A New Class of Sequential and Adaptive Algorithms with Application to Noise Cancellation *

S12.8

by

Meir Feder,[1] Ehud Weinstein[2] and Alan V. Oppenheim[1]

Abstract

In this paper we present a class of sequential and adaptive algorithms for parameter estimation. These algorithms are based on the iterative Estimate-Maximize (EM) algorithm. In some cases we will be able to derive sequential algorithms that perform an exact EM step in each recursion; an example for these cases will be given for the linear least-squares problem. In general, however, we will have to approximate the EM iteration in order to develop sequential algorithms. Possible application of this new class of algorithm to the two-microphone noise cancellation problem will be presented.

1 Introduction

The Estimate-Maximize (EM) method is a class of iterative batch algorithms for parameter estimation, [1]. In the EM algorithm, the observations are considered "incomplete" with respect to a more convenient "complete data" measurements. The algorithm iterates between estimating the sufficient statistics of the "complete data" given the observations and a current estimate of the parameters (the E step) and maximizing the likelihood of the complete data, using the estimated sufficient statistics (the M step).

This EM method has been applied to several signal processing problems, e.g. [2], [3] and recently [4], [5] and [6]. In many signal processing applications it is desirable to use sequential or adaptive algorithms; the data may arrive sequentially, a sequential processing may be more efficient and we may want to track varying parameters. For these reasons, we will suggest and investigate in this paper a new class of sequential and adaptive algorithms, based on the EM concept.

Many sequential and adaptive algorithms are based on a given iterative algorithm. A well known example is the stochastic gradient algorithm, which is an adaptive version of the iterative gradient algorithm. As another example, the recursive least-squares (RLS) algorithm and the (extended) Kalman algorithm are sequential algorithms based on the iterative Newton-Raphson method. Similarly, as will be demonstrated in this paper, the iterative EM algorithm also suggests sequential and adaptive algorithms.

The paper is organized as follows: In section 2 we will discuss sequential algorithms which exactly implement an EM recursion. These algorithms, however, may be applied only when the underlying estimation problem has a special structure. In section 3 we will use approximations and develop sequential and adaptive algorithms, based on the EM method, that may be applied in general. In section 4, we will suggest an application of the sequential EM algorithms to the noise cancellation problem. This suggestion is motivated by the successful application of the batch EM algorithm to this problem, [6], [7].

*This work has been supported in part by the Advanced Research Projects Agency monitored by ONR under contract N00014-81-K-0742 and the National Science Foundation under grant ECS-8407285 at MIT, and in part by ONR under contract N00014-85-K-0272 at WHOI.

[1]Meir Feder and Alan V. Oppenheim are with the Department of Electrical Engineering and Computer Science, Research Laboratory of Electronics, Room 36-615, Massachusetts Institute of Technology, Cambridge, MA. 02139

[2]Ehud Weinstein is with the Department of Electronic systems, Faculty of Engineering, Tel-Aviv University, Tel-Aviv 69978, Israel. He is also an adjunct scientist with the Department of Ocean Engineering, Woods Hole Oceanographic Institution, Woods Hole, MA. 02543.

2 Sequential EM with exact EM mapping

Throughout this paper, we will consider the observed data as blocks, $\underline{y}_1, \underline{y}_2, \ldots, \underline{y}_n, \ldots$, to be processed sequentially. The complete data is denoted $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n, \ldots$, and is chosen so that each block of observed data, $\underline{y}_n$, corresponds to a block of complete data, $\underline{x}_n$, by

$$\underline{y}_n = T_n(\underline{x}_n) \tag{1}$$

where $T_n(\cdot)$ is a non-invertible transformation.

In this environment the log-likelihood of the observations, after $n+1$ data blocks have been observed, is given by,

$$L_{n+1}(\underline{\theta}) = \log f_{Y_{n+1}\cdots Y_1}(\underline{y}_{n+1}, \cdots, \underline{y}_1; \underline{\theta}) \tag{2}$$

Using the complete data, $\underline{x}_1, \cdots, \underline{x}_{n+1}$, and following the basic identity of the EM algorithm (see [1], equation (3.2)), the log-likelihood of the observations may be written as,

$$L_{n+1}(\underline{\theta}) = Q_{n+1}(\underline{\theta}; \underline{\theta}') - H_{n+1}(\underline{\theta}; \underline{\theta}') \tag{3}$$

where

$$Q_{n+1}(\underline{\theta}, \underline{\theta}') = \tag{4}$$

$$= E\left\{\log f_{X_{n+1}\cdots X_1}(\underline{x}_{n+1}, \cdots, \underline{x}_1; \underline{\theta}) \,\Big|\, \underline{y}_{n+1}, \cdots, \underline{y}_1; \underline{\theta}'\right\}$$

and

$$H_{n+1}(\underline{\theta}, \underline{\theta}') = \tag{5}$$

$$= E\left\{\log f_{X_{n+1}\cdots X_1/\underline{y}_{n+1}\cdots \underline{y}_1}(\underline{x}_{n+1}, \cdots, \underline{x}_1; \underline{\theta}) \,\Big|\, \underline{y}_{n+1}, \cdots, \underline{y}_1; \underline{\theta}'\right\}$$

$E\{\cdot\}$ denotes statistical expectation.

An EM algorithm for solving the maximum likelihood problem, given these $n+1$ blocks of data, using the above definition of complete data, is given by the following iteration,

$$\underline{\theta}^{(k+1)} = \arg\max_{\underline{\theta}\in\Theta} Q_{n+1}(\underline{\theta}; \underline{\theta}^{(k)}) = \tag{6}$$

$$= \arg\max_{\underline{\theta}\in\Theta} E\left\{\log f_{X_{n+1}\cdots X_1}(\underline{x}_{n+1}, \cdots, \underline{x}_1; \underline{\theta}) \,\Big|\, \underline{y}_{n+1}, \cdots, \underline{y}_1, \underline{\theta}^{(k)}\right\}$$

where k denotes the iteration index and n the data index.

A sequential EM algorithm with exact EM mapping is a method that recalculates in each iteration, as more data is processed, the exact steps of the EM algorithm for maximizing the new likelihood function. For convenience, suppose we perform a single EM iteration for each new observed data block, i.e. the iteration and the data indices are equivalent. This mapping is given by (7) where k is replaced by n. This EM mapping is, in general, a function of all given observed data blocks; thus, it may written abstractly as,

$$\underline{\theta}^{(n+1)} = M_{n+1}(\underline{\theta}^{(n)}; \underline{y}_{n+1}, \cdots, \underline{y}_1) \tag{7}$$

The exact EM iteration may be implemented recursively, when the effect of the past data blocks, $\underline{y}_n, \cdots, \underline{y}_1$, can be summarized into a small number of simple quantities. We may algebraically manipulate the given expression for the EM iteration and achieve an equivalent expression, that may be written abstractly as the mapping,

$$\underline{\theta}^{(n+1)} = M'_{n+1}\left(\underline{\theta}^{(n)}; \underline{y}_{n+1}, \underline{g}(\underline{y}_n, \cdots, \underline{y}_1)\right) \tag{8}$$

where $\underline{g}$ indicates easily stored and updated functions of the past observations.

We will assume that the structure of (8) may be achieved for all n. In this case, we suggest the following sequential EM algorithm:

- Start, $n = 0$: Guess $\underline{\theta}^{(0)}$. Initialize $\underline{g}(\cdot, \cdots) = 0$
- For each new data block, $\underline{y}_{n+1}$.
  - **Exact EM mapping:** Update parameters,

$$\underline{\theta}^{(n+1)} = M'_{n+1}\left(\underline{\theta}^{(n)}; \underline{y}_{n+1}, \underline{g}(\underline{y}_n, \cdots, \underline{y}_1)\right) \quad (9)$$

  - Update and record $\underline{g}(\underline{y}_{n+1}, \cdots, \underline{y}_1)$ for the next step
  - $n \Longrightarrow n+1$

In each step, this algorithm implements the exact EM mapping for maximizing the new likelihood $L_{n+1}(\underline{\theta})$, and thus,

$$L_{n+1}(\underline{\theta}^{(n+1)}) \geq L_{n+1}(\underline{\theta}^{(n)})$$

This algorithm has been presented abstractly so far. As an illustration, consider a simple example, in which a linear least squares problem is solved recursively using this algorithm.

Example: Sequential Least Squares EM algorithm

It is well known that the linear least-squares problem may be posed as a statistical maximum likelihood problem, in the following way. Suppose we observe a vector, $\underline{y} = (y_1, \cdots, y_n)^T$, given by,

$$\underline{y} = A \cdot \underline{\theta} + \underline{n} \quad (10)$$

where $\underline{\theta} = (\theta_1, \cdots, \theta_k)^T$ is the unknown parameter vector, $\underline{n} = (n_1, \cdots, n_n)^T$ is the noise vector, where $\{n_i\}$ are i.i.d random variables distributed normally with zero mean and variance σ^2, and A is a given $(n \times k)$ matrix, which may be written by columns as $A = [\underline{a}_1, \cdots, \underline{a}_k]$ or by rows as $A^T = [\underline{\alpha}_1, \cdots, \underline{\alpha}_n]$. In this case maximizing the likelihood of the observation yield a least-squares problem as,

$$\hat{\underline{\theta}}_{ML} = \arg\max_{\underline{\theta}} \log f_Y(\underline{y}; \underline{\theta}) = \arg\min_{\underline{\theta}} \frac{1}{2\sigma^2} ||\underline{y} - A \cdot \underline{\theta}||^2 \quad (11)$$

An iterative EM algorithm for this ML problem is as follows. Let the complete data be the vectors $\{\underline{x}_j\}_{j=1}^k$ where,

$$\underline{x}_j = \underline{a}_j \cdot \theta_j + \underline{n}_j \quad (12)$$

$\underline{n}_j$ is $(n \times 1)$ noise vector, whose components n_{ji} are zero mean Gaussian i.i.d random variables with variance $\beta_j \sigma^2$. Assuming that $\{\underline{n}_j\}$ are uncorrelated and that $\sum_{j=1}^k \beta_j = 1$, we have

$$\underline{y} = \sum_{j=1}^{k} \underline{x}_j \quad (13)$$

As shown in [4], the E and M steps of an EM algorithm for solving the least-squares problem of (11), using the complete data above, are given by,

- E step:

$$\underline{x}_j^{(n+1)} = \underline{x}_j^{(n)} + \beta_j \left(\underline{y} - A \cdot \underline{\theta}^{(n)}\right), \quad j = 1, \cdots, k \quad (14)$$

- M step

$$\theta_j^{(n+1)} = \arg\min_{\theta_j} ||\underline{x}_j^{(n)} - \underline{a}_j \cdot \theta_j||^2 = \frac{\underline{a}_j^T \underline{x}_j^{(n)}}{||\underline{a}_j||^2}, \quad j = 1, \cdots, k \quad (15)$$

Combining these two steps we get the iteration,

$$\underline{\theta}^{(n+1)} = \underline{\theta}^{(n)} + diag\left(\frac{\beta_1}{||\underline{a}_1||^2}, \cdots, \frac{\beta_k}{||\underline{a}_k||^2}\right) \cdot A^T \cdot \left(\underline{y} - A \cdot \underline{\theta}^{(n)}\right) \quad (16)$$

where $diag(\cdot, \cdots, \cdot)$ is a diagonal matrix.

A sequential algorithm, based on the iteration (16), according to the exact EM mapping method, may now be easily developed. Define a "correlation matrix", $\mathcal{A}_n$, and a "cross-correlation vector", $\underline{p}_n$, for the least squares problem of order n in the following way,

$$\begin{aligned} \mathcal{A}_n = \frac{1}{n} A^T A &= \frac{1}{n} \sum_{i=1}^{n} \underline{\alpha}_i \underline{\alpha}_i^T \\ \underline{p}_n = \frac{1}{n} A^T \underline{y} &= \frac{1}{n} \sum_{i=1}^{n} \underline{\alpha}_i \cdot y_i \end{aligned} \quad (17)$$

Given a new measurement, y_{n+1}, we can update $\mathcal{A}_n$ and $\underline{p}_n$ recursively, as,

$$\begin{aligned} \mathcal{A}_{n+1} &= \frac{n}{n+1} \mathcal{A}_n + \frac{1}{n+1} \underline{\alpha}_{n+1} \underline{\alpha}_{n+1}^T \\ \underline{p}_{n+1} &= \frac{n}{n+1} \underline{p}_n + \frac{1}{n+1} \underline{\alpha}_{n+1} \cdot y_{n+1} \end{aligned} \quad (18)$$

The exact EM iteration (16) may be written as,

$$\underline{\theta}^{(n+1)} = \underline{\theta}^{(n)} + diag\left(\frac{\beta_1}{\mathcal{A}_{n+1}(1,1)}, \cdots, \frac{\beta_k}{\mathcal{A}_{n+1}(k,k)}\right) \cdot \left(\underline{p}_{n+1} - \mathcal{A}_{n+1} \cdot \underline{\theta}^{(n)}\right) \quad (19)$$

which can be calculated recursively, since all required quantities are calculated recursively. The sequential least squares EM algorithm (SLSEM) is completely specified by (18) and (19).

3 Sequential EM based on approximations

The sequential algorithms presented so far were suggested by assuming that the underlying problem had a special structure. In this section, we will address the general situation. Unfortunately, sequential algorithms may not be derived directly from the EM algorithm in the general case. We will therefore suggest algorithms that approximate the EM iteration, in order to get a recursive implementation.

3.1 General sequential considerations

The log-likelihood of the observations, given $n+1$ data blocks, is given by (2). Define,

$$L_{n+1/n}(\underline{\theta}) = \log f_{Y_{n+1}/Y_n \cdots Y_1}(\underline{y}_{n+1}/\underline{y}_n, \cdots, \underline{y}_1; \underline{\theta}) \quad (20)$$

The log-likelihood of the observations, eq. (2), may be written recursively as,

$$L_{n+1}(\underline{\theta}) = L_n(\underline{\theta}) + L_{n+1/n}(\underline{\theta}) \quad (21)$$

or as,

$$L_{n+1}(\underline{\theta}) = L_1(\underline{\theta}) + \sum_{i=1}^{n} L_{i+1/i}(\underline{\theta}) \quad (22)$$

We note that analogous to (3), the term L_n may be written as,

$$L_n(\underline{\theta}) = Q_n(\underline{\theta}; \underline{\theta}') - H_n(\underline{\theta}; \underline{\theta}') \quad (23)$$

where the complete data is defined to be $\underline{x}_1, \cdots, \underline{x}_n$.

One approach for developing recursive EM algorithm will refer to the recursive formula of the log-likelihood (21). For the term $L_{n+1/n}$, the complete data is $\underline{x}_{n+1}$ and following the same considerations which lead to (23), we may write,

$$L_{n+1/n}(\underline{\theta}) = Q_{n+1/n}(\underline{\theta}; \underline{\theta}') - H_{n+1/n}(\underline{\theta}; \underline{\theta}') \quad (24)$$

where

$$Q_{n+1/n}(\underline{\theta}, \underline{\theta}') = \tag{25}$$

$$= E\left\{\log f_{X_{n+1}/Y_n \cdots Y_1}(\underline{x}_{n+1}/\underline{y}_n, \cdots, \underline{y}_1; \underline{\theta}) \,\middle|\, \underline{y}_{n+1}, \cdots, \underline{y}_1; \underline{\theta}'\right\}$$

and

$$H_{n+1/n}(\underline{\theta}, \underline{\theta}') = \tag{26}$$

$$= E\left\{\log f_{X_{n+1}/Y_{n+1} \cdots Y_1}(\underline{x}_{n+1}/\underline{y}_{n+1}, \cdots, \underline{y}_1; \underline{\theta}) \,\middle|\, \underline{y}_{n+1}, \cdots, \underline{y}_1; \underline{\theta}'\right\}$$

Therefore,

$$L_{n+1}(\underline{\theta}) = L_n(\underline{\theta}) + L_{n+1/n}(\underline{\theta}) = \tag{27}$$

$$= Q_n(\underline{\theta}; \underline{\theta}') + Q_{n+1/n}(\underline{\theta}; \underline{\theta}') - \left[H_n(\underline{\theta}; \underline{\theta}') + H_{n+1/n}(\underline{\theta}; \underline{\theta}')\right]$$

and we have,

$$H_n(\underline{\theta}; \underline{\theta}') \leq H_n(\underline{\theta}'; \underline{\theta}') \quad \text{and} \quad H_{n+1/n}(\underline{\theta}; \underline{\theta}') \leq H_{n+1/n}(\underline{\theta}'; \underline{\theta}') \tag{28}$$

One could try to achieve a recursive algorithm by maximizing either,

$$Q_n(\underline{\theta}; \underline{\theta}^{(n)}) + Q_{n+1/n}(\underline{\theta}; \underline{\theta}^{(n)}) \tag{29}$$

or,

$$Q_1(\underline{\theta}; \underline{\theta}^{(n)}) + Q_{2/1}(\underline{\theta}; \underline{\theta}^{(n)}) + \cdots + Q_{n+1/n}(\underline{\theta}; \underline{\theta}^{(n)}) \tag{30}$$

since maximizing either (29) or (30) will generate a new value $\underline{\theta}^{(n+1)}$ that increases the likelihood $L_{n+1}(\underline{\theta})$. However, despite their seemingly recursive structure, these maximizations cannot be performed sequentially in general, because:

- Calculating $Q_{n+1/n}$ involves the past data $\underline{y}_n, \cdots, \underline{y}_1$
- For each new parameter value, the conditional expectations needed for the terms $Q_1, Q_{2/1}, \cdots, Q_{n/n-1}$ or the term Q_n, should be recalculated. This requires using the past data samples.

Alternative approach

The batch EM algorithm updates the parameter estimate by maximizing $Q_{n+1}(\underline{\theta}; \underline{\theta}')$, where Q_{n+1} is defined in (4). However, using recursive formulas for the likelihood of the complete data we may write,

$$Q_{n+1}(\underline{\theta}; \underline{\theta}') = \tilde{Q}_{n+1/n} + Q_n(\underline{\theta}; \underline{\theta}') = \sum_{i=1}^{n} \tilde{Q}_{i+1/i}(\underline{\theta}; \underline{\theta}') + Q_1(\underline{\theta}; \underline{\theta}') \tag{31}$$

where

$$\tilde{Q}_{i+1/i}(\underline{\theta}; \underline{\theta}') = \tag{32}$$

$$E\left\{\log f_{X_{i+1}/X_i \cdots X_1}(\underline{x}_{i+1}/\underline{x}_i, \cdots, \underline{x}_1; \underline{\theta}) \,\middle|\, \underline{y}_{n+1}, \cdots, \underline{y}_1; \underline{\theta}'\right\}$$

One may try to achieve a recursive algorithm by maximizing (31). Again, however, despite the seemingly recursive structure of this maximization, it cannot be performed sequentially in general, since:

- Calculating $\tilde{Q}_{i+1/i}$ involves the observed data $\underline{y}_n, \cdots, \underline{y}_1$, and all the past complete data $\underline{x}_i, \cdots, \underline{x}_1$.
- For each new parameter value, the conditional expectations needed for the terms $\tilde{Q}_1, \tilde{Q}_{2/1}, \cdots, \tilde{Q}_{n/n-1}$, should be recalculated. This requires using the past data samples.

3.2 Approximate sequential algorithms

The problems mentioned above occurred in both approaches when we tried to calculate sequentially the exact EM iteration for the general case. To overcome these problems and to achieve sequential algorithms, we will approximate the desired EM iteration. The resulting algorithms are no longer EM algorithms; nevertheless, as shown in [4], these algorithms are related to the method of stochastic approximation and thus, the convergence results and the asymptotic behavior of stochastic approximation methods are applied to these algorithms.

Consider the following sequential procedure,

- Start, $n = 0$: Initialize $\Psi_0(\underline{\theta}) = 0$. Guess $\underline{\theta}^{(0)}$
- For each new data $\underline{y}_{n+1}$,
  - **E-step:** calculate

$$Q^a_{n+1/n}(\underline{\theta}, \underline{\theta}^{(n)}) = \tag{33}$$

$$E\left\{\log f_{X_{n+1}/X_n \cdots X_{n-q}}(\underline{x}_{n+1}/\underline{x}_n, \cdots, \underline{x}_{n-q}; \underline{\theta}) \,\middle|\, \underline{y}_{n+1}, \underline{y}_n, \cdots, \underline{y}_{n-m}; \underline{\theta}^{(n)}\right\}$$

  - **M-step:** solve

$$\underline{\theta}^{(n+1)} = \arg\max_{\underline{\theta} \in \Theta} \left[Q^a_{n+1/n}(\underline{\theta}, \underline{\theta}^{(n)}) + \beta_n \cdot \Psi_n(\underline{\theta})\right] \tag{34}$$

  - Record for next step

$$\Psi_{n+1}(\underline{\theta}) = Q^a_{n+1/n}(\underline{\theta}, \underline{\theta}^{(n)}) + \beta_n \cdot \Psi_n(\underline{\theta})$$

  - $n \Longrightarrow n + 1$

This algorithm approximates the procedures of maximizing (30) or (31) as follows. First, the term $Q^a_{n+1/n}(\underline{\theta}; \underline{\theta}^{(n)})$, given by (33), approximates $Q_{n+1/n}(\underline{\theta}; \underline{\theta}^{(n)})$ or $\tilde{Q}_{n+1/n}(\underline{\theta}; \underline{\theta}^{(n)})$. We will use q past complete data samples $\underline{x}_n, \cdots, \underline{x}_{n-q}$ so that the conditional likelihood of the complete data is a simple function of the parameters. We will use in this approximation m past observations values, $\underline{y}_n, \cdots, \underline{y}_{n-m}$, as long as $Q^a_{n+1/n}$ is calculated recursively. We note that, if the different observation blocks are independent, $Q_{n+1/n} = \tilde{Q}_{n+1/n} = Q^a_{n+1/n}$. In general, the weaker the successive observations blocks are correlated, the better this approximation becomes. Second, the previous terms are not recalculated. We calculate each $Q^a_{i+1/i}$, using the corresponding parameter value, $\underline{\theta}^{(i)}$, and we simply accumulate these functions and generate $\Psi_n(\underline{\theta})$ recursively. Also using this algorithm, the previous terms may be weighted, according to the choice of β_n. By an appropriate choice, we may reduce the contribution of the past data and track varying parameters in the adaptive situation, or we may weight the past data more heavily, to guarantee convergence and consistency, for a sequential algorithm.

4 Application to the noise cancellation problem

A new approach for solving the two-microphone noise cancellation problem has been suggested in [7] and [6]. In this approach, a statistical ML problem has been formulated for estimating the parameters required for the cancellation, and an iterative algorithm, based on the EM method, has been suggested and applied to solve the ML problem. The observation model in this formulation is summarized in Figure 1, i.e.

$$y_1(t) = s(t) + A\{w(t)\} + e_1(t) \tag{35}$$

$$y_2(t) = B\{s(t)\} + w(t) + e_2(t) \tag{36}$$

where, $y_1(t)$ and $y_2(t)$ are the observed signals, $s(t)$ is the desired (speech) signal, $w(t)$ is the noise source signal, and $e_1(t)$ and $e_2(t)$ are the measurement and modeling error signals in the two microphones. For the statistical formulation, $s(t)$ is modeled as a sample from a Gaussian AR random process, while $w(t)$, $e_1(t)$ and $e_2(t)$ are white Gaussian noise processes. The systems A and B are assumed to be linear FIR filters, whose orders are q_a and q_b respectively. The mathematics and the algorithms are formulated in terms of discrete time signals with the independent variable t representing normalized sample time.

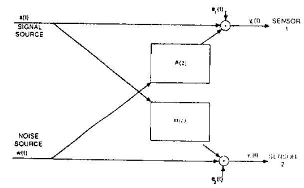

Figure 1: The observations model

The suggested iterative EM algorithm is summarized in Figure 2. This is an intuitively appealing scheme; the algorithm iterates between a two-channel Wiener filtering that estimates $s(t)$ and $w(t)$ given the parameters of the systems and solving sets of linear equations for updating the parameters.

This approach improves upon the currently used methods for noise cancellation, e.g. [8]. However, the algorithm of Figure 2 is a batch algorithm, which imposes certain difficulties. For example it cannot smoothly track changing parameters in a non-stationary environment and it processes the signals in blocks which leads to discontinuities and other block-effect distortions in the processed signals.

To overcome these problems, a sequential (adaptive) algorithm is suggested for this two-microphone noise cancellation problem, following the general sequential procedure of (33) and (34). We will use the statistical model of Figure 1 above, and we choose the complete data that led to the algorithm of Figure 2 i.e.

$$x(t) = \{y_1(t), y_2(t), s(t), w(t)\} \tag{37}$$

As in the case of the batch algorithm, the unknown parameters will be the coefficients $\{a_k\}$ and $\{b_k\}$ of the systems A and B, and the spectral parameters (LPC parameters) of the desired (speech) signal $s(t)$.

The sequential algorithm calculates in the E step the conditional expectation of the conditional log-likelihood of the complete data. This conditional log-likelihood depends on q past samples of the complete data. A natural choice for q, in the noise cancellation problem, is the maximum of the orders of the FIR filters A and B, i.e. $q = \max[q_a, q_b]$. We may now write,

$$\log f(x(t)/x(t-1), \cdots, x(t-q); \underline{\theta}) = \tag{38}$$

$$\underbrace{\log f(y_1(t)/s(t), \cdots, s(t-q), w(t), \cdots, w(t-q); \underline{\theta})}_{I} +$$

$$+ \underbrace{\log f(y_2(t)/s(t), \cdots, s(t-q), w(t), \cdots, w(t-q); \underline{\theta})}_{II} +$$

$$+ \underbrace{\log f(s(t)/s(t-1), \cdots, s(t-q); \underline{\theta})}_{III} +$$

$$+ \underbrace{\log f(w(t)/w(t-1), \cdots, w(t-q); \underline{\theta})}_{IV}$$

The term I depends only on $\{a_k\}$ as,

$$-\frac{1}{\sigma_{e_1}^2}\left[y_1(t) - s(t) - \sum_{k=0}^{q_a} a_k w(t-k)\right]^2 \tag{39}$$

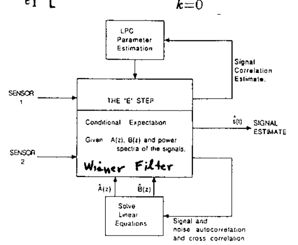

Figure 2: Noise cancellation using the EM algorithm

Similarly, the term II depends on $\{b_k\}$ as,

$$-\frac{1}{\sigma_{e_2}^2}\left[y_2(t) - w(t) - \sum_{k=0}^{q_b} b_k s(t-k)\right]^2 \tag{40}$$

The terms III and IV depends on the spectral parameters of the $s(t)$ and $w(t)$ respectively. The desired signal $s(t)$ is assumed to be an AR process of order r with unknown parameters G (the gain) and $\{h_k\}$ (the AR coefficients). The term III will thus be,

$$-\frac{1}{G}\left[s(t) - \sum_{k=1}^{r} h_k s(t-k)\right]^2 \tag{41}$$

For implementing the E step we have to take the conditional expectation of (39)-(41) given the observed data. We note that this requires the expectation of the signal samples $s(t), \cdots, s(t-q)$ the noise samples $w(t), \cdots, w(t-q)$ and cross terms such as $s(t) \cdot s(t-k)$, $w(t) \cdot w(t-k)$ and $s(t) \cdot w(t-k)$. All these terms may be calculated *recursively* from the entire past observation data, using Kalman filter and Kalman smoothing formulas, and the associated error covariance matrix formulas. This Kalman filter is determined, of course, by the current parameter values.

The sequential M step, given by (34), requires adding the expected values of (39)-(41) to the similar terms accumulated so far, and update the parameters by maximizing the appropriate terms (i.e. update $\{a_k\}$ by maximizing the accumulation of the terms similar to (39)). This M step is reduced to solving a set of linear equations. In order to avoid solving large sets of linear equations in each step, we can use the fact that we have, in each step, a current value of the parameters. Thus, we can perform in the M step a gradient step, a Newton-Raphson step, or even a step of the SLSEM algorithm, i.e. the EM algorithm suggested previously for the linear least-squares problem, instead of re-solving the linear least-squares problem for each new observed data.

To summarize, the suggested sequential scheme for the noise cancellation problem has the attractive structure of Figure 2, where a recursive two-channel Kalman filter block replaces the Wiener filter block, and recursive least-squares blocks replace the batch least-squares blocks of the original batch EM algorithm.

References

[1] 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*, Ser. 39:1–38, 1977.

[2] J. S. Lim and A. V. Oppenheim. All pole modeling of degraded speech. *IEEE Trans. Acoustics, Speech, and Signal Processing*, ASSP-26:197–210, 1978.

[3] B. R. Musicus. *Iterative algorithms for optimal signal reconstruction and parameter identification given noisy and incomplete data.* PhD thesis, Massachusetts Institute of Technology, September 1982.

[4] M. Feder. *Iterative Algorithms for Parameter Estimation with applications to Signal Processing.* ScD thesis, Massachusetts Institute of Technology, 1987.

[5] M. Feder and E. Weinstein. Parameter estimation of superimposed signals using the EM algorithm. *IEEE Trans. Acoustics, Speech, and Signal Processing*, April, 1988.

[6] M. Feder, Alan V. Oppenheim, and E. Weinstein. Maximum likelihood noise cancellation using the EM algorithm. *IEEE Trans. Acoustics, Speech, and Signal Processing*, accepted for publication, 1988.

[7] M. Feder, Alan V. Oppenheim, and E. Weinstein. Methods for noise cancellation based on the EM algorithm. In *Proceedings of the 1987 International Conference on Acoustics Speech and Signal Processing*, 1987.

[8] B. Widrow et al. Adaptive noise canceling: principles and applications. *Proc. IEEE*, 63:1692–1716, 1975.