^{\mathcal{H}} P_{k-1|k}^{\diamond} \tilde{\boldsymbol{\mu}}_{k|k-1})^{-1} \\ &\cdot (\tilde{\mu}_{k|k}^l - \tilde{\mathbf{f}}_{|k-1}^l \tilde{\boldsymbol{\mu}}_{k-1|k}) \end{aligned} \tag{33}$$

$$\begin{aligned} \tilde{R}_{g^l|k} \simeq{}& \tilde{R}_{g^l|k-1} + \frac{1}{k} (\tilde{R}_{g^l|k-1} - |\tilde{\mu}_{k|k}^l - \tilde{\mathbf{f}}_{|k-1}^l \tilde{\boldsymbol{\mu}}_{k-1|k}|^2), \\ & 0 \le l \le L \end{aligned} \tag{34}$$

where $P_{k-1|k}^{\diamond} = (\Sigma_{j=1}^{k-1} \tilde{\boldsymbol{\mu}}_{j-1|j} \tilde{\boldsymbol{\mu}}_{j-1|j}^{\mathcal{H}})^{-1}$ and $P_{k|k+1}^{\diamond}$ is given by

$$\begin{aligned} P_{k|k+1}^{\diamond} = P_{k-1|k}^{\diamond} - P_{k-1|k}^{\diamond} \tilde{\boldsymbol{\mu}}_{k-1|k} (1 + \tilde{\boldsymbol{\mu}}_{k-1|k}^{\mathcal{H}} P_{k-1|k}^{\diamond} \\ \cdot \tilde{\boldsymbol{\mu}}_{k|k-1})^{-1} \tilde{\boldsymbol{\mu}}_{k-1|k}^{\mathcal{H}} P_{k-1|k}^{\diamond}. \end{aligned} \tag{35}$$

All conditional mean values in (33) and (34) can be obtained from (28). Meanwhile, the approximation procedure used to generate the recursive formula (33) is similar to RLS algorithm. For starting the algorithm proposed in Section III-C, such that the assumed approximations become valid we recommend that in the first step the stochastic RLS algorithm, (23), be used for estimating the initial values of $\boldsymbol{\mu}_{j-1|j}$ and $\boldsymbol{\mu}_{j|j}$ for a short time period, then (33) and (34) be used to estimate the initial values of $\mathbf{f}^l$ and R_{g^l} and then finally the complete procedure proposed in Section III-C be applied.

IV. Conclusions

The recursive (online) EM algorithm was presented for estimating time-invariant/variant parameters by using the generalized form of Titterington's approach. From the recursive EM algorithm, we have obtained different types of RLS, Kalman and combined RLS/Kalman-type algorithms some of which were not available before. These algorithms were derived directly based on the EM approach which emerged as a powerful tool for the unification of different types of adaptive algorithms. Meanwhile, the channel estimation algorithms proposed in the paper have yielded good results when simulated in the framework of MLSD receivers [15], [18].

References

[1] 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, pp. 1–38, 1977.

[2] T. K. Moon, "The expectation-maximization algorithm," *IEEE Signal Processing Mag.*, vol. 13, pp. 47–60, Nov. 1996.

[3] M. Feder and E. Weinstein, "Parameter estimation of superimposed signal using the EM algorithm," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 36, pp. 477–489, Apr. 1988.

[4] E. Weinstein, M. Feder, and A. Oppenheim, "Sequential algorithm for parameter estimation based on the Kullback-Leibler information measure," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, pp. 1652–1654, Sept. 1990.

[5] E. Weinstein, A. Oppenheim, M. Feder, and J. R. Buck, "Iterative and sequential algorithms for multisensor signal enhancement," *IEEE Trans. Signal Processing*, vol. 42, pp. 846–859, Aug. 1994.

[6] M. Segal and E. Weinstein, "A new method evaluating the log-likelihood gradient, the Hessian, and the Fisher information matrix for linear dynamic systems," *IEEE Trans. Inform. Theory*, vol. 35, pp. 682–687, May 1989.

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

[8] V. Krishnamurthy, "On-line estimation of dynamic shock-error models based on the Kullback–Leibler information measure," *IEEE Trans. Automat. Contr.*, pp. 1129–1135, May 1994.

[9] R. H. Shumway and D. S. Stoffer, "An approach to time series smoothing and forecasting using the em algorithm," *J. Time Series Anal.*, vol. 3, no. 4, pp. 253–264, 1982.

[10] L. B. Nelson and H. V. Poor, "Iterative multiuser receivers for CDMA channels: An EM-based approach," *IEEE Trans. Commun.*, vol. 44, pp. 1700–1710, Dec. 1996.

[11] G. K. Kaleh and R. Vallet, "Joint parameter estimation and symbol detection for linear or nonlinear unknown channels," *IEEE Trans. Commun.*, vol. 42, pp. 2406–2413, July 1994.

[12] C. N. Georghiades and J. C. Han, "Sequence estimation in presence of random parameters via the EM algorithm," *IEEE Trans. Commun.*, vol. 45, pp. 300–308, Mar. 1997.

[13] L. Ljung and T. Soderstrom, *Theory and Practice of Recursive Identification*. London, U.K.: The MIT Press, 1986.

[14] D. M. Titterington, "Recursive parameter estimation using incomplete data," *J. Roy. Stat. Soc. B*, vol. 46, no. 2, pp. 257–267, 1984.

[15] H. Zamiri-Jafarian and S. Pasupathy, "Adaptive MLSDE using the EM algorithm," *IEEE Trans. Commun.*, vol. 47, pp. 1181–1193, July 1999.

[16] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1993.

[17] J. A. Fessler and A. O. Hero, "Space-alternating generalized expectation-maximization algorithm," *IEEE Trans. Signal Processsing*, vol. 42, pp. 2664–2677, Oct. 1994.

[18] H. Zamiri-Jafarian, "Adaptive MLSDE receivers for wireless communications," Ph.D. dissertation, Dept. Elect. Comp. Eng., Univ. of Toronto, Toronto, Canada, 1998.