Signal Enhancement Using Single and Multi-Sensor Measurements

RLE Technical Report No. 560

E. Weinstein, A.V. Oppenheim and M. Feder

November 1990

Research Laboratory of Electronics
Massachusetts Institute of Technology
Cambridge, MA 02139 USA

Signal Enhancement Using Single and Multi-Sensor Measurements

RLE Technical Report No. 560

E. Weinstein,[1] A.V. Oppenheim[2] and M. Feder[3]

November 1990

Research Laboratory of Electronics
Massachusetts Institute of Technology
Cambridge, MA 02139 USA

This work was supported in part by the Defense Advanced Research Projects Agency monitored by the Office of Naval Research under Grant Number N00014-89-J-1489 at MIT, in part by Lockheed/Sanders Inc. at MIT, and in part by the Office of Naval Research under Grant Number N00014-90-J-1109 at Woods Hole Oceanographic Institution.

---

[1] Department of Electrical Engineering-Systems, Faculty of Engineering, Tel-Aviv University, Tel-Aviv 69978, Israel; adjunct scientist with the Department of Applied Ocean Physics and Engineering, Woods Hole Oceanographic Institution, Woods Hole, MA 02543.

[2] Department of Electrical Engineering and Computer Science and the Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139.

[3] Department of Electrical Engineering-Systems, Faculty of Engineering, Tel-Aviv University, Tel-Aviv 69978, Israel.

Abstract

A time-domain approach to signal enhancement based on single and multiple sensor measurements is developed. The procedure is based on the iterative Estimate-Maximize (EM) algorithm for maximum likelihood estimation. On each iteration, in the M step of the algorithm, parameter values are estimated based on the signal estimates obtained in the E step of the prior iteration. The E step is then applied using these parameter estimates to obtain a refined estimate of the signal. In our formulation, the E step is implemented in the time domain using a Kalman smoother. This enables us to avoid many of the computational and conceptual difficulties with prior frequency domain formulations. Furthermore, the time domain formulation leads naturally to a time-adaptive algorithm by replacing the Kalman smoother with a Kalman filter and in place of successive iterations on each data block, the algorithm proceeds sequentially through the data with exponential weighting applied to permit the algorithm to adapt to changes in the structure of the data. A particularly efficient implementation of the time-adaptive algorithm is formulated for both the single- and two-sensor cases by exploiting the structure of the Kalman filtering equations. In addition an approach to avoiding matrix inversions in the modified M step is proposed based on gradient search techniques.

Acknowledgements

The authors would like to thank Mr. John Buck and Mr. Kambiz Zangi for implementing and verifying the proposed algorithms.

The work reported here was supported in part by the Defense Advanced Research Projects Agency monitored by the Office of Naval Research under Grant No. N00014-89-J-1489 at M.I.T., in part by Lockheed/Sanders Inc. at M.I.T., and in part by ONR Grant No. N00014-90-J-1109 at the Woods Hole Oceanographic Institution.

Contents

I Introduction

Problems of signal enhancement and noise cancellation have been of considerable interest for many years. In one class of such problems, there may or may not be a signal present. With no signal present, the objective is to eliminate noise (or, more generally, an unwanted signal). This may arise, for example, in noisy factory environments, in quieting of active acoustic signatures of ships, etc. In other problems, there is assumed to be a desired signal present, and the objective is to reduce or eliminate background noise or some other unwanted signal, i.e., to enhance the desired signal. A well-studied example is the enhancement of speech in the presence of noise or competing speakers. In either of these contexts, when the unwanted component is estimated in some way and used to generate a cancelling signal, the process is typically referred to as noise cancellation. The cancelling signal may be "inserted", i.e., added to the total signal acoustically or electrically in real time or through postprocessing, but in this specific class of problems, it is assumed that a cancelling signal is generated which is then added to the total signal. A somewhat different but clearly overlapping class of problems is that of signal enhancement in which it is explicitly assumed that a desired signal is present. Signal enhancement may then be accomplished explicitly by noise cancellation or by estimating and generating the desired signal in some other way.

In this report, we develop a class of time-adaptive algorithms for signal enhancement, in which the desired signal is estimated from either single or multiple sensor measurements. The estimation procedure is based on the Estimate-Maximize (EM) algorithm for maximum likelihood parameter estimation [1]. Previously, the EM algorithm has been proposed for two-sensor signal enhancement [2], and an algorithm closely related to the EM algorithm has been proposed for single-sensor signal enhancement [3]. These algorithms are based on the use of a non-causal Wiener filter, implemented iteratively in the frequency domain on consecutive data blocks. However, the approach in [2] and [3] implies the use of a sliding window over which the signal and noise are assumed to be stationary. In order to avoid a computationally complex implementation of the Wiener filter, it is further assumed that the window is sufficiently long so that the Fourier transform coefficients of the received data at different frequencies are approximately uncorrelated. For most realistic situations, these are strongly competing requirements.

In this paper, we reformulate the EM algorithm for the single- and two-sensor problems directly in the time domain. This approach will enable us to avoid the competing assumptions indicated above. The resulting algorithms are similar in structure to those in [2] and [3] but, instead of the non-causal Wiener filter, we employ the Kalman smoother. These time domain algorithms are then converted to sequential/adaptive algorithms by replacing the Kalman smoother with a Kalman filter, and in place of successive iterations on each block, the algorithm moves sequentially through the data. In place of the windowing operation, exponential weighting is incorporated into the algorithm. The specific structure of the Kalman filtering equations is also exploited to simplify the computations involved.

In Section II, we review the EM algorithm for the convenience of the reader. In Section III, we consider the single-sensor problem, and in Section IV, we consider the two-sensor problem.

II The EM Algorithm

The Estimate-Maximize (EM) algorithm [1] is an iterative method for finding Maximum Likelihood (ML) parameter estimates. It works with the notion of "complete" data, and iterates between estimating the log-likelihood of the complete data using the observed (incomplete) data and the current parameter estimate (E-step), and maximizing the estimated log-likelihood function (M-step) to obtain the new parameter estimate.

More specifically, let $\boldsymbol{z}$ denote the observed data, with the probability density function (p.d.f.) $f_{\boldsymbol{Z}}(\boldsymbol{z};\boldsymbol{\theta})$, indexed by the vector of unknown parameters $\theta \in \Theta \subseteq \mathbb{R}^k$. The ML estimate $\widehat{\boldsymbol{\theta}}_{\mathrm{ML}}$ of $\boldsymbol{\theta}$ is defined by:

$$\widehat{\boldsymbol{\theta}}_{\mathrm{ML}} = \arg\max_{\theta\in\Theta} \log f_{\boldsymbol{Z}}(\boldsymbol{z};\boldsymbol{\theta}) \tag{1}$$

Let $\boldsymbol{y}$ denote the complete data, related to the observed (incomplete) data $\boldsymbol{z}$ by

$$H(\boldsymbol{y}) = \boldsymbol{z} \tag{2}$$

where $H(\cdot)$ is a non-invertible (many-to-one) transformation. With $f_{\boldsymbol{Y}}(\boldsymbol{y};\boldsymbol{\theta})$ denoting the p.d.f. of $\boldsymbol{y}$, and $f_{\boldsymbol{Y}|\boldsymbol{Z}}(\boldsymbol{y}|\boldsymbol{z};\boldsymbol{\theta})$ denoting the conditional p.d.f. of $\boldsymbol{y}$ given $\boldsymbol{z}$,

$$f_{\boldsymbol{Y}}(\boldsymbol{y};\boldsymbol{\theta}) = f_{\boldsymbol{Z}}(\boldsymbol{z};\boldsymbol{\theta}) f_{\boldsymbol{Y}|\boldsymbol{Z}}(\boldsymbol{y}|\boldsymbol{z};\boldsymbol{\theta}) \qquad \forall H(\boldsymbol{y}) = \boldsymbol{z} \tag{3}$$

Equivalently,

$$\log f_{\boldsymbol{Z}}(\boldsymbol{z};\boldsymbol{\theta}) = \log f_{\boldsymbol{Y}}(\boldsymbol{y};\boldsymbol{\theta}) - \log f_{\boldsymbol{Y}|\boldsymbol{Z}}(\boldsymbol{y}|\boldsymbol{z};\boldsymbol{\theta}) \tag{4}$$

Taking the conditional expectation given $\boldsymbol{z}$ at a parameter value $\boldsymbol{\theta}'$ (that is, multiplying both sides of (4) by $f_{\boldsymbol{Y}|\boldsymbol{Z}}(\boldsymbol{y}|\boldsymbol{z};\boldsymbol{\theta}')$ and integrating over $\boldsymbol{y}$,

$$f_{\boldsymbol{Z}}(\boldsymbol{z};\boldsymbol{\theta}) = E_{\boldsymbol{\theta}'}\{\log f_{\boldsymbol{Y}}(\boldsymbol{y};\boldsymbol{\theta})|\boldsymbol{z}\} - E_{\boldsymbol{\theta}'}\{\log f_{\boldsymbol{Y}|\boldsymbol{Z}}(\boldsymbol{y}|\boldsymbol{z};\boldsymbol{\theta})|\boldsymbol{z}\} \tag{5}$$

where $E_{\boldsymbol{\theta}'}\{\cdot|\boldsymbol{z}\}$ denotes the conditional expectation given $\boldsymbol{z}$ computed using the parameter value $\boldsymbol{\theta}'$.

For convenience we define

$$\begin{aligned} Q(\boldsymbol{\theta},\boldsymbol{\theta}') &= E_{\boldsymbol{\theta}'}\{\log f_{\boldsymbol{Y}}(\boldsymbol{y};\boldsymbol{\theta})|\boldsymbol{z}\} & (6)\\ P(\boldsymbol{\theta},\boldsymbol{\theta}') &= E_{\boldsymbol{\theta}'}\{\log f_{\boldsymbol{Y}|\boldsymbol{Z}}(\boldsymbol{y}|\boldsymbol{z};\boldsymbol{\theta})|\boldsymbol{z}\} & (7) \end{aligned}$$

So that Eq. (5) becomes

$$\log f_{\boldsymbol{Z}}(\boldsymbol{z};\boldsymbol{\theta}) = Q(\boldsymbol{\theta},\boldsymbol{\theta}') - P(\boldsymbol{\theta},\boldsymbol{\theta}') \tag{8}$$

By Jensen's inequality[1]

$$P(\boldsymbol{\theta},\boldsymbol{\theta}') \le P(\boldsymbol{\theta}',\boldsymbol{\theta}') \tag{9}$$

[1]Jensen's inequality asserts that for any pair of p.d.f.'s $f(\boldsymbol{x})$ and $g(\boldsymbol{x})$ defined over the probability space Ω of points $\boldsymbol{x}$,

$$\int_{\Omega} f(\boldsymbol{x}) \log g(\boldsymbol{x})\, d\boldsymbol{x} \le \int_{\Omega} f(\boldsymbol{x}) \log f(\boldsymbol{x})\, d\boldsymbol{x}$$

Therefore,

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}') > Q(\boldsymbol{\theta}', \boldsymbol{\theta}') \qquad \text{implies} \qquad \log f_{\boldsymbol{Z}}(\boldsymbol{z}; \boldsymbol{\theta}) > \log f_{\boldsymbol{Z}}(\boldsymbol{z}; \boldsymbol{\theta}') \tag{10}$$

The relation in (10) forms the basis for the EM algorithm. Denote by $\boldsymbol{\theta}^{(\ell)}$ the estimate of $\boldsymbol{\theta}$ after ℓ iterations of the algorithm. Then, the next iteration cycle is specified in two steps as follows:

E-step: Compute

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(\ell)}) = E_{\boldsymbol{\theta}^{(\ell)}}\{\log f_{\boldsymbol{Y}}(\boldsymbol{y}; \boldsymbol{\theta}) | \boldsymbol{z}\} \tag{11}$$

M-step:

$$\max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(\ell)}) \rightarrow \boldsymbol{\theta}^{(\ell+1)} \tag{12}$$

If $Q(\boldsymbol{\theta}, \boldsymbol{\theta}')$ is continuous in both $\boldsymbol{\theta}$ and $\boldsymbol{\theta}'$, the algorithm converges to a stationary point of the observed log-likelihood $\log f_{\boldsymbol{Z}}(\boldsymbol{z}; \boldsymbol{\theta})$, where the maximization in (12) ensures that each iteration cycle increases the likelihood of the estimated parameters. Specifically, as in all "hill climbing" algorithms, the stationary point may not be the global maximum, and thus several starting points or an initial grid search may be needed.

We note that the transformation $H(\cdot)$ is not uniquely defined. Specifically, there are many complete data specifications $\boldsymbol{y}$ that will generate an observed $\boldsymbol{z}$. The final point of convergence of the EM algorithm is essentially independent of the complete data specification. However, the choice of $\boldsymbol{y}$ may strongly affect the rate of convergence of the algorithm, and the computations involved.

As we shall see, for the set of problems of interest to us, there is a natural choice of complete data, leading to a simple intuitive algorithm for estimating the unknown signal and noise parameters. As a by-product of the algorithm, we also obtain an estimate of the desired (speech) signal. For the purpose of signal enhancement, it is the signal estimate that we are eventually interested in.

III Single-Sensor Signal Enhancement

3.1 Time-Domain EM Algorithm

We assume that the signal measured as the sensor output is of the form:

$$z(t) = s(t) + \epsilon(t) \tag{13}$$

where $\epsilon(t)$ is the measurement noise modelled as a zero-mean white Gaussian process with spectral level of $E\{\epsilon^2(t)\} = g_\epsilon$, and $s(t)$ is the desired signal modelled as an autoregressive (AR), or all-pole, process of order p:

$$s(t) = -\sum_{k=1}^{p} \alpha_k s(t-k) + \sqrt{g_s}\, u(t) \tag{14}$$

where $u(t)$ is a normalized (zero mean unit variance) white Gaussian process. We assume that $\epsilon(t)$ and $u(t)$ are mutually uncorrelated.

The vector $\boldsymbol{\theta}$ of unknown parameters associated with this model is:

$$\boldsymbol{\theta} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_p \\ g_s \\ g_\epsilon \end{bmatrix} \tag{15}$$

Given the observed data $z(t) \quad t = 1, 2, \ldots N$, we want to find the best possible estimate of the desired signal $s(t)$. Under the critereon of minimizing the mean square error (m.s.e.), the optimal signal estimate is obtained by performing the conditional expectation of $s(t)$ given the observed data. However, this conditional expectation requires prior knowledge of $\boldsymbol{\theta}$. Since $\boldsymbol{\theta}$ is unknown, one must deal with a more complicated problem of joint signal and parameter estimation.

The basic approach is to apply the EM algorithm, where the complete data is defined as the samples of the observed signal $z(t)$ together with the samples of the desired signal $s(t)$. This complete data specification will lead us to a simple iterative algorithm for extracting the ML parameter estimates. Furthermore, as a by-product of the algorithm, we will obtain the desired signal estimate which we are ultimately interested in.

More specifically, let

$$\boldsymbol{z} = \{z(t) : t = 1, 2, \ldots N\} \tag{16}$$

denote the "observed" data, let

$$\boldsymbol{s} = \{s(t) \ : \ t = -p+1, -p+2, \ldots N\} \tag{17}$$

be a collection of signal samples, and let

$$\boldsymbol{y} = \begin{bmatrix} \boldsymbol{z} \\ \boldsymbol{s} \end{bmatrix} \tag{18}$$

denote the complete data. Invoking Bayes's rule,

$$f_{\boldsymbol{Y}}(\boldsymbol{y}; \boldsymbol{\theta}) = f_{\boldsymbol{S}}(\boldsymbol{s}; \boldsymbol{\theta}) \cdot f_{\boldsymbol{Z}|\boldsymbol{S}}(\boldsymbol{z}|\boldsymbol{s}; \boldsymbol{\theta}) \tag{19}$$

and equivalently,

$$\log f_{\boldsymbol{Y}}(\boldsymbol{y}; \boldsymbol{\theta}) = \log f_{\boldsymbol{S}}(\boldsymbol{s}; \boldsymbol{\theta}) + \log f_{\boldsymbol{Z}|\boldsymbol{S}}(\boldsymbol{z}|\boldsymbol{s}; \boldsymbol{\theta}) \tag{20}$$

From (14),

$$\log f_{\boldsymbol{S}}(\boldsymbol{s}; \boldsymbol{\theta}) = \log f(\boldsymbol{s}_{p-1}(0)) - \frac{N}{2} \log 2\pi g_s - \frac{1}{2g_s} \sum_{t=1}^{N} [s(t) + \boldsymbol{\alpha}^{\top} \boldsymbol{s}_{p-1}(t-1)]^2 \tag{21}$$

where we define:

$$\boldsymbol{s}_p(t) = \begin{bmatrix} s(t-p) \\ s(t-p+1) \\ \vdots \\ s(t) \end{bmatrix} \tag{22}$$

$$\boldsymbol{\alpha} = \begin{bmatrix} \alpha_p \\ \alpha_{p-1} \\ \vdots \\ \alpha_1 \end{bmatrix} \tag{23}$$

and $f(\boldsymbol{s}_{p-1}(0))$ denotes the p.d.f. of $\boldsymbol{s}_{p-1}(0)$.

From (13)

$$\log f_{\boldsymbol{Z}|\boldsymbol{S}}(\boldsymbol{z}|\boldsymbol{s};\boldsymbol{\theta}) = -\frac{N}{2}\log 2\pi g_\epsilon - \frac{1}{2g_\epsilon}\sum_{t=1}^{N}[z(t)-s(t)]^2 \tag{24}$$

Substituting (21) and (24) into (20), and assuming that $\log f(\boldsymbol{s}_{p-1}(0))$ is pre-specified, or that $N \gg p$ so that its contribution is negligible,

$$\begin{aligned} \log f_{\boldsymbol{Y}}(\boldsymbol{y};\boldsymbol{\theta}) &= C - \frac{N}{2}\log g_s - \frac{1}{2g_s}\sum_{t=1}^{N}[s(t) + \boldsymbol{\alpha}^\top \boldsymbol{s}_{p-1}(t-1)]^2 \\ &\quad - \frac{N}{2}\log g_\epsilon - \frac{1}{2g_\epsilon}\sum_{t=1}^{N}[z(t)-s(t)]^2 \end{aligned} \tag{25}$$

where C is a constant independent of $\boldsymbol{\theta}$. Taking the conditional expectation given $\boldsymbol{z}$ at a parameter value $\boldsymbol{\theta}^{(\ell)}$

$$\begin{aligned} Q(\boldsymbol{\theta},\boldsymbol{\theta}^{(\ell)}) &= E_{\boldsymbol{\theta}^{(\ell)}}\{\log f_{\boldsymbol{Y}}(\boldsymbol{y};\boldsymbol{\theta})|\boldsymbol{z}\} \\ &= -\frac{N}{2}\log g_s - \frac{1}{2g_s}\sum_{t=1}^{N}\left[s^2(t) + 2\boldsymbol{\alpha}^\top \widehat{\boldsymbol{s}_{p-1}(t-1)s(t)}^{(\ell)} + \boldsymbol{\alpha}^\top \widehat{\boldsymbol{s}_{p-1}(t-1)\boldsymbol{s}_{p-1}^\top(t-1)}^{(\ell)}\boldsymbol{\alpha}^\top\right] \\ &\quad - \frac{N}{2}\log g_\epsilon - \frac{1}{2g_\epsilon}\sum_{t=1}^{N}\left[z^2(t) - 2z(t)\widehat{s}^{(\ell)}(t) + \widehat{s^2}^{(\ell)}(t)\right] \end{aligned} \tag{26}$$

where we define:

$$\widehat{(\cdot)}^{(\ell)} \triangleq E_{\boldsymbol{\theta}^{(\ell)}}\{\cdot|\boldsymbol{z}\} \tag{27}$$

Thus, the computation of $Q(\boldsymbol{\theta},\boldsymbol{\theta}^{(\ell)})$ (E-step) only requires te computation of the indicated conditional expectations. Furthermore, the maximization of $Q(\boldsymbol{\theta},\boldsymbol{\theta}^{(\ell)})$ with respect to $\boldsymbol{\theta}$ (M-step) can be solved analytically. Altogether, we obtain the following algorithm:

E-step: For $t = 1, 2, \ldots, N$ compute:

$$\widehat{\boldsymbol{s}}_p^{(\ell)}(t) = \begin{bmatrix} \widehat{\boldsymbol{s}}_{p-1}^{(\ell)}(t-1) \\ \hline \widehat{s}^{(\ell)}(t) \end{bmatrix} \tag{28}$$

$$\widehat{s_p(t)s_p^\top(t)}^{(\ell)} = \left[\begin{array}{c|c} \widehat{s_{p-1}(t-1)s_{p-1}^\top(t-1)}^{(\ell)} & \widehat{s_{p-1}(t-1)s(t)}^{(\ell)} \\ \hline \widehat{s(t)s_{p-1}^\top(t-1)}^{(\ell)} & \widehat{s^2(t)}^{(\ell)} \end{array}\right] \tag{29}$$

M-step:

$$\hat{\alpha}^{(\ell+1)} = -\left[\sum_{t=1}^{N} \widehat{s_{p-1}(t-1)s_{p-1}^\top(t-1)}^{(\ell)}\right]^{-1} \sum_{t=1}^{N} \widehat{s_{p-1}(t-1)s(t)}^{(\ell)} \tag{30}$$

$$\hat{g}_s^{(\ell+1)} = \frac{1}{N}\sum_{t=1}^{N}\left[\widehat{s^2(t)}^{(\ell)} + \hat{\alpha}^{(\ell+1)\top} \cdot \widehat{s_{p-1}(t-1)s(t)}^{(\ell)}\right] \tag{31}$$

$$\hat{g}_\epsilon^{(\ell+1)} = \frac{1}{N}\sum_{t=1}^{N}\left[z^2(t) - 2z(t)\hat{s}^{(\ell)}(t) + \widehat{s^2(t)}^{(\ell)}\right] \tag{32}$$

It is important to note that although the EM algorithm is directed at finding the ML parameter estimates, as a by-product of the algorithm we also obtain the estimate $\hat{s}^{(\ell)}(t)$ of the signal $s(t)$. For the purpose of signal enhancement, it is the signal estimate that we are eventually interested in.

Note the intuitive form of the parameter estimation update. Equation (30) is the Yule-Walker solution for the AR parameters, where the first and second order statistics of the signal are substituted by their current estimate, and the gain parameters in (31) and (32) are re-estimated as the sample average of the corresponding power levels. The algorithm iterates using the current parameter estimate to improve the estimate of the signal (and its sufficient statistics) and then uses this signal estimate to improve the next parameter estimate.

The computation of the conditional expectations in (28) and (29) can be carried out using the Kalman smoothing equations. To do that, we represent Eqs. (13) and (14) in state-space form:

$$s_p(t) = \Phi s_p(t-1) + g u(t) \tag{33}$$

$$z(t) = h^\top s_p(t) + \epsilon(t) \tag{34}$$

where the state vector $s_p(t)$ is the $(p+1)\times 1$ vector of signal samples defined in (22), Φ is the $(p+1)\times(p+1)$ matrix

$$\Phi = \begin{bmatrix} 0 & 1 & 0 & \cdots & \cdots & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & & & \vdots \\ \vdots & & \ddots & \ddots & \ddots & & \vdots \\ \vdots & & & \ddots & \ddots & \ddots & \vdots \\ \vdots & & & & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & \cdots & \cdots & 0 & 1 \\ 0 & -\alpha_p & \cdots & \cdots & \cdots & \cdots & -\alpha_1 \end{bmatrix} \tag{35}$$

$\boldsymbol{g}$ is the $(p+1)\times 1$ vector

$$\boldsymbol{g}^{\top} = [\ 0\ \ldots\ 0\ \sqrt{g_s}\ ] \tag{36}$$

and $\boldsymbol{h}$ is the $(p+1)\times 1$ vector

$$\boldsymbol{h}^{\top} = [\ 0\ \ldots\ 0\ 1\ ] \tag{37}$$

In this setting, $\boldsymbol{s}_p(t)$ is the state vector. Now, define

$$\boldsymbol{\mu}_{t|n}^{(\ell)} = E_{\boldsymbol{\theta}^{(\ell)}}\{\boldsymbol{s}_p(t)|\boldsymbol{z}(1),\boldsymbol{z}(2),\ldots,\boldsymbol{z}(n)\} \tag{38}$$

$$P_{t|n}^{(\ell)} = E_{\boldsymbol{\theta}^{(\ell)}}\{[\boldsymbol{s}_p(t)-\boldsymbol{\mu}_{t|n}^{(\ell)}][\boldsymbol{s}_p(t)-\boldsymbol{\mu}_{t|n}^{(\ell)}]^{\top}|\boldsymbol{z}(1),\boldsymbol{z}(2),\ldots,\boldsymbol{z}(n)\} \tag{39}$$

Clearly, the conditional expectations in (28) and (29) are given by:

$$\widehat{\boldsymbol{s}}_p^{(\ell)}(t) = \boldsymbol{\mu}_{t|N}^{(\ell)} \tag{40}$$

$$\widehat{\boldsymbol{s}_p(t)\boldsymbol{s}_p^{\top}(t)}^{(\ell)} = \boldsymbol{\mu}_{t|N}\boldsymbol{\mu}_{t|N}^{\top} + P_{t|N}^{(\ell)} \tag{41}$$

We also denote by $\Phi^{(\ell)}$, $\boldsymbol{g}^{(\ell)}$, and $\boldsymbol{h}^{(\ell)}$ the matrices Φ, $\boldsymbol{g}$, and $\boldsymbol{h}$ computed at the current parameter estimate $\boldsymbol{\theta}=\boldsymbol{\theta}^{(\ell)}$. Then, the Kalman smoothing equations compute $\boldsymbol{\mu}_{t|N}^{(\ell)}$ and $P_{t|N}^{(\ell)}$ in three stages as follows:

Propagation Equations

For $t = 1,2,\ldots,N$

$$\boldsymbol{\mu}_{t|t-1}^{(\ell)} = \Phi^{(\ell)}\boldsymbol{\mu}_{t-1|t-1}^{(\ell)} \qquad ,\boldsymbol{\mu}_{0|0}^{(\ell)} \tag{42}$$

$$P_{t|t-1}^{(\ell)} = \Phi^{(\ell)}P_{t-1|t-1}^{(\ell)}\Phi^{(\ell)^{\top}} + \boldsymbol{g}^{(\ell)}\boldsymbol{g}^{(\ell)^{\top}} \qquad ,P_{0|0}^{(\ell)} \tag{43}$$

where the initial conditions $\boldsymbol{\mu}_{0|0}^{(\ell)}$ and $P_{0|0}^{(\ell)}$ are the mean and the covariance matrix of the initial state $\boldsymbol{s}_p(0)$, computed using the current estimate $\boldsymbol{\theta}=\boldsymbol{\theta}^{(\ell)}$.

Updating Equations

For $t = 1,2,\ldots,N$

$$\boldsymbol{\mu}_{t|t}^{(\ell)} = \boldsymbol{\mu}_{t|t-1}^{(\ell)} + \boldsymbol{k}_t^{(\ell)}\left[z(t) - \boldsymbol{h}^{\top}\boldsymbol{\mu}_{t|t-1}\right] \tag{44}$$

$$P_{t|t}^{(\ell)} = \left(I - \boldsymbol{k}_t^{(\ell)}\boldsymbol{h}^{\top}\right)P_{t|t-1}^{(\ell)} \tag{45}$$

where

$$\boldsymbol{k}_t^{(\ell)} = \frac{1}{\boldsymbol{h}^{\top}P_{t|t-1}^{(\ell)}\boldsymbol{h} + g_{\epsilon}^{(\ell)}}\cdot P_{t|t-1}^{(\ell)}\boldsymbol{h} \tag{46}$$

Smoothing Equations

For $t = N, N-1,\ldots,1$

$$\boldsymbol{\mu}_{t-1|N}^{(\ell)} = \boldsymbol{\mu}_{t-1|t-1}^{(\ell)} + S_{t-1}^{(\ell)}\left[\boldsymbol{\mu}_{t|N}^{(\ell)} - \Phi^{(\ell)}\boldsymbol{\mu}_{t-1|t-1}^{(\ell)}\right] \tag{47}$$

$$P_{t-1|N}^{(\ell)} = P_{t-1|t-1}^{(\ell)} + S_{t-1}^{(\ell)}\left[P_{t|N}^{(\ell)} - P_{t|t-1}^{(\ell)}\right]S_{t-1}^{(\ell)^{\top}} \tag{48}$$

where

$$S_{t-1}^{(\ell)} = P_{t-1|t-1}^{(\ell)} \Phi^{(\ell)^{\top}} P_{t|t-1}^{(\ell)^{-1}} \tag{49}$$

The values $\mu_{t|N}^{(\ell)}$ and $P_{t|N}^{(\ell)}$, $t = 1, 2, \ldots, N$ resulting from the recursions in (47) and (48) are then substituted into (40) and (41) to yield the conditional expectations needed in the E-step of the algorithm.

3.2 Sequential/Adaptive Algorithms

The algorithm thus far was developed under the assumption that the signal and the noise are stationary during the observation period. In practice, this is an unrealistic assumption. A typical speech signal is a non-stationary process with time-varying parameters, and the statistical properties (spectral level) of the additive noise may also be changing in time. Basically, we have a situation in which the observed signal depends on a time-varying parameter vector $\boldsymbol{\theta}(t)$, and we want to convert the iterative batch EM algorithm into an adaptive algorithm that tracks the varying parameters.

The conservative approach is to assume that the observed signals are stationary over a fixed time window, and apply the proposed EM algorithm on consecutive data blocks. An alternative approach is suggested by the structure of the EM algorithm. As it stands, in the E-step of the algorithm we use the Kalman smoother to generate the state (signal) estimate. The idea is to replace the Kalman smoother with the Kalman filter that only involves the propogation equations (42) (43) followed by the updating equations (44) (45). In this way, the signal at a particular time instant t is estimated using only the past and present data samples, and we have removed the smoothing equations (47) (48) that are computationally the most expensive.

To obtain a recursive/sequential algorithm, we further replace the iteration index by the time index — which is a standard procedure in stochastic approximation. With this additional modification, the estimate of the state and its error covariance are generated using a single Kalman filter whose matrices are continuously updated using the current parameter estimates (instead of applying the Kalman filter iteratively).

More specifically, denote by $\widehat{\boldsymbol{s}}_p(t|t)$ and $\widehat{\boldsymbol{s}_p(t|t)\boldsymbol{s}_p^{\top}(t|t)}$ the estimate of $\boldsymbol{s}_p(t)$ and $\boldsymbol{s}_p(t)\boldsymbol{s}_p^{\top}(t)$ based on the observed data to time t and the current parameter estimate. Also, define

$$\mu_{t|t} \triangleq \widehat{\boldsymbol{s}}_p(t|t) \tag{50}$$

and

$$P_{t|t} \triangleq \widehat{\boldsymbol{s}_p(t|t)\boldsymbol{s}_p^{\top}(t|t)} \tag{51}$$

and $\widehat{\boldsymbol{\theta}}(t)$ as the current parameter estimate, i.e.

$$\widehat{\boldsymbol{\theta}}(t) = \begin{bmatrix} \alpha(t) \\ \widehat{g}_s(t) \\ \widehat{g}_\epsilon(t) \end{bmatrix} \tag{52}$$

Also, denote by $\widehat{\Phi}_t$ and $\widehat{\boldsymbol{g}}_t$ the matrices Φ and $\boldsymbol{g}$ computed at $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}(t)$:

$$\widehat{\Phi}_t = \left[\begin{array}{c|c} \mathbf{0} & \overset{\longleftarrow p \longrightarrow}{I} \\ \hline 0 & -\widehat{\boldsymbol{\alpha}}^\top(t) \end{array}\right] \updownarrow p \tag{53}$$

$$\widehat{\boldsymbol{g}}_t = \left[\begin{array}{c} \mathbf{0} \\ \hline \sqrt{\widehat{g}_s(t)} \end{array}\right] \updownarrow p \tag{54}$$

Then, $\boldsymbol{\mu}_{t|t}$ and $P_{t|t}$ are computed recursively as follows:

Propagation Equations

$$\boldsymbol{\mu}_{t|t-1} = \widehat{\Phi}_t \boldsymbol{\mu}_{t-1|t-1} \qquad , \boldsymbol{\mu}_{0|0} \tag{55}$$

$$P_{t|t-1} = \widehat{\Phi}_t P_{t-1|t-1} \widehat{\Phi}_t^\top + \widehat{\boldsymbol{g}}_t \widehat{\boldsymbol{g}}_t^\top \qquad , P_{0|0} \tag{56}$$

Updating Equations

$$\boldsymbol{\mu}_{t|t} = \boldsymbol{\mu}_{t|t-1} + \widehat{\boldsymbol{k}}_t [z(t) - \boldsymbol{h}^\top \boldsymbol{\mu}_{t|t-1}] \tag{57}$$

$$P_{t|t} = \left(I - \widehat{\boldsymbol{k}}_t \boldsymbol{h}^\top\right) P_{t|t-1} \tag{58}$$

where

$$\widehat{\boldsymbol{k}}_t = \frac{1}{\boldsymbol{h}^\top P_{t|t-1} \boldsymbol{h} + \widehat{g}_\epsilon(t)} P_{t|t-1} \boldsymbol{h} \tag{59}$$

These recursions can be simplified by exploiting the structure of $\widehat{\Phi}_t$ and $\widehat{\boldsymbol{g}}_t$. To this end, we observe that $\widehat{\boldsymbol{s}}_{p-1}(t-1|t-1)$ is the lower $p \times 1$ sub-vector or $\boldsymbol{\mu}_{t-1|t-1}$:

$$\boldsymbol{\mu}_{t-1|t-1} = \left[\begin{array}{c} \\ \hline \widehat{\boldsymbol{s}}_{p-1}(t-1|t-1) \end{array}\right] \updownarrow p \tag{60}$$

Denote by Λ_{t-1} the covariance of $\widehat{\boldsymbol{s}}_{p-1}(t-1|t-1)$, that is the lower $p \times p$ sub-matrix of $P_{t-1|t-1}$:

$$P_{t-1|t-1} = \left[\begin{array}{c|c} & \\ \hline & \Lambda_{t-1} \end{array}\right] \tag{61}$$

Substituting (53) and (60) into (55) and performing the indicated matrix multiplication, we obtain:

$$\boldsymbol{\mu}_{t|t-1} = \left[\begin{array}{c} \widehat{\boldsymbol{s}}_{p-1}(t-1|t-1) \\ \hline -\widehat{\boldsymbol{\alpha}}^\top(t) \widehat{\boldsymbol{s}}_{p-1}(t-1|t-1) \end{array}\right] \tag{62}$$

Using (53), (54), and (60) in (56), we obtain:

$$P_{t|t-1} = \left[\begin{array}{c|c} \Lambda_{t-1} & -\Lambda_{t-1} \widehat{\boldsymbol{\alpha}}(t) \\ \hline -\widehat{\boldsymbol{\alpha}}^\top(t) \Lambda_{t-1} & \widehat{\boldsymbol{\alpha}}^\top(t) \Lambda_{t-1} \widehat{\boldsymbol{\alpha}}(t) + \widehat{g}_s(t) \end{array}\right] \tag{63}$$

Substituting (37) and (63) into (59) and performing the indicated matrix manipulations,

$$\hat{\boldsymbol{k}}_t = \frac{1}{\eta(t)}\left[\frac{-\Lambda_{t-1}\hat{\boldsymbol{\alpha}}(t)}{\hat{\boldsymbol{\alpha}}^\top(t)\Lambda_{t-1}\hat{\boldsymbol{\alpha}}(t)+\hat{g}_s(t)}\right] \tag{64}$$

where

$$\eta(t) \triangleq \hat{\boldsymbol{\alpha}}^\top(t)\Lambda_{t-1}\hat{\boldsymbol{\alpha}}(t)+\hat{g}_s(t)+\hat{g}_\epsilon(t) \tag{65}$$

Using (37) and (62)–(64) in (57) and (58), we finally obtain:

$$\begin{aligned}\boldsymbol{\mu}_{t|t} &= \left[\frac{\hat{\boldsymbol{s}}_{p-1}(t-1|t-1)}{-\hat{\boldsymbol{\alpha}}^\top(t)\hat{\boldsymbol{s}}_{p-1}(t-1|t-1)}\right] \\ &\quad + \frac{z(t)+\hat{\boldsymbol{\alpha}}^\top(t)\hat{\boldsymbol{s}}_{p-1}(t-1|t-1)}{\eta(t)}\left[\frac{-\Lambda_{t-1}\hat{\boldsymbol{\alpha}}(t)}{\hat{\boldsymbol{\alpha}}^\top(t)\Lambda_{t-1}\hat{\boldsymbol{\alpha}}(t)+\hat{g}_s(t)}\right]\end{aligned} \tag{66}$$

$$P_{t|t} = \left[\begin{array}{c|c} \Lambda_{t-1}-\frac{1}{\eta(t)}\Lambda_{t-1}\hat{\boldsymbol{\alpha}}(t)\hat{\boldsymbol{\alpha}}^\top(t)\Lambda_{t-1} & -\frac{\hat{g}_\epsilon(t)}{\eta(t)}\Lambda_{t-1}^\top\hat{\boldsymbol{\alpha}}(t) \\ \hline -\frac{\hat{g}_\epsilon(t)}{\eta(t)}\hat{\boldsymbol{\alpha}}^\top(t)\Lambda_{t-1} & \frac{\hat{g}_\epsilon(t)}{\eta(t)}[\hat{\boldsymbol{\alpha}}^\top(t)\Lambda_{t-1}\hat{\boldsymbol{\alpha}}(t)+\hat{g}_s(t)] \end{array}\right] \tag{67}$$

To update the parameter estimates, we modify Eqs. (30)–(32) in a similar fashion, by replacing the iteration index (ℓ) by the time index (t), using the data only up to the current time t:

$$\hat{\boldsymbol{\alpha}}(t+1) = -\left[\sum_{\tau=1}^{t}\widehat{\boldsymbol{s}_{p-1}(\tau-1)\boldsymbol{s}_{p-1}^\top(\tau-1)}\right]^{-1}\sum_{\tau=1}^{t}\widehat{\boldsymbol{s}_{p-1}(\tau-1)s(\tau)} \tag{68}$$

$$\hat{g}_s(t+1) = \frac{1}{t}\sum_{\tau=1}^{t}\left[\widehat{s^2}(\tau|\tau)+\hat{\boldsymbol{\alpha}}^\top(t+1)\widehat{\boldsymbol{s}_{p-1}(\tau-1|\tau)s(\tau|\tau)}\right] \tag{69}$$

$$\hat{g}_\epsilon(t+1) = \frac{1}{t}\sum_{\tau=1}^{t}[z^2(\tau)-2z(\tau)\hat{s}(\tau|\tau)+\widehat{s^2}(\tau|\tau)] \tag{70}$$

We note that $\hat{\boldsymbol{s}}_p(\tau)$ and $\widehat{\boldsymbol{s}_p(\tau)\boldsymbol{s}_p^\top(\tau)}$ are generated using the data observed to time τ. Therefore, under stationary conditions, the quality of these estimates improves as τ gets closer to the current time instant t. Furthermore, in a non-stationary environment, when the parameters may vary in time, current data and the current signal estimates provide even more information concerning the current parameter values than past data samples. This suggests introducing an exponential weighting into (68)–(70):

$$\hat{\boldsymbol{\alpha}}(t+1) = -\left[\sum_{\tau=1}^{t}\gamma_s^{t-\tau}\widehat{\boldsymbol{s}_{p-1}(\tau-1|\tau)\boldsymbol{s}_{p-1}^\top(\tau-1|\tau)}\right]^{-1}\sum_{\tau=1}^{t}\gamma_s^{t-\tau}\widehat{\boldsymbol{s}_{p-1}(\tau-1|\tau)s(\tau|\tau)} \tag{71}$$

$$\hat{g}_s(t+1) = \frac{1}{\sum_{\tau=1}^{t}\gamma_s^{t-\tau}}\sum_{\tau=1}^{t}\gamma_s^{t-\tau}\left[\widehat{s^2}(\tau|\tau)+\hat{\boldsymbol{\alpha}}^\top(t+1)\widehat{\boldsymbol{s}_{p-1}(\tau-1|\tau)s(\tau|\tau)}\right] \tag{72}$$

$$\hat{g}_\epsilon(t+1) = \frac{1}{\sum_{\tau=1}^{t}\gamma_\epsilon^{t-\tau}}\sum_{\tau=1}^{t}\gamma_\epsilon^{t-\tau}\left[z^2(\tau)-2z(\tau)\hat{s}(\tau|\tau)+\widehat{s^2}(\tau|\tau)\right] \tag{73}$$

where γ_s and γ_ϵ are exponential "forgetting" factors, which are real numbers between 0 and 1. If we choose these factors to be close to 1, then (71)–(73) closely approach (68)–(70). If we choose the forgetting factors to be smaller than 1, then the new vector parameter estimates $\widehat{\boldsymbol{\theta}}(t+1)$ will depend more heavily on the current data and the current signal estimate. Consequently, the next signal estimate at time $(t+1)$ will depend more heavily on the current parameter value, and we have an adaptive algorithm. For that reason, we shall utilize (71)–(73). An attractive feature of these equations is that they can be computed recursively in time.

In order to develop the recursive form of these equations, we define:

$$\begin{aligned} R_{11}(t) &= \sum_{\tau=1}^{t} \gamma_s^{t-\tau} \widehat{\boldsymbol{s}_{p-1}(\tau-1|\tau)\boldsymbol{s}_{p-1}^{\mathrm{T}}(\tau-1|\tau)} \\ &= \gamma_s R_{11}(t-1) + \widehat{\boldsymbol{s}_{p-1}(t-1|t)\boldsymbol{s}_{p-1}^{\mathrm{T}}(t-1|t)} \end{aligned} \tag{74}$$

$$\begin{aligned} R_{12}(t) &= \sum_{\tau=1}^{t} \gamma_s^{t-\tau} \widehat{\boldsymbol{s}_{p-1}(\tau-1|\tau)s(\tau|\tau)} \\ &= \gamma_s R_{12}(t-1) + \widehat{\boldsymbol{s}_{p-1}(t-1)s(t)} \end{aligned} \tag{75}$$

$$R_{22}(t) = \sum_{\tau=1}^{t} \gamma_s^{t-\tau} \widehat{s^2}(\tau|\tau) = \gamma_s R_{22}(t-1) + \widehat{s^2}(t|t) \tag{76}$$

$$\begin{aligned} Q_{22}(t) &= \sum_{\tau=1}^{t} \gamma_\epsilon^{t-\tau} \left[z^2(\tau) - 2z(\tau)\widehat{s}(\tau|\tau) + \widehat{s^2}(\tau|\tau) \right] \\ &= \gamma_\epsilon Q_{22}(t-1) + \left[z^2(t) - 2z(t)\widehat{s}(t|t) + \widehat{s^2}(t|t) \right] \end{aligned} \tag{77}$$

Then,

$$\begin{aligned} \widehat{\boldsymbol{\alpha}}(t+1) &= -R_{11}^{-1}(t)R_{12}(t) = -R_{11}^{-1}(t)\left[\gamma_s R_{12}(t-1) + \widehat{\boldsymbol{s}_{p-1}(t-1|t)s(t|t)}\right] \\ &= -R_{11}^{-1}(t)\left[-\gamma_s R_{11}(t-1)\widehat{\boldsymbol{\alpha}}(t) + \widehat{\boldsymbol{s}_{p-1}(t-1|t)s(t|t)}\right] \\ &= R_{11}^{-1}(t)\left\{\left[R_{11}(t) - \widehat{\boldsymbol{s}_{p-1}(t-1|t)\boldsymbol{s}_{p-1}^{\mathrm{T}}(t-1|t)}\right]\widehat{\boldsymbol{\alpha}}(t) - \widehat{\boldsymbol{s}_{p-1}(t-1|t)s(t|t)}\right\} \\ &= \widehat{\boldsymbol{\alpha}}(t) - R_{11}^{-1}(t)\left[\widehat{\boldsymbol{s}_{p-1}(t-1|t)s(t|t)} + \widehat{\boldsymbol{s}_{p-1}(t-1|t)\boldsymbol{s}_{p-1}^{\mathrm{T}}(t-1|t)}\widehat{\boldsymbol{\alpha}}(t)\right] \end{aligned} \tag{78}$$

Using (74)–(78), in (72) and (73), and observing that

$$\sum_{\tau=1}^{t} \gamma^{t-\tau} = \frac{1-\gamma^t}{1-\gamma} \qquad ,|\gamma| < 1$$

we obtain:

$$\widehat{g}_s(t+1) = \frac{1-\gamma_s}{1-{\gamma_s}^t}[R_{22}(t) + \widehat{\alpha}^\top(t+1)R_{12}(t)] \tag{79}$$

$$\widehat{g}_\epsilon(t+1) = \frac{1-\gamma_\epsilon}{1-{\gamma_\epsilon}^t} Q_{22}(t) \tag{80}$$

Equations (74)–(80) can be used for recursive updating of the vector parameter estimates. We note, in passing, that similar recursions can be developed for (68)–(70).

Equations (60), (61), (66) and (67), together with equations (74)–(80), specify the proposed sequential/adaptive algorithm for the joint signal and parameter estimation.

3.3 Comments

By definition (see (22) and (50)), the last component of $\boldsymbol{\mu}_{t|t}$ is the signal estimate at time t based on data to that time, and on the current parameter estimate $\widehat{\boldsymbol{\theta}}(t)$.

The first component of $\boldsymbol{\mu}_{t|t}$ is the signal estimate at time $(t-p)$ based on data to time t. Thus, if we are willing to accomodate a delay in the signal estimate, we may prefer to use the smoothed signal estimate in order to improve statistical stability.

The last component of $\boldsymbol{\mu}_{t|t-1}$ is the predicted value of $s(t)$ given the data to time $(t-1)$. From (62) it is given by $-\widehat{\alpha}^\top(t)\widehat{\mathbf{s}}_{p-1}(t-1|t-1)$. This may be useful, for example, in noise cancellation scenarios, where we want to predict the next value of the signal/noise for the purpose of active cancellation.

IV Two-sensor Signal Enhancement

4.1 Time Domain EM Algorithm

The basic system of interest consists of a desired (speech) signal source and a noise source both existing in the same acoustic environment, say a living room or an office. Ideally, we want to install two microphones in such a way that one measures mainly the speech signal while the other measures mainly the noise. Unfortunately, the signal and the noise are typically both coupled into each microphone by the acoustic field in this environment. The mathematical model for the received signals is given by:

$$z_1(t) = C\{s(t)\} + A\{w(t)\} + e_1(t) \tag{81}$$

$$z_2(t) = B\{s(t)\} + D\{w(t)\} + e_2(t) \tag{82}$$

where $s(t)$ denotes the desired (speech) signal, $w(t)$ denotes the noise source signal, and A, B, C, and D represent the acoustic transfer functions between the sources and the microphones.

The additional noises $e_1(t)$ and $e_2(t)$ are included to represent modelling errors, microphone and measurement noise, etc. We shall make the following assumptions:

(a) The desired signal $s(t)$ is modelled as an Auto-Regressive (AR) process of order p, satisfying the following difference equation:

$$s(t) = -\sum_{k=1}^{p} \alpha_k s(t-k) + \sqrt{g_s} \cdot u_s(t) \tag{83}$$

where $u_s(t)$ is a normalized (zero-mean unit variance) white Gaussian process.

(b) The noise signal $w(t)$ is modelled by:

$$w(t) = \sqrt{g_w} \cdot u_w(t) \tag{84}$$

where $u_w(t)$ is a normalized white Gaussian process.

(c) $e_1(t)$ and $e_2(t)$ are statistically independent zero-mean white Gaussian processes with spectral levels of g_1 and g_2, respectively. We assume that $u_s(t)$, $u_w(t)$, $e_1(t)$, and $e_2(t)$ are mutually independent.

(d) The acoustic transfer functions A, B, C, and D are constrained to be causal linear time-invariant finite impulse response (FIR) filters. Since the signal source is assumed to be located near the primary microphone, and the noise source is assumed to be located near the reference microphone, we assume that C and D are identity (all pass) filters.

Under these assumptions, equations (81) and (82) can be written in the form:

$$z_1(t) = s(t) + \sum_{k=0}^{q} a_k w(t-k) + e_1(t) \tag{85}$$

$$z_2(t) = \sum_{k=0}^{r} b_k s(t-k) + w(t) + e_2(t) \tag{86}$$

where the a_k's are the filter coefficients of $A(z)$, the b_k's are the filter coefficients of $B(z)$, and q and r are the respective filter orders.

We shall find it convenient to define:

$$\boldsymbol{s}_r(t) = \begin{bmatrix} s(t-r) \\ s(t-r+1) \\ \vdots \\ s(t) \end{bmatrix} \tag{87}$$

$$\boldsymbol{w}_q(t) = \begin{bmatrix} w(t-q) \\ w(t-q+1) \\ \vdots \\ w(t) \end{bmatrix} \tag{88}$$

$$\boldsymbol{\alpha} = \begin{bmatrix} \alpha_p \\ \alpha_{p-1} \\ \vdots \\ \alpha_1 \end{bmatrix} \tag{89}$$

$$\boldsymbol{a} = \begin{bmatrix} a_q \\ a_{q-1} \\ \vdots \\ a_0 \end{bmatrix} \tag{90}$$

and

$$\boldsymbol{b} = \begin{bmatrix} b_r \\ b_{r-1} \\ \vdots \\ b_0 \end{bmatrix} \tag{91}$$

Note that (87) and (89) are consistent with (22) and (23). With these definitions, (85) and (86) can be written in the form:

$$z_1(t) = s(t) + \boldsymbol{a}^\top \boldsymbol{w}_q(t) + e_1(t) \tag{92}$$

$$z_2(t) = \boldsymbol{b}^\top \boldsymbol{s}_r(t) + w(t) + e_2(t) \tag{93}$$

and (83) becomes:

$$s(t) = -\boldsymbol{a}^\top \boldsymbol{s}_{p-1}(t-1) + \sqrt{g_s} \cdot u_s(t) \tag{94}$$

Denote by $\boldsymbol{\theta}$ the vector of unknown parameters

$$\boldsymbol{\theta} = \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{a} \\ \boldsymbol{b} \\ g_s \\ g_w \\ g_1 \\ g_2 \end{bmatrix} \tag{95}$$

Given observed signals $z_1(t)$ and $z_2(t)$, $t = 1, 2, \ldots, N$, we want to estimate the desired signal $s(t)$ jointly with the unknown parameters $\boldsymbol{\theta}$. Again, the idea is to use the EM algorithm with the following complete data specification:

$$\boldsymbol{y} = \begin{bmatrix} z \\ s \\ w \end{bmatrix} \tag{96}$$

where $\boldsymbol{z}$ is the observed data vector:

$$\boldsymbol{z} = \{z_1(t), z_2(t) : t = 1, 2, \cdots, N\} \tag{97}$$

$\boldsymbol{s}$ is defined by

$$\boldsymbol{s} = \{s(t) : t = -r+1, -r+2, \cdots, N\} \tag{98}$$

and $\boldsymbol{w}$ is defined by:

$$\boldsymbol{w} = \{w(t) : t = -q+1, -q+2, \cdots, N\} \tag{99}$$

Typically, the order r of the transfer function B is much greater than the order p of the speech process. Therefore, the vectors $\boldsymbol{s}$ and $\boldsymbol{w}$ contain all of the signal and noise samples that affect the observed data $\boldsymbol{z}$.

Invoking Bayes' rule

$$\begin{aligned} f_{\boldsymbol{y}}(\boldsymbol{y};\boldsymbol{\theta}) &= f_{\boldsymbol{s},\boldsymbol{w}}(\boldsymbol{s},\boldsymbol{w};\boldsymbol{\theta}) \cdot f(\boldsymbol{z}|\boldsymbol{s},\boldsymbol{w};\boldsymbol{\theta}) \\ &= f_{\boldsymbol{s}}(\boldsymbol{s};\boldsymbol{\theta}) \cdot f_{\boldsymbol{w}}(\boldsymbol{w};\boldsymbol{\theta}) \cdot f(\boldsymbol{z}|\boldsymbol{s},\boldsymbol{w};\boldsymbol{\theta}) \end{aligned} \tag{100}$$

where in the transition from the first line of (100) to its second line we invoked the statistical independence of $s(t)$ and $w(t)$.

Taking the logarithm on both sides of the equation,

$$\log f_{\boldsymbol{y}}(\boldsymbol{y};\boldsymbol{\theta}) = \log f_{\boldsymbol{s}}(\boldsymbol{s};\boldsymbol{\theta}) + \log f_{\boldsymbol{w}}(\boldsymbol{w};\boldsymbol{\theta}) + \log f(\boldsymbol{z}|\boldsymbol{s},\boldsymbol{w};\boldsymbol{\theta}) \tag{101}$$

By (94),

$$\log f_{\boldsymbol{s}}(\boldsymbol{s};\boldsymbol{\theta}) = \log f(\boldsymbol{s}_{p-1}(0)) - \frac{N}{2}\log 2\pi g_s - \frac{1}{2g_s}\sum_{t=1}^{N}[s(t) + \boldsymbol{\alpha}^\top \boldsymbol{s}_{p-1}(t-1)]^2 \tag{102}$$

By the assumption that $w(t)$ is a white Gaussian noise process,

$$\log f_{\boldsymbol{w}}(\boldsymbol{w};\boldsymbol{\theta}) = \log f(\boldsymbol{w}_q(0)) - \frac{N}{2}\log 2\pi g_w - \frac{1}{2g_w}\sum_{t=1}^{N} w^2(t) \tag{103}$$

and by (92), (93)

$$\begin{aligned} \log f(\boldsymbol{z}|\boldsymbol{s},\boldsymbol{w};\boldsymbol{\theta}) &= -\frac{N}{2}\log 2\pi g_1 - \frac{1}{2g_1}\sum_{t=1}^{N}[z_1(t) - s(t) - \boldsymbol{a}^\top \boldsymbol{w}_q(t)]^2 \\ &\quad - \frac{N}{2}\log 2\pi g_2 - \frac{1}{2g_2}\sum_{t=1}^{N}[z_2(t) - \boldsymbol{b}^\top \boldsymbol{s}_r(t) - w(t)]^2 \end{aligned} \tag{104}$$

Substituting (102), (103), and (104) into (101), and assuming that $N \gg p, q$ so that the contributions of $\log f(\boldsymbol{s}_{p-1}(0))$ and $\log f(\boldsymbol{w}_q(0))$ are negligible,

$$\log f_{\boldsymbol{y}}(\boldsymbol{y};\boldsymbol{\theta}) = C - \frac{N}{2}\log g_s - \frac{1}{2g_s}\sum_{t=1}^{N}[s(t) + \boldsymbol{\alpha}^\top \boldsymbol{s}_{p-1}(t-1)]^2$$

$$
\begin{aligned}
& -\frac{N}{2}\log g_w - \frac{1}{2g_w}\sum_{t=1}^{N} w^2(t) \\
& -\frac{N}{2}\log g_1 - \frac{1}{2g_1}\sum_{t=1}^{N}[z_1(t) - s(t) - \boldsymbol{a}^\top \boldsymbol{w}_q(t)]^2 \\
& -\frac{N}{2}\log g_2 - \frac{1}{2g_2}\sum_{t=1}^{N}[z_2(t) - \boldsymbol{b}^\top \boldsymbol{s}_r(t) - w(t)]^2
\end{aligned}
\tag{105}
$$

where C is a constant independent of $\boldsymbol{\theta}$. Taking the conditional expectation given $\boldsymbol{z}$ at a parameter value $\boldsymbol{\theta}^{(\ell)}$,

$$
\begin{aligned}
Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(\ell)}) &= E_{\boldsymbol{\theta}^{(\ell)}}\{\log f_{\boldsymbol{y}}(\boldsymbol{y};\boldsymbol{\theta})|\boldsymbol{z}\} \\
&= C - \frac{N}{2}\log g_s - \frac{1}{2g_s}\sum_{t=1}^{N}[\widehat{s^2(t)}^{(\ell)} + 2\boldsymbol{\alpha}^\top \widehat{\boldsymbol{s}_{p-1}(t-1)\ s(t)}^{(\ell)} \\
&\qquad + \boldsymbol{\alpha}^\top \widehat{\boldsymbol{s}_{p-1}(t-1)\ \boldsymbol{s}_{p-1}^\top(t-1)}^{(\ell)}\ \boldsymbol{\alpha}] \\
&\quad - \frac{N}{2}\log g_w - \frac{1}{2g_w}\sum_{t=1}^{N}\widehat{w^2}^{(\ell)}(t) \\
&\quad - \frac{N}{2}\log g_1 - \frac{1}{2g_1}\sum_{t=1}^{N}[z_1^2(t) - 2\widehat{s}^{(\ell)}(t)z_1(t) + \widehat{s^2}^{(\ell)}(t) \\
&\qquad - 2\boldsymbol{a}^\top \widehat{\boldsymbol{w}}_q^{(\ell)}(t)\ z_1(t) + 2\boldsymbol{a}^\top \widehat{\boldsymbol{w}_q(t)s(t)}^{(\ell)} \\
&\qquad + \boldsymbol{a}^\top \widehat{\boldsymbol{w}_q(t)\boldsymbol{w}_q^\top(t)}^{(\ell)}\ \boldsymbol{a}] \\
&\quad - \frac{N}{2}\log g_2 - \frac{1}{2g_2}\sum_{t=1}^{N}[z_2^2(t) - 2\widehat{w}^{(\ell)}(t)z_2(t) + \widehat{w^2}^{(\ell)}(t) \\
&\qquad - 2\boldsymbol{b}^\top \widehat{\boldsymbol{s}}_r^{(\ell)}(t)z_2(t) + 2\boldsymbol{b}^\top \widehat{\boldsymbol{s}_r(t)w(t)}^{(\ell)} \\
&\qquad + \boldsymbol{b}^\top \widehat{\boldsymbol{s}_r(t)\boldsymbol{s}_r^\top(t)}^{(\ell)}\ \boldsymbol{b}]
\end{aligned}
\tag{106}
$$

where, as before,

$$
\widehat{(\cdot)}^{(\ell)} \triangleq E_{\boldsymbol{\theta}^{(\ell)}}\{\cdot|\boldsymbol{z}\}
\tag{107}
$$

Thus, the computation of $Q(\boldsymbol{\theta},\boldsymbol{\theta}^{(\ell)})$ (E-step) only requires the computation of $\widehat{\boldsymbol{x}}^{(\ell)}(t)$ and $\widehat{\boldsymbol{x}(t)\boldsymbol{x}^{\top}(t)}^{(\ell)}$, where $\boldsymbol{x}(t)$ is defined by:

$$\boldsymbol{x}(t)=\left[\begin{array}{c}\boldsymbol{s}_r(t)\\ \hline \boldsymbol{w}_q(t)\end{array}\right]=\left[\begin{array}{c}s(t-r)\\ s(t-r+1)\\ \vdots\\ s(t)\\ w(t-q)\\ w(t-q+1)\\ \vdots\\ w(t)\end{array}\right] \tag{108}$$

We further observe that the first line of (106) depends only on the desired signal parameters $(\boldsymbol{\alpha},g_s)$, the second line depends only on the spectral level g_w of the noise signal, the third line depends only on $(\boldsymbol{a},g_1)$, and the fourth line depends only on $(\boldsymbol{b},g_2)$. Therefore, the maximization of $Q(\boldsymbol{\theta},\boldsymbol{\theta}^{(\ell)})$ (M-step) decouples into the separate maximizations of each line of (106) with respect to the associated parameters. These maximizations can be solved analytically. Altogether, we obtain the following algorithm:

E-step: For $t=1,2,\ldots,N$ compute:

$$\widehat{\boldsymbol{x}}^{(\ell)}(t)=E_{\boldsymbol{\theta}^{(\ell)}}\{\boldsymbol{x}(t)|\boldsymbol{z}\} \tag{109}$$

$$\widehat{\boldsymbol{x}(t)\boldsymbol{x}^{\top}(t)}^{(\ell)}=E_{\boldsymbol{\theta}^{(\ell)}}\{\boldsymbol{x}(t)\boldsymbol{x}^{\top}(t)|\boldsymbol{z}\} \tag{110}$$

M-step: Compute

$$\widehat{\boldsymbol{\alpha}}^{(\ell+1)} = -\left[\sum_{t=1}^{N}\widehat{\boldsymbol{s}_{p-1}(t-1)\boldsymbol{s}_{p-1}^{\top}}^{(\ell)}\right]^{-1}\sum_{t=1}^{N}\widehat{\boldsymbol{s}_{p-1}(t-1)s(t)}^{(\ell)} \tag{111}$$

$$\widehat{g}_s^{(\ell+1)} = \frac{1}{N}\left\{\sum_{t=1}^{N}\widehat{s^2}^{(\ell)}(t)+\widehat{\boldsymbol{\alpha}}^{(\ell+1)\top}\cdot\sum_{t=1}^{N}\widehat{\boldsymbol{s}_{p-1}(t-1)s(t)}^{(\ell)}\right\} \tag{112}$$

$$\widehat{g}_w^{(\ell+1)} = \frac{1}{N}\sum_{t=1}^{N}\widehat{w^2}^{(\ell)}(t) \tag{113}$$

$$\widehat{\boldsymbol{a}}^{(\ell+1)} = \left[\sum_{t=1}^{N}\widehat{\boldsymbol{w}_q(t)\boldsymbol{w}_q^{\top}(t)}^{(\ell)}\right]^{-1}\sum_{t=1}^{N}\left[\widehat{\boldsymbol{w}}_q^{(\ell)}(t)z_1(t)-\widehat{\boldsymbol{w}_q(t)s(t)}^{(\ell)}\right] \tag{114}$$

$$\begin{aligned}\widehat{g}_1^{(\ell+1)} &= \frac{1}{N}\sum_{t=1}^{N}\left[z_1^2(t)-2\widehat{s}^{(\ell)}(t)z_1(t)+\widehat{s^2}^{(\ell)}(t)\right]\\ &\quad-\widehat{\boldsymbol{a}}^{(\ell+1)\top}\cdot\frac{1}{N}\sum_{t=1}^{N}\left[\widehat{\boldsymbol{w}}_q^{(\ell)}(t)z_1(t)-\widehat{\boldsymbol{w}_q(t)s(t)}^{(\ell)}\right]\end{aligned} \tag{115}$$

$$\hat{\boldsymbol{b}}^{(\ell+1)} = \left[\sum_{t=1}^{N} \widehat{\boldsymbol{s}_r(t)\boldsymbol{s}_r^{\top}(t)}^{(\ell)}\right]^{-1} \sum_{t=1}^{N} \left[\hat{\boldsymbol{s}}_r^{(\ell)}(t) z_2(t) - \widehat{\boldsymbol{s}_r(t) w(t)}^{(\ell)}\right] \tag{116}$$

$$\begin{aligned}
\hat{g}_2^{(\ell+1)} &= \frac{1}{N}\left\{\sum_{t=1}^{N}\left[z_2^2(t) - 2\hat{w}^{(\ell)}(t)z_2(t) + \widehat{w^2}^{(\ell)}(t)\right]\right. \\
&\quad \left. - \hat{\boldsymbol{b}}^{(\ell+1)\top} \sum_{t=1}^{N}\left[\hat{\boldsymbol{s}}_r^{(\ell)}(t) z_2(t) - \widehat{\boldsymbol{s}_r(t) w(t)}^{(\ell)}\right]\right\}
\end{aligned} \tag{117}$$

The algorithm has a nice intuitive form. In the E-step, we use the current parameter estimate $\boldsymbol{\theta}^{(\ell)}$ to estimate the sufficient statistics of the desired signal and the undesired (noise) signal. The M-step of the algorithm decouples as follows: Equation (111) is the Yule-Walker solution for the AR parameters where the sufficient statistics of the signal are substituted by their current estimates. Equations (114) and (116) are the least squares solutions for the $\boldsymbol{a}$ and $\boldsymbol{b}$ parameters, respectively, based on the estimated sufficient statistics. The gain parameters in (112), (113), (115), and (117) are re-estimated as the sample average of the corresponding power levels. We note that if some of the gain parameters are known *a priori* and need not be estimated, we simply eliminate the corresponding equations.

Since the algorithm is based on the EM method, it will converge monotonically to the ML estimate of $\boldsymbol{\theta}$ (or, at least, to a stationary point of the log-likelihood function), and it also provides the desired signal estimate $\hat{s}^{(\ell)}(t)$, which is the $(r+1)$st component of $\hat{\boldsymbol{x}}^{(\ell)}(t)$. We note again that for the purpose of signal enhancement, it is the signal estimate that we are essentially interested in.

As in the single sensor case, the computation of the conditional expectations in (109) and (110) can be carried out using the Kalman smoothing equations. To do that, we represent Eqs. (92)-(94) in state-space form:

$$\begin{aligned}
\boldsymbol{x}(t) &= \Phi \boldsymbol{x}(t-1) + G\boldsymbol{u}(t) && (118)\\
\boldsymbol{z}(t) &= H\boldsymbol{x}(t) + \boldsymbol{e}(t) && (119)
\end{aligned}$$

where state vector $\boldsymbol{x}(t)$ is defined in (108),

$$\boldsymbol{z}(t) = \begin{bmatrix} z_1(t) \\ z_2(t) \end{bmatrix} \tag{120}$$

$$\boldsymbol{u}(t) = \begin{bmatrix} u_s(t) \\ u_w(t) \end{bmatrix} \tag{121}$$

$$\boldsymbol{e}(t) = \begin{bmatrix} e_1(t) \\ e_2(t) \end{bmatrix} \tag{122}$$

$$G^{\top} = \begin{array}{c} \begin{array}{ccccccccc} & & & (r+1) & & & & & \\ & & & \downarrow & & & & & \end{array} \\ \left[\begin{array}{ccccccccc} 0 & \cdots & 0 & \sqrt{g_s} & 0 & \cdots & \cdots & 0 & \\ 0 & \cdots & \cdots & \cdots & \cdots & \cdots & 0 & \sqrt{g_w} \end{array} \right] \\ \begin{array}{ccccccccc} & & & & & & & & \uparrow \\ & & & & & & & & (r+q+2) \end{array} \end{array} \tag{123}$$

$$H = \begin{bmatrix} 0 & \cdots & \cdots & 0 & 1 & a_q & a_{q-1} & \cdots & \cdots & a_0 \\ b_r & b_{r-1} & \cdots & \cdots & b_0 & 0 & \cdots & \cdots & 0 & 1 \end{bmatrix} \tag{124}$$

and

$$\Phi = \left[\begin{array}{c|c} \Phi_s & 0 \\ \hline 0 & \Phi_w \end{array} \right] \tag{125}$$

where

$$\Phi_s = \begin{bmatrix} 0 & 1 & 0 & \cdots & \cdots & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & & & \vdots \\ \vdots & & \ddots & \ddots & \ddots & & \vdots \\ \vdots & & & \ddots & \ddots & \ddots & \vdots \\ \vdots & & & & \ddots & \ddots & 0 \\ \vdots & & & & & 0 & 1 \\ \vdots & \cdots & 0 & -\alpha_p & -\alpha_{p-1} & \cdots & \alpha_1 \end{bmatrix} \tag{126}$$

and

$$\Phi_w = \begin{bmatrix} 0 & 1 & 0 & \cdots & \cdots & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & & & \vdots \\ \vdots & & \ddots & \ddots & \ddots & & \vdots \\ \vdots & & & \ddots & \ddots & \ddots & \vdots \\ \vdots & & & & \ddots & \ddots & 0 \\ \vdots & & & & & 0 & 1 \\ \vdots & \cdots & 0 & \cdots & \cdots & \cdots & 0 \end{bmatrix} \tag{127}$$

We also define by R the 2×2 covariance matrix of $\boldsymbol{e}(t)$:

$$R = \begin{bmatrix} g_1 & 0 \\ 0 & g_2 \end{bmatrix} \tag{128}$$

Now, let us define:

$$\boldsymbol{\mu}_{t|n}^{(\ell)} = E_{\boldsymbol{\theta}^{(\ell)}}\{\boldsymbol{x}(t)|\boldsymbol{z}(1), \boldsymbol{z}(2), \ldots, \boldsymbol{z}(n)\} \tag{129}$$

$$\begin{aligned} P_{t|n}^{(\ell)} &= E_{\boldsymbol{\theta}^{(\ell)}}\{[\boldsymbol{x}(t) - \boldsymbol{\mu}_{t|n}^{(\ell)}][\boldsymbol{x}(t) - \boldsymbol{\mu}_{t|n}^{(\ell)}]^{\top}|\boldsymbol{z}(1), \boldsymbol{z}(2), \ldots, \boldsymbol{z}(n)\} \\ &= E_{\boldsymbol{\theta}^{(\ell)}}\{\boldsymbol{x}(t)\boldsymbol{x}^{\top}(t)|\boldsymbol{z}(1), \boldsymbol{z}(2), \ldots, \boldsymbol{z}(n)\} - \boldsymbol{\mu}_{t|n}^{(\ell)}\boldsymbol{\mu}_{t|n}^{(\ell)\top} \end{aligned} \tag{130}$$

Clearly, the conditional expectations in (109) and (110) are given by:

$$\widehat{\boldsymbol{x}}^{(\ell)}(t) = \boldsymbol{\mu}_{t|N}^{(\ell)} \tag{131}$$

$$\widehat{\boldsymbol{x}(t)\boldsymbol{x}^{\top}(t)}^{(\ell)} = \boldsymbol{\mu}_{t|N}^{(\ell)}\boldsymbol{\mu}_{t|N}^{(\ell)\top} + P_{t|N}^{(\ell)} \tag{132}$$

Denote by $\Phi^{(\ell)}$, $G^{(\ell)}$, $H^{(\ell)}$, and $R^{(\ell)}$ the matrices Φ, G, H, and R computed at the current parameter estimate $\boldsymbol{\theta} = \boldsymbol{\theta}^{(\ell)}$. Then, the Kalman smoothing equations compute $\boldsymbol{\mu}_{t|N}^{(\ell)}$ and $P_{t|N}^{(\ell)}$ in three stages as follows:

Propagation Equations

For $t = 1, 2, \ldots, N$

$$\boldsymbol{\mu}_{t|t-1}^{(\ell)} = \Phi^{(\ell)}\boldsymbol{\mu}_{t-1|t-1} \qquad , \boldsymbol{\mu}_{0|0}^{(\ell)} \tag{133}$$

$$P_{t|t-1}^{(\ell)} = \Phi^{(\ell)}P_{t-1|t-1}^{(\ell)}\Phi^{(\ell)\top} + G^{(\ell)}G^{(\ell)\top} \qquad , P_{0|0}^{(\ell)} \tag{134}$$

Updating Equations

For $t = 1, 2, \ldots, N$

$$\boldsymbol{\mu}_{t|t}^{(\ell)} = \boldsymbol{\mu}_{t|t-1}^{(\ell)} + K_t^{(\ell)}[z(t) - H^{(\ell)}\boldsymbol{\mu}_{t|t-1}^{(\ell)}] \tag{135}$$

$$P_{t|t}^{(\ell)} = [I - K_t^{(\ell)}H^{(\ell)}]\, P_{t|t-1}^{(\ell)} \tag{136}$$

where I is the identity matrix, and $K_t^{(\ell)}$ is the Kalman gain:

$$K_t^{(\ell)} = P_{t|t-1}^{(\ell)}H^{(\ell)\top}\left[H^{(\ell)}P_{t|t-1}^{(\ell)}H^{(\ell)\top} + R^{(\ell)}\right]^{-1} \tag{137}$$

Smoothing Equations

For $t = N, N-1, \ldots, 1$

$$\boldsymbol{\mu}_{t-1|N}^{(\ell)} = \boldsymbol{\mu}_{t-1|t-1}^{(\ell)} + S_{t-1}^{(\ell)}[\boldsymbol{\mu}_{t|N}^{(\ell)} - \Phi^{(\ell)}\boldsymbol{\mu}_{t-1|t-1}^{(\ell)}] \tag{138}$$

$$P_{t-1|N}^{(\ell)} = P_{t-1|t-1}^{(\ell)} + S_{t-1}^{(\ell)}[P_{t|N}^{(\ell)} - P_{t|t-1}^{(\ell)}]S_{t-1}^{(\ell)\top} \tag{139}$$

where

$$S_{t-1}^{(\ell)} \triangleq P_{t-1|t-1}^{(\ell)}\Phi^{(\ell)\top}P_{t|t-1}^{(\ell)-1} \tag{140}$$

The outcomes $\boldsymbol{\mu}_{t|N}^{(\ell)}$ and $P_{t|N}^{(\ell)}$, $t = 1, 2, \ldots, N$ of the recursions in (138) and (139) are then substituted in (131) and (132) to yield the conditional expectations that are required in the E-step of the algorithm.

4.2 Sequential/Adaptive Algorithms

In complete analogy with the single sensor case, the EM algorithm for the two-sensor case can be converted into a sequential/adaptive algorithm by substituting the time index for the iteration index, replacing the Kalman smoother with the Kalman filter, and incorporating exponential weighting into the parameter estimation update.

More specifically, denote by $\hat{\boldsymbol{x}}(t|t)$ and $\widehat{\boldsymbol{x}(t|t)\boldsymbol{x}^\top(t|t)}$ the estimate of $\boldsymbol{x}(t)$ and $\boldsymbol{x}(t)\boldsymbol{x}^\top(t)$ based on the observed data to time t and the current parameter estimate $\hat{\boldsymbol{\theta}}(t)$. Also, define

$$\boldsymbol{\mu}_{t|t} \triangleq \hat{\boldsymbol{x}}(t|t) \tag{141}$$

and

$$P_{t|t} \triangleq \widehat{\boldsymbol{x}(t|t)\boldsymbol{x}^\top(t|t)} - \boldsymbol{\mu}_{t|t}\boldsymbol{\mu}_{t|t}^\top \tag{142}$$

Then, $\boldsymbol{\mu}_{t|t}$ and $P_{t|t}$ are computed recursively as follows:

Propagation Equation:

$$\boldsymbol{\mu}_{t|t-1} = \hat{\Phi}_t\boldsymbol{\mu}_{t-1|t-1} \qquad , \boldsymbol{\mu}_{0|0} \tag{143}$$

$$P_{t|t-1} = \hat{\Phi}_t P_{t-1|t-1}\hat{\Phi}_t^\top + \hat{G}_t\hat{G}_t^\top \qquad , P_{0|0} \tag{144}$$

Updating equations

$$\boldsymbol{\mu}_{t|t} = \boldsymbol{\mu}_{t|t-1} + \hat{K}_t[\boldsymbol{z}(t) - \hat{H}_t\boldsymbol{\mu}_{t|t-t}] \tag{145}$$

$$P_{t|t} = [I - \hat{K}_t\hat{H}_t]P_{t|t-1} \tag{146}$$

where

$$\hat{K}_t = P_{t|t-1}\hat{H}_t^\top\left[\hat{H}_t P_{t|t-1}\hat{H}_t^\top + \hat{R}_t\right]^{-1} \tag{147}$$

where $\hat{\Phi}_t$, $\hat{G}_t$, $\hat{H}_t$, and $\hat{R}_t$ are the matrices Φ, G, H, and R computed at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}(t)$. In this way, the state (signal) estimate is generated using a single Kalman filter whose parameters are continuously updated.

For the application we have in mind, the dimensions r and q of the FIR filters characterizing the acoustic transfer function tend to be very large (on the order of several hundred coefficients). Consequently, the computation of the Kalman filtering equations involve the multiplication of matrices that are very large in dimension. This computation can be greatly simplified by exploiting the structure of the matrices $\hat{\Phi}_t$, $\hat{G}_t$, $\hat{H}_t$, and $\hat{R}_t$. The derivation is given in Appendix A. To present the results, let $\boldsymbol{\mu}_{t-1|t-1}$ be partitioned as follows:

$$\boldsymbol{\mu}_{t-1|t-1} = \begin{bmatrix} \boldsymbol{\mu}_s \\ \hline \boldsymbol{\mu}_w \end{bmatrix} \begin{matrix} \updownarrow r+1 \\ \updownarrow q+1 \end{matrix} = \begin{bmatrix} \mu_1 \\ \hline \boldsymbol{\mu}_2 \\ \hline \mu_3 \\ \hline \boldsymbol{\mu}_4 \end{bmatrix} \begin{matrix} \updownarrow 1 \\ \updownarrow r \\ \updownarrow 1 \\ \updownarrow q \end{matrix} \tag{148}$$

Let $\boldsymbol{\mu}_p$ be the lower $p \times 1$ subvector of $\boldsymbol{\mu}_2$:

$$\boldsymbol{\mu}_2 = \begin{bmatrix} \\ \hline \boldsymbol{\mu}_p \end{bmatrix} \updownarrow p \tag{149}$$

Let $P_{t-1|t-1}$ be partitioned as follows:

$$\begin{aligned}
P_{t-1|t-1} &= \begin{array}{cl}
\left[\begin{array}{c|c} P_{ss} & P_{sw} \\ \hline P_{sw}^{\top} & P_{ww} \end{array}\right] & \begin{array}{l} \updownarrow r+1 \\ \updownarrow q+1 \end{array} \\
\begin{array}{cc} \overleftrightarrow{\;r+1\;} & \overleftrightarrow{\;q+1\;} \end{array} &
\end{array} \\
&= \begin{array}{cl}
\left[\begin{array}{c|c|c|c} P_{11} & P_{12} & P_{13} & P_{14} \\ \hline P_{12}^{\top} & P_{22} & P_{23} & P_{24} \\ \hline P_{13}^{\top} & P_{23}^{\top} & P_{33} & P_{34} \\ \hline P_{14}^{\top} & P_{24}^{\top} & P_{34}^{\top} & P_{44} \end{array}\right] & \begin{array}{l} \updownarrow 1 \\ \updownarrow r \\ \updownarrow 1 \\ \updownarrow q \end{array} \\
\begin{array}{cccc} \overleftrightarrow{\;1\;} & \overleftrightarrow{\;r\;} & \overleftrightarrow{\;1\;} & \overleftrightarrow{\;q\;} \end{array} &
\end{array}
\end{aligned} \tag{150}$$

Let Γ_p be the following sub-matrix of P_{22}:

$$P_{22} = \begin{array}{cl} \left[\begin{array}{c|c} \quad\quad & \Gamma_p \end{array}\right] & \updownarrow r \\ \begin{array}{cc} \quad\quad & \overleftrightarrow{\;p\;} \end{array} & \end{array} \tag{151}$$

and let Γ_{pp} be the following sub-matrix of Γ_p:

$$\Gamma_p = \begin{array}{cl} \left[\begin{array}{c} \quad\quad \\ \hline \Gamma_{pp} \end{array}\right] & \begin{array}{l} \\ \updownarrow p \end{array} \\ \overleftrightarrow{\;p\;} & \end{array} \tag{152}$$

(so that Γ_{pp} is the lower right $p \times p$ sub-matrix of P_{22}).

Let Λ_p be the following sub-matrix of P_{24}:

$$P_{24} = \begin{array}{cl} \left[\begin{array}{c} \quad\quad \\ \hline \Lambda_p \end{array}\right] & \begin{array}{l} \\ \updownarrow p \end{array} \\ \overleftrightarrow{\;q\;} & \end{array} \tag{153}$$

Let $\boldsymbol{a}$ and $\boldsymbol{b}$ be partitioned as follows:

$$\boldsymbol{a} = \left[\begin{array}{c} a_1 \\ \hline a_0 \end{array}\right] \begin{array}{l} \updownarrow q \\ \updownarrow 1 \end{array} \tag{154}$$

$$\boldsymbol{b} = \left[\begin{array}{c} \boldsymbol{b}_1 \\ \hline b_0 \end{array}\right] \begin{array}{l} \updownarrow r \\ \updownarrow 1 \end{array} \tag{155}$$

For convenience, we shall use $\boldsymbol{\theta}$ instead of $\widehat{\boldsymbol{\theta}}(t)$. With this notation, the Kalman filtering equations are given by:

Propagation Equations

$$\boldsymbol{\mu}_{t|t-1} = \left[\begin{array}{c} \boldsymbol{\mu}_2 \\ \hline -\boldsymbol{\alpha}^{\top}\boldsymbol{\mu}_p \\ \hline \boldsymbol{\mu}_4 \\ \hline 0 \end{array}\right] \begin{array}{l} \updownarrow r \\ \updownarrow 1 \\ \updownarrow q \\ \updownarrow 1 \end{array} \tag{156}$$

$$
P_{t|t-1} = \begin{array}{cl}
\left[\begin{array}{c|c|c|c}
P_{22} & -\Gamma_p \boldsymbol{\alpha} & P_{24} & \mathbf{0} \\ \hline
 & \boldsymbol{\alpha}^\top \Gamma_{pp} \boldsymbol{\alpha} + g_s & -\boldsymbol{\alpha}^\top \Lambda_p & 0 \\ \hline
 & & P_{44} & \mathbf{0} \\ \hline
 & & & g_w
\end{array}\right] & \begin{array}{l} \updownarrow r \\ \updownarrow 1 \\ \updownarrow q \\ \updownarrow 1 \end{array} \\
\begin{array}{cccc} \overset{\longleftrightarrow}{r} & \overset{\longleftrightarrow}{1} & \overset{\longleftrightarrow}{q} & \overset{\longleftrightarrow}{1} \end{array} &
\end{array} \tag{157}
$$

where we note that $P_{t|t-1}$ is a symmetric matrix.

Updating Equations

$$
\boldsymbol{\mu}_{t|t} = \boldsymbol{\mu}_{t|t-1} + K_t \cdot \left[\begin{array}{c} z_1(t) + \boldsymbol{\alpha}^\top \boldsymbol{\mu}_p - \boldsymbol{a}_1^\top \boldsymbol{\mu}_4 \\ \hline z_2(t) + b_0 \boldsymbol{\alpha}^\top \boldsymbol{\mu}_p - \boldsymbol{b}_1^\top \boldsymbol{\mu}_2 \end{array} \right] \tag{158}
$$

$$
P_{t|t} = P_{t|t-1} - K_t D_t^\top \tag{159}
$$

where K_t (the Kalman gain) is given by:

$$
K_t = D_t F_t^{-1} \tag{160}
$$

where

$$
D_t = \begin{array}{cl}
\left[\begin{array}{c|c}
-\Gamma_p \boldsymbol{\alpha} + P_{24} \boldsymbol{a}_1 & P_{22} \boldsymbol{b}_1 - \Gamma_p \boldsymbol{\alpha} \cdot b_0 \\ \hline
\boldsymbol{\alpha}^\top \Gamma_{pp} \boldsymbol{\alpha} + g_s - \boldsymbol{\alpha}^\top \Lambda_p \boldsymbol{a}_1 & -\boldsymbol{\alpha}^\top \Gamma_p^\top \boldsymbol{b}_1 + b_0 (\boldsymbol{\alpha}^\top \Gamma_{pp} \boldsymbol{\alpha} + g_s) \\ \hline
-\Lambda_p^\top \boldsymbol{\alpha} + P_{44} \boldsymbol{a}_1 & P_{24}^\top \boldsymbol{b}_1 - \Lambda_p^\top \boldsymbol{\alpha} \cdot b_0 \\ \hline
a_0 \cdot g_w & g_w
\end{array}\right] & \begin{array}{l} \updownarrow r \\ \updownarrow 1 \\ \updownarrow q \\ \updownarrow 1 \end{array} \\
\begin{array}{cc} \overset{\longleftrightarrow}{1} & \overset{\longleftrightarrow}{1} \end{array} &
\end{array} \tag{161}
$$

and F_t is the 2×2 symmetric matrix:

$$
F_t = \begin{bmatrix} f_{11} & f_{12} \\ f_{12} & f_{22} \end{bmatrix} \tag{162}
$$

where

$$
f_{11} = \boldsymbol{a}_1^\top P_{44} \boldsymbol{a}_1 - 2 \boldsymbol{a}_1^\top \Lambda_p^\top \boldsymbol{\alpha} + \boldsymbol{\alpha}^\top \Gamma_{pp} \boldsymbol{\alpha} + g_s + a_0^2 g_w + g_1 \tag{163}
$$

$$
f_{22} = \boldsymbol{b}_1^\top P_{22} \boldsymbol{b}_1 - 2 b_0 \cdot \boldsymbol{b}_1^\top \Gamma_p \boldsymbol{\alpha} + b_0^2 (\boldsymbol{\alpha}^\top \Gamma_{pp} \boldsymbol{\alpha} + g_s) + g_w + g_2 \tag{164}
$$

$$
f_{12} = \boldsymbol{a}_1^\top P_{24}^\top \boldsymbol{b}_1 - b_0 \cdot \boldsymbol{a}_1^\top \Lambda_p^\top \boldsymbol{\alpha} - \boldsymbol{b}_1^\top \Gamma_p \boldsymbol{\alpha} + b_0 (\boldsymbol{\alpha}^\top \Gamma_{pp} \boldsymbol{\alpha} + g_s) + a_0 g_w \tag{165}
$$

To update the parameter estimates, we modify Eqs. (111)–(117) by replacing the iteration index (ℓ) by the time index (t), use the data only up to the current time t, and incorporate exponential weighting. With that, we obtain:

$$
\hat{\boldsymbol{\alpha}}(t+1) = -\left[\sum_{\tau=1}^{t} \gamma_s^{t-\tau} \widehat{\boldsymbol{s}_{p-1}(\tau-1|\tau) \boldsymbol{s}_{p-1}^\top(\tau-1|\tau)} \right]^{-1} \sum_{\tau=1}^{t} \gamma_s^{t-\tau} \widehat{\boldsymbol{s}_{p-1}(\tau-1|\tau) s(\tau|\tau)} \tag{166}
$$

$$\hat{g}_s(t+1) = \frac{1}{\sum_{\tau=1}^{t}\gamma_s^{t-\tau}}\left[\sum_{\tau=1}^{t}\gamma_s^{t-\tau}\widehat{s^2}(\tau|\tau) + \hat{\boldsymbol{a}}^{\top}(t+1)\sum_{\tau=1}^{t}\gamma_s^{t-\tau}\widehat{\boldsymbol{s}_{p-1}(\tau-1|\tau)s(\tau|\tau)}\right] \quad (167)$$

$$\hat{g}_w(t+1) = \frac{1}{\sum_{\tau=1}^{t}\gamma_w^{t-\tau}}\cdot\sum_{\tau=1}^{t}\gamma_w^{t-\tau}\widehat{w^2}(\tau|\tau) \quad (168)$$

$$\hat{\boldsymbol{a}}(t+1) = \left[\sum_{\tau=1}^{t}\gamma_a^{t-\tau}\widehat{\boldsymbol{w}_q(\tau|\tau)\boldsymbol{w}_q^{\top}(\tau|\tau)}\right]^{-1}\sum_{\tau=1}^{t}\gamma_a^{t-\tau}\left[\hat{\boldsymbol{w}}_q(\tau|\tau)z_1(\tau) - \widehat{\boldsymbol{w}_q(\tau|\tau)s(\tau|\tau)}\right] \quad (169)$$

$$\begin{aligned}\hat{g}_1(t+1) &= \frac{1}{\sum_{\tau=1}^{t}\gamma_a^{t-\tau}}\left\{\sum_{\tau=1}^{t}\gamma_a^{t-\tau}\left[z_1^2(\tau) - 2\hat{s}(\tau|\tau)z_1(\tau) + \widehat{s^2}(\tau|\tau)\right]\right. \\ &\quad \left. - \hat{\boldsymbol{a}}^{\top}(t+1)\sum_{\tau=1}^{t}\gamma_a^{t-\tau}\left[\hat{\boldsymbol{w}}_q(\tau|\tau)z_1(\tau) - \widehat{\boldsymbol{w}_q(\tau|\tau)s(\tau|\tau)}\right]\right\} \quad (170)\end{aligned}$$

$$\hat{\boldsymbol{b}}(t+1) = \left[\sum_{\tau=1}^{t}\gamma_b^{t-\tau}\widehat{\boldsymbol{s}_r(\tau|\tau)\boldsymbol{s}_r^{\top}(\tau|\tau)}\right]^{-1}\sum_{\tau=1}^{t}\gamma_b^{t-\tau}\left[\widehat{\boldsymbol{s}_r}(\tau|\tau)z_2(\tau) - \widehat{\boldsymbol{s}_r(\tau|\tau)w(\tau|\tau)}\right] \quad (171)$$

$$\begin{aligned}\hat{g}_2(t+1) &= \frac{1}{\sum_{\tau=1}^{t}\gamma_b^{t-\tau}}\left\{\sum_{\tau=1}^{t}\gamma_b^{t-\tau}[z_2^2(\tau) - 2\hat{w}(\tau|\tau)z_2(\tau) + \widehat{w^2}(\tau|\tau)]\right. \\ &\quad \left. - \hat{\boldsymbol{b}}^{\top}(t+1)\sum_{\tau=1}^{t}\gamma_b^{t-\tau}\left[\hat{\boldsymbol{s}}_r(\tau|\tau)z_2(\tau) - \widehat{\boldsymbol{s}_r(\tau|\tau)w(\tau|\tau)}\right]\right\} \quad (172)\end{aligned}$$

These equations can be computed recursively as follows:

$$\begin{aligned}R_{11}(t) &= \sum_{\tau=1}^{t}\gamma_s^{t-\tau}\widehat{\boldsymbol{s}_{p-1}(\tau-1|\tau)\boldsymbol{s}_{p-1}^{\top}(\tau-1|\tau)} \\ &= \gamma_s R_{11}(t-1|t) + \widehat{\boldsymbol{s}_{p-1}(t-1|t)\boldsymbol{s}_{p-1}^{\top}(t-1|t)} \quad (173)\end{aligned}$$

$$\begin{aligned}R_{12}(t) &= \sum_{\tau=1}^{t}\gamma_s^{t-\tau}\widehat{\boldsymbol{s}_{p-1}(\tau-1|\tau)s(\tau|\tau)} \\ &= \gamma_s R_{12}(t-1|t) + \widehat{\boldsymbol{s}_{p-1}(t-1|t)s(t|t)} \quad (174)\end{aligned}$$

$$R_{22}(t) = \sum_{\tau=1}^{t}\gamma_s^{t-\tau}\widehat{s^2}(\tau|\tau) = \gamma_s R_{22}(t-1) + \widehat{s^2}(t|t) \quad (175)$$

$$Q_{11}(t) = \sum_{\tau=1}^{t}\gamma_w^{t-\tau}\widehat{w^2}(\tau|\tau) = \gamma_w Q_{11}(t-1) + \widehat{w^2}(t|t) \quad (176)$$

$$A_{11}(t) = \sum_{\tau=1}^{t}\gamma_a^{t-\tau}\widehat{\boldsymbol{w}_q(\tau|\tau)\boldsymbol{w}_q^{\top}(\tau|\tau)}$$

$$= \gamma_a A_{11}(t-1) + \widehat{\boldsymbol{w}_q(t|t)\boldsymbol{w}_q^\top(t|t)} \tag{177}$$

$$A_{12}(t) = \sum_{\tau=1}^{t} \gamma_a^{t-\tau}\left[\widehat{\boldsymbol{w}}_q(\tau|\tau)z_1(\tau) - \widehat{\boldsymbol{w}_q(\tau|\tau)s(\tau|\tau)}\right]$$

$$= \gamma_a A_{12}(t-1) + \widehat{\boldsymbol{w}}_q(t|t)z_1(t) - \widehat{\boldsymbol{w}_q(t|t)s(t|t)} \tag{178}$$

$$A_{22}(t) = \sum_{\tau=1}^{t} \gamma_a^{t-\tau}[z_1^2(\tau) - 2\widehat{s}(\tau|\tau)z_1(\tau) + \widehat{s^2}(\tau|\tau)]$$

$$= \gamma_a A_{22}(t-1) + z_1^2(t) - 2\widehat{s}(t|t)z_1(t) + \widehat{s^2}(t|t) \tag{179}$$

$$B_{11}(t) = \sum_{\tau=1}^{t} \gamma_b^{t-\tau}\widehat{\boldsymbol{s}_r(\tau|\tau)\boldsymbol{s}_r^\top(\tau|\tau)} = \gamma_b B_{11}(t-1) + \widehat{\boldsymbol{s}_r(t|t)\boldsymbol{s}_r^\top(t|t)} \tag{180}$$

$$B_{12}(t) = \sum_{\tau=1}^{t} \gamma_b^{t-\tau}\left[\widehat{\boldsymbol{s}}_r(\tau|\tau)z_2(\tau) - \widehat{\boldsymbol{s}_r(\tau)w(\tau|\tau)}\right]$$

$$= \gamma_b B_{12}(t-1) + \widehat{\boldsymbol{s}}_r(t|t)z_2(t) - \widehat{\boldsymbol{s}_r(t|t)w(t|t)} \tag{181}$$

$$B_{22}(t) = \sum_{\tau=1}^{t} \gamma_b^{t-\tau}[z_2^2(\tau) - 2\widehat{w}(\tau|\tau)z_2(\tau) + \widehat{w^2}(\tau|\tau)]$$

$$= \gamma_b B_{22}(t-1) + z_2^2(t) - 2\widehat{w}(t|t)z_2(t) + \widehat{w^2}(t|t) \tag{182}$$

Then:

$$\widehat{\boldsymbol{\alpha}}(t+1) = -R_{11}^{-1}(t)R_{12}(t)$$

$$= \widehat{\boldsymbol{\alpha}}(t) - R_{11}^{-1}(t)\left[\widehat{\boldsymbol{s}_{p-1}(t-1|t)s(t|t)} + \widehat{\boldsymbol{s}_{p-1}(t-1|t)\boldsymbol{s}_{p-1}^\top(t-1|t)}\,\widehat{\boldsymbol{\alpha}}(t)\right] \tag{183}$$

$$\widehat{g}_s(t+1) = \frac{1-\gamma_s}{1-\gamma_s^t}[R_{22}(t) + \widehat{\boldsymbol{\alpha}}^\top(t+1)R_{12}(t)] \tag{184}$$

$$\widehat{g}_w(t+1) = \frac{1-\gamma_w}{1-\gamma_w^t}Q_{11}(t) \tag{185}$$

$$\widehat{\boldsymbol{a}}(t+1) = A_{11}^{-1}(t)A_{12}(t)$$

$$= \widehat{\boldsymbol{a}}(t) + A_{11}^{-1}(t)\left[\widehat{\boldsymbol{w}}_q(t|t)z_1(t) - \widehat{\boldsymbol{w}_q(t|t)s(t|t)} - \widehat{\boldsymbol{w}_q(t|t)\boldsymbol{w}_q^\top(t|t)}\,\widehat{\boldsymbol{a}}(t)\right] \tag{186}$$

$$\widehat{g}_1(t+1) = \frac{1-\gamma_a}{1-\gamma_a^t}[A_{22}(t) - \widehat{\boldsymbol{a}}^\top(t+1)A_{12}(t)] \tag{187}$$

$$\widehat{\boldsymbol{b}}(t+1) = B_{11}^{-1}(t)B_{12}(t)$$

$$= \widehat{\boldsymbol{b}}(t) + B_{11}^{-1}(t)\left[\widehat{\boldsymbol{s}}_r(t|t)z_2(t) - \widehat{\boldsymbol{s}_r(t|t)w(t|t)} - \widehat{\boldsymbol{s}_r(t|t)\boldsymbol{s}_r^\top(t|t)\widehat{\boldsymbol{b}}(t)}\right] \tag{188}$$

$$\widehat{g}_2(t) = \frac{1-\gamma_b}{1-\gamma_b^t}[B_{22}(t) - \widehat{\boldsymbol{b}}^\top(t+1)B_{12}(t)] \tag{189}$$

We note that the recursions in (186) and in (188) involve the inversion of the matrices $A_{11}(t)$ and $B_{11}(t)$ at each step of the algorithm. Since these matrices may be very large in dimension, it may be computationally expensive. However, as it turns out, these matrices (also $R_{11}(t)$) have weak dependence in t since they quickly converge to their expected value. After these matrices approach their steady state, we may fix them and obtain computationally much simpler recursions.

We also note that if the various parameters are slowly varying, we do not have to update their estimates each time sample. Thus, to update the slowly varying parameters of the speech signal, we may compute $R_{11}(t)$ and $R_{12}(t)$ recursively in time using (173) and (174), respectively; but, re-estimate the AR parameters by multiplying $R_{11}^{-1}(t)$ with $R_{12}(t)$ only occasionally. The same procedure can be applied when estimating the coefficients $\boldsymbol{a}$ and $\boldsymbol{b}$ of the transfer functions (under quasi-stationary assumption), with a considerable savings in computation.

V Gradient-Based Algorithms

As an alternative to the EM algorithm, consider the following gradient-search algorithm for solving the maximization in (1):

$$\theta_i^{(\ell+1)} = \theta_i^{(\ell)} + \delta_i \cdot \frac{1}{N} \cdot \frac{\partial \log f_{\boldsymbol{Z}}(\boldsymbol{z};\boldsymbol{\theta})}{\partial \theta_i}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^{(\ell)}} \tag{190}$$

where $\theta_i^{(\ell)}$ is the estimate of θ_i (the i^{th} component of $\boldsymbol{\theta}$) after ℓ iteration cycles, δ_i is the step-size (it may vary in dimension and size depending on the index i), and N corresponds to the number of data samples. For sufficiently small step-size, this algorithm converges to the ML solution, or at least to a stationary point of the log-likelihood function.

To compute the partial derivatives in (190), we observe from (9) that

$$\frac{\partial}{\partial \theta_i} P(\boldsymbol{\theta},\boldsymbol{\theta}')\Big|_{\boldsymbol{\theta}'=\boldsymbol{\theta}} = 0 \qquad i = 1,2,\cdots \tag{191}$$

(since $P(\boldsymbol{\theta},\boldsymbol{\theta}')$ obtains its maximum at $\boldsymbol{\theta}' = \boldsymbol{\theta}$). Therefore, differentiating (8) with respect to θ_i at $\boldsymbol{\theta}' = \boldsymbol{\theta}$,

$$\begin{aligned} \frac{\partial}{\partial \theta_i} \log f_{\boldsymbol{Z}}(\boldsymbol{z};\boldsymbol{\theta}) &= \frac{\partial}{\partial \theta_i} Q(\boldsymbol{\theta},\boldsymbol{\theta}')\Big|_{\boldsymbol{\theta}'=\boldsymbol{\theta}} \\ &\underset{\text{by (6)}}{=} E_{\boldsymbol{\theta}}\left\{ \frac{\partial}{\partial \theta_i} \log f_{\boldsymbol{Y}}(\boldsymbol{y};\boldsymbol{\theta}) \Big| \boldsymbol{z} \right\} \end{aligned} \tag{192}$$

where in the transition from the first line of (192) to the second line we have assumed that the operations of differentiation and expectation (integration) are interchangable. This identity was first presented in [4], and more recently in [1], [5], and [6]. It asserts that the derivatives $\frac{\partial}{\partial \theta_i} \log f_{\boldsymbol{Z}}(\boldsymbol{z};\boldsymbol{\theta})$, where $\boldsymbol{z}$ denotes the observed (incomplete) data, can be calculated by taking the conditional expectation of the derivatives $\frac{\partial}{\partial \theta_i} \log f_{\boldsymbol{Y}}(\boldsymbol{y};\boldsymbol{\theta})$, where $\boldsymbol{y}$ denotes the complete data. This identity appears to be very useful for a set of problems of interest to us, since the differentiation of $\log f_{\boldsymbol{Y}}(\boldsymbol{y};\boldsymbol{\theta})$ is much easier than the differentiation of $\log f_{\boldsymbol{Z}}(\boldsymbol{z};\boldsymbol{\theta})$, as indicated in [7], [8].

5.1 Single-Sensor Case: Iterative Gradient Algorithm

Differentiating (25),

$$\frac{\partial}{\partial \boldsymbol{\alpha}} \log f_{\boldsymbol{Y}}(\boldsymbol{y};\boldsymbol{\theta}) = -\frac{1}{g_s}\sum_{t=1}^{N}\left[s(t)\boldsymbol{s}_{p-1}^{\mathsf{T}}(t-1) + \boldsymbol{\alpha}^{\mathsf{T}}\boldsymbol{s}_{p-1}(t-1)\boldsymbol{s}_{p-1}^{\mathsf{T}}(t-1)\right] \tag{193}$$

$$\frac{\partial}{\partial g_s} \log f_{\boldsymbol{Y}}(\boldsymbol{y};\boldsymbol{\theta}) = -\frac{N}{2g_s} + \frac{1}{2g_s^2}\sum_{t=1}^{N}\left[s(t) + \boldsymbol{\alpha}^{\mathsf{T}}\boldsymbol{s}_{p-1}(t-1)\right]^2 \tag{194}$$

$$\frac{\partial}{\partial g_\epsilon} \log f_{\boldsymbol{Y}}(\boldsymbol{y};\boldsymbol{\theta}) = -\frac{N}{2g_\epsilon} + \frac{1}{2g_\epsilon^2}\sum_{t=1}^{N}[z(t) - s(t)]^2 \tag{195}$$

Invoking (192),

$$\frac{\partial}{\partial \boldsymbol{\alpha}} \log f_{\boldsymbol{Z}}(\boldsymbol{z};\boldsymbol{\theta}) = -\frac{1}{g_s}\sum_{t=1}^{N}\left[\widehat{s(t)\boldsymbol{s}_{p-1}^{\mathsf{T}}(t-1)} + \boldsymbol{\alpha}^{\mathsf{T}}\widehat{\boldsymbol{s}_{p-1}(t-1)\boldsymbol{s}_{p-1}^{\mathsf{T}}(t-1)}\right] \tag{196}$$

$$\begin{aligned}\frac{\partial}{\partial g_s} \log f_{\boldsymbol{Z}}(\boldsymbol{z};\boldsymbol{\theta}) = & -\frac{N}{2g_s^2} + \frac{1}{2g_s^2}\sum_{t=1}^{N}\Big[\widehat{s^2}(t) + 2\boldsymbol{\alpha}^{\mathsf{T}}\widehat{\boldsymbol{s}_{p-1}(t-1)s(t)} + \\ & + \boldsymbol{\alpha}^{\mathsf{T}}\widehat{\boldsymbol{s}_{p-1}(t-1)\boldsymbol{s}_{p-1}^{\mathsf{T}}(t-1)}\boldsymbol{\alpha}\Big]\end{aligned} \tag{197}$$

$$\frac{\partial}{\partial g_\epsilon} \log f_{\boldsymbol{Z}}(\boldsymbol{z};\boldsymbol{\theta}) = -\frac{N}{2g_\epsilon} + \frac{1}{2g_\epsilon^2}\sum_{t=1}^{N}\left[z^2(t) - 2z(t)\widehat{s}(t) + \widehat{s^2}(t)\right] \tag{198}$$

where we define:

$$\widehat{(\cdot)} \triangleq E_{\boldsymbol{\theta}}\{\cdot|\boldsymbol{z}\} \tag{199}$$

Equations (196)–(199) specify the components of the gradient needed for the algorithm in (190).

As before, we denote by:

$$\boldsymbol{\theta}^{(\ell)} = \begin{bmatrix} \widehat{\boldsymbol{\alpha}}^{(\ell)} \\ \widehat{g}_s^{(\ell)} \\ \widehat{g}_\epsilon^{(\ell)} \end{bmatrix} \tag{200}$$

and by

$$\widehat{(\cdot)}^{(\ell)} = E_{\boldsymbol{\theta}^{(\ell)}}\{\cdot|\boldsymbol{z}\} \tag{201}$$

Then, the algorithm in (190) can be put in a form very similar to the EM algorithm:

Signal Estimation: for $t = 1, 2, \ldots, N$ compute:

$$\widehat{\boldsymbol{s}}_p^{(\ell)} = \begin{bmatrix} \widehat{\boldsymbol{s}}_{p-1}^{(\ell)}(t-1) \\ \widehat{s}^{(\ell)}(t) \end{bmatrix} \tag{202}$$

$$\widehat{s_p(t)s_p^\top(t)}^{(\ell)} = \left[\begin{array}{c|c} \widehat{s_{p-1}(t-1)s_{p-1}^\top(t-1)}^{(\ell)} & \widehat{s_{p-1}(t-1)s(t)}^{(\ell)} \\ \hline \widehat{s(t)s_{p-1}^\top(t-1)}^{(\ell)} & \widehat{s^2}^{(\ell)}(t) \end{array}\right] \tag{203}$$

Parameter Estimation:

$$\hat{\alpha}^{(\ell+1)} = \hat{\alpha}^{(\ell)} - \delta_\alpha \cdot \frac{1}{\hat{g}_s^{(\ell)}} \cdot \frac{1}{N} \sum_{t=1}^{N} \left[\widehat{s_{p-1}(t-1)s(t)}^{(\ell)} + \widehat{s_{p-1}(t-1)s_{p-1}^\top(t-1)}^{(\ell)} \hat{\alpha}^{(\ell)} \right] \tag{204}$$

$$\begin{aligned} \hat{g}_s^{(\ell+1)} = \hat{g}_s^{(\ell)} + \delta_s \Bigg\{ -\frac{1}{2\hat{g}_s^{(\ell)}} + \frac{1}{2\hat{g}_s^{(\ell)^2}} \cdot \frac{1}{N} \sum_{t=1}^{N} \Bigg[& \widehat{s^2}^{(\ell)}(t) + 2\hat{\alpha}^{(\ell)^\top} \cdot \widehat{s_{p-1}(t-1)s(t)}^{(\ell)} \\ & + \hat{\alpha}^{(\ell)^\top} \widehat{s_{p-1}(t-1)s_{p-1}^\top(t-1)}^{(\ell)} \hat{\alpha}^{(\ell)} \Bigg] \Bigg\} \end{aligned} \tag{205}$$

$$\hat{g}_\epsilon^{(\ell+1)} = \hat{g}_\epsilon^{(\ell)} + \delta_\epsilon \cdot \left\{ -\frac{1}{2\hat{g}_\epsilon^{(\ell)}} + \frac{1}{2\hat{g}_\epsilon^{(\ell)^2}} \cdot \frac{1}{N} \sum_{t=1}^{N} [z^2(t) - 2z(t)\hat{s}^{(\ell)}(t) + \widehat{s^2}^{(\ell)}(t)] \right\} \tag{206}$$

Equations (202) and (203) are identical to Eqs. (28) and (29). Thus, the signal estimation is identical to the E-step of the EM algorithm. The difference between the EM algorithm and the gradient algorithm is in the parameter updating.

We note that δ_α is dimension-free since α is dimension-free, δ_s has the dimension of g_s^2, and δ_ϵ has the dimension of g_ϵ^2. It therefore suggests normalizing δ_s and δ_ϵ as follows:

$$\delta_s = \hat{g}_s^{(\ell)^2} \cdot \tilde{\delta}_s \tag{207}$$

$$\delta_\epsilon = \hat{g}_\epsilon^{(\ell)^2} \cdot \tilde{\delta}_\epsilon \tag{208}$$

where $\tilde{\delta}_s$ and $\tilde{\delta}_\epsilon$ are dimension-free. In this setting, Eqs. (205) and (206) become:

$$\begin{aligned} \hat{g}_s^{(\ell+1)} = \left(1 - \frac{\tilde{\delta}_s}{2}\right) \hat{g}_s^{(\ell)} + \frac{\tilde{\delta}_s}{2} \frac{1}{N} \sum_{t=1}^{N} \Bigg[& \widehat{s^2}^{(\ell)}(t) + 2\hat{\alpha}^{(\ell)^\top} \widehat{s_{p-1}(t-1)s(t)}^{(\ell)} \\ & + \hat{\alpha}^{(\ell)^\top} \widehat{s_{p-1}(t-1)s_{p-1}^\top(t-1)}^{(\ell)} \hat{\alpha}^{(\ell)} \Bigg] \end{aligned} \tag{209}$$

$$\hat{g}_\epsilon^{(\ell+1)} = \left(1 - \frac{\tilde{\delta}_\epsilon}{2}\right) \hat{g}_\epsilon^{(\ell)} + \frac{\tilde{\delta}_\epsilon}{2} \frac{1}{N} \sum_{t=1}^{N} \left[z^2(t) - 2z(t)\hat{s}^{(\ell)}(t) + \widehat{s^2}^{(\ell)}(t) \right] \tag{210}$$

We note that δ_α, $\tilde{\delta}_s$, and $\tilde{\delta}_\epsilon$ need not be the same.

5.2 Single-Sensor Case: Sequential Adaptive Algorithm

We can convert the iterative algorithm into a sequential/adaptive algorithm. Thus, instead of estimating the signal by applying the Kalman smoother iteratively, it is estimated using a forward Kalman filter whose parameters are continuously updated. The recursive formulas for the signal (state) estimation are the same as before, and are given by (62)-(63).

To update the parameter estimates, we replace the iteration index by the time index, and substitute the sample averages $\frac{1}{N}\sum(\cdot)$ in (204), (209), and (210) by their most current term:

$$\widehat{\boldsymbol{\alpha}}(t+1) = \widehat{\boldsymbol{\alpha}}(t) - \delta_\alpha \cdot \frac{1}{\widehat{g}_s(t)}\left[\widehat{\boldsymbol{s}_{p-1}(t-1|t)s(t|t)} + \widehat{\boldsymbol{s}_{p-1}(t-1|t)\boldsymbol{s}_{p-1}^\top(t-1|t)}\,\widehat{\boldsymbol{\alpha}}(t)\right] \tag{211}$$

$$\begin{aligned}\widehat{g}_s(t+1) &= \left(1-\frac{\tilde{\delta}_s}{2}\right)\widehat{g}_s(t) + \frac{\tilde{\delta}_s}{2}\left[\widehat{s^2}(t|t) + 2\widehat{\alpha}^\top(t)\widehat{\boldsymbol{s}_{p-1}(t-1|t)s(t|t)}\right. \\ &\qquad \left. + \,\widehat{\boldsymbol{\alpha}}^\top(t)\widehat{\boldsymbol{s}_{p-1}(t-1|t)\boldsymbol{s}_{p-1}^\top(t-1|t)}\,\widehat{\boldsymbol{\alpha}}(t)\right]\end{aligned} \tag{212}$$

$$\widehat{g}_\epsilon(t+1) = \left(1-\frac{\tilde{\delta}_\epsilon}{2}\right)\widehat{g}_\epsilon(t) + \frac{\tilde{\delta}_\epsilon}{2}\left[z^2(t) - 2z(t)\widehat{s}(t|t) + \widehat{s^2}(t|t)\right] \tag{213}$$

We may choose δ_α, $\tilde{\delta}_s$, and $\tilde{\delta}_\epsilon$ to be functions of time, i.e. $\delta_\alpha = \delta_\alpha(t)$. We note that the effect of these coefficients is similar to the effect of the forgetting factors in the previous development.

Invoking the special structure of (62) and (63), and following straightforward algebra manipulations (see Appendix B for details), Eqs. (211)-(213) can be represented in the form:

$$\widehat{\boldsymbol{\alpha}}(t+1) = \widehat{\boldsymbol{\alpha}}(t) - \frac{\delta_\alpha}{\eta(t)}\cdot\left[\widehat{\boldsymbol{s}}_{p-1}(t-1|t)e(t) + \Lambda_{t-1}\widehat{\boldsymbol{\alpha}}(t)\right] \tag{214}$$

$$\widehat{g}_s(t+1) = \widehat{g}_s(t) + \frac{\delta_s}{2\eta^2(t)}\left[e^2(t) - \eta(t)\right] \tag{215}$$

$$\widehat{g}_\epsilon(t+1) = \widehat{g}_\epsilon(t) + \frac{\delta_\epsilon}{2\eta^2(t)}\left[e^2(t) - \eta(t)\right] \tag{216}$$

where

$$e(t) \triangleq z(t) + \widehat{\boldsymbol{\alpha}}^\top(t)\widehat{\boldsymbol{s}}_{p-1}(t-1|t) \tag{217}$$

We note that $-\widehat{\boldsymbol{\alpha}}^\top(t)\widehat{\boldsymbol{s}}_{p-1}(t-1|t)$ can be interpreted as the predicted value of the signal at time t based on observations to time $(t-1)$, that is:

$$\widehat{s}(t|t-1) = -\widehat{\boldsymbol{\alpha}}^\top(t)\widehat{\boldsymbol{s}}_{p-1}(t-1|t)$$

Therefore,

$$e(t) = z(t) - \widehat{s}(t|t-1)$$

can be interpreted as the new information, or the innovation, carried by the new measurement.

5.3 Two-Sensor Case: Iterative Gradient Algorithm

Substituting (101) into (192) and performing the indicated differentiation and expectation operations, the components of the log-likelihood gradient are:

$$\frac{\partial}{\partial \boldsymbol{\alpha}} \log f_{\boldsymbol{Z}}(\boldsymbol{z};\boldsymbol{\theta}) = -\frac{1}{g_s}\sum_{t=1}^{N}\left[\widehat{s(t)\boldsymbol{s}_{p-1}^{\top}(t-1)} + \boldsymbol{\alpha}^{\top}\widehat{\boldsymbol{s}_{p-1}(t-1)\boldsymbol{s}_{p-1}^{\top}(t-1)}\right] \tag{218}$$

$$\begin{aligned}\frac{\partial}{\partial g_s} \log f_{\boldsymbol{Z}}(\boldsymbol{z};\boldsymbol{\theta}) = -\frac{N}{2g_s} + \frac{1}{1g_s^2}\sum_{t=1}^{N}\Big[&\widehat{s^2}(t) + 2\boldsymbol{\alpha}^{\top}\widehat{\boldsymbol{s}_{p-1}(t-1)s(t)} \\ &+ \boldsymbol{\alpha}^{\top}\widehat{\boldsymbol{s}_{p-1}(t-1)\boldsymbol{s}_{p-1}^{\top}(t-1)}\,\boldsymbol{\alpha}\Big]\end{aligned} \tag{219}$$

$$\frac{\partial}{\partial g_w} \log f_{\boldsymbol{Z}}(\boldsymbol{z};\boldsymbol{\theta}) = -\frac{N}{2g_w} + \frac{1}{2g_w^2}\sum_{t=1}^{N}\widehat{w^{2(\ell)}}(t) \tag{220}$$

$$\frac{\partial}{\partial \boldsymbol{a}} \log f_{\boldsymbol{Z}}(\boldsymbol{z};\boldsymbol{\theta}) = \frac{1}{g_1}\sum_{t=1}^{N}[z_1(t)\widehat{\boldsymbol{w}}_q^{\top}(t) - \widehat{s(t)\boldsymbol{w}_q^{\top}(t)} - \boldsymbol{a}^{\top}\widehat{\boldsymbol{w}_q(t)\boldsymbol{w}_q^{\top}(t)}] \tag{221}$$

$$\begin{aligned}\frac{\partial}{\partial g_1} \log f_{\boldsymbol{Z}}(\boldsymbol{z};\boldsymbol{\theta}) = -\frac{N}{2g_1} + \frac{1}{2g_1^2}\sum_{t=1}^{N}\Big[&z_1^2(t) - 2z_1(t)\widehat{s}(t) + \widehat{s^2}(t) - 2\boldsymbol{a}^{\top}\widehat{\boldsymbol{w}}_q(t)z_1(t) \\ &+ 2\boldsymbol{a}^{\top}\widehat{\boldsymbol{w}_q(t)s(t)} + \boldsymbol{a}^{\top}\widehat{\boldsymbol{w}_q(t)\boldsymbol{w}_q^{\top}(t)}\,\boldsymbol{a}\Big]\end{aligned} \tag{222}$$

$$\frac{\partial}{\partial \boldsymbol{b}} \log f_{\boldsymbol{Z}}(\boldsymbol{z};\boldsymbol{\theta}) = \frac{1}{g_2}\sum_{t=1}^{N}\left[z_2(t)\widehat{\boldsymbol{s}}_r^{\top}(t) - \widehat{w(t)\boldsymbol{s}_r(t)}^{\top} - \boldsymbol{b}^{\top}\widehat{\boldsymbol{s}_r(t)\boldsymbol{s}_r^{\top}(t)}\right] \tag{223}$$

$$\begin{aligned}\frac{\partial}{\partial g_2} \log f_{\boldsymbol{Z}}(\boldsymbol{z};\boldsymbol{\theta}) = -\frac{N}{2g_2} + \frac{1}{2g_2^2}\sum_{t=1}^{N}\Big[&z_2^2(t) - 2z_2(t)\widehat{w}(t) + \widehat{w^2}(t) - 2\boldsymbol{b}^{\top}\widehat{\boldsymbol{s}}_r(t)z_2(t) \\ &+ 2\boldsymbol{b}^{\top}\widehat{\boldsymbol{s}_r(t)w(t)} + \boldsymbol{b}^{\top}\widehat{\boldsymbol{s}_r(t)\boldsymbol{s}_r^{\top}(t)}\,\boldsymbol{b}\Big]\end{aligned} \tag{224}$$

where $\widehat{(\cdot)} \triangleq E_{\boldsymbol{\theta}}\{\cdot|\boldsymbol{z}\}$ as in (199). Thus, the computation of the log-likelihood gradient requires the computation of $\widehat{\boldsymbol{x}}(t)$ and $\widehat{\boldsymbol{x}(t)\boldsymbol{x}^{\top}(t)}$, where $\boldsymbol{x}(t)$ is the state vector defined by (104).

Substituting (217)–(222) into (190), and performing the iteration cycles by first calculating $\boldsymbol{\theta} = \boldsymbol{\theta}^{(\ell)}$, followed by a parameter update step, the resulting gradient algorithm can be put in a form very similar to the EM algorithm:

Signal Estimation:
for $t = 1, 2, \ldots, N$ compute:

$$\widehat{\boldsymbol{x}}^{(\ell)}(t) = E_{\boldsymbol{\theta}^{(\ell)}}\{\boldsymbol{x}(t)|\boldsymbol{z}\} \tag{225}$$

$$\widehat{\boldsymbol{x}(t)\boldsymbol{x}^{\top}(t)}^{(\ell)} = E_{\boldsymbol{\theta}^{(\ell)}}\{\boldsymbol{x}(t)\boldsymbol{x}^{\top}(t)|z\} \tag{226}$$

Parameter Estimation:

$$\widehat{\boldsymbol{\alpha}}^{(\ell+1)} = \widehat{\boldsymbol{\alpha}}^{(\ell)} - \frac{\delta_\alpha}{\widehat{g}_s^{(\ell)}} \cdot \frac{1}{N} \sum_{t=1}^{N} [\widehat{\boldsymbol{s}_{p-1}(t-1)s(t)}^{(\ell)} + \widehat{\boldsymbol{s}_{p-1}(t-1)\boldsymbol{s}_{p-1}^{\top}(t-1)}\,\widehat{\boldsymbol{\alpha}}^{(\ell)}] \tag{227}$$

$$\widehat{g}_s^{(\ell+1)} = \left(1 - \frac{\tilde{\delta}_s}{2}\right) \widehat{g}_s^{(\ell)} + \frac{\tilde{\delta}_s}{2} \cdot \frac{1}{N} \sum_{t=1}^{N} \Big[\widehat{s^2}^{(\ell)}(t) + 2\widehat{\boldsymbol{\alpha}}^{(\ell)\top} \widehat{\boldsymbol{s}_{p-1}(t-1)s(t)}^{(\ell)} + \widehat{\boldsymbol{\alpha}}^{(\ell)\top} \widehat{\boldsymbol{s}_{p-1}(t-1)\boldsymbol{s}_{p-1}^{\top}(t-1)}\,\widehat{\boldsymbol{\alpha}}^{(\ell)} \Big] \tag{228}$$

$$\widehat{g}_w^{(\ell+1)} = \left(1 - \frac{\tilde{\delta}_w}{2}\right) \widehat{g}_w^{(\ell)} + \frac{\tilde{\delta}_w}{2} \cdot \frac{1}{N} \sum_{t=1}^{N} \widehat{w^2}^{(\ell)}(t) \tag{229}$$

$$\widehat{\boldsymbol{a}}^{(\ell+1)} = \widehat{\boldsymbol{a}}^{(\ell)} + \frac{\tilde{\delta}_a}{\widehat{g}_1^{(\ell)}} \cdot \frac{1}{N} \sum_{t=1}^{N} [\widehat{\boldsymbol{w}}_q^{(\ell)}(t) z_1(t) - \widehat{\boldsymbol{w}_q(t)s(t)}^{(\ell)} - \widehat{\boldsymbol{w}_q(t)\boldsymbol{w}_q^{\top}(t)}\,\widehat{\boldsymbol{a}}^{(\ell)}] \tag{230}$$

$$\widehat{g}_1^{(\ell+1)} = \left(1 - \frac{\tilde{\delta}_1}{2}\right) \widehat{g}_1^{(\ell)} + \frac{\tilde{\delta}_1}{2} \cdot \frac{1}{N} \sum_{t=1}^{N} \Big[z_1^2(t) - 2z_1(t)\widehat{s}^{(\ell)}(t) + \widehat{s^2}^{(\ell)}(t) - 2\widehat{\boldsymbol{a}}^{(\ell)\top} \widehat{\boldsymbol{w}}_q^{(\ell)}(t) z_1(t) + 2\widehat{\boldsymbol{a}}^{(\ell)\top} \widehat{\boldsymbol{w}_q(t)s(t)}^{(\ell)} + \widehat{\boldsymbol{a}}^{(\ell)\top} \widehat{\boldsymbol{w}_q(t)\boldsymbol{w}_q^{\top}(t)}\,\widehat{\boldsymbol{a}}^{(\ell)} \Big] \tag{231}$$

$$\widehat{\boldsymbol{b}}^{(\ell+1)} = \widehat{\boldsymbol{b}}^{(\ell)} + \frac{\tilde{\delta}_b}{\widehat{g}_2^{(\ell)}} \cdot \frac{1}{N} \sum_{t=1}^{N} [\widehat{\boldsymbol{s}}_r^{(\ell)}(t) z_2(t) - \widehat{\boldsymbol{s}_r(t)w(t)}^{(\ell)} - \widehat{\boldsymbol{s}_r(t)\boldsymbol{s}_r^{\top}(t)}^{(\ell)}\,\widehat{\boldsymbol{b}}] \tag{232}$$

$$\widehat{g}_2^{(\ell+1)} = \left(1 - \frac{\tilde{\delta}_2}{2}\right) \widehat{g}_2^{(\ell)} + \frac{\tilde{\delta}_2}{2} \cdot \frac{1}{N} \sum_{t=1}^{N} \Big[z_2^2(t) - 2z_2(t)\widehat{w}^{(\ell)}(t) + \widehat{w^2}^{(\ell)}(t) - 2\widehat{\boldsymbol{b}}^{(\ell)\top} \widehat{\boldsymbol{s}}_r^{(\ell)}(t) z_2(t) + 2\widehat{\boldsymbol{b}}^{(\ell)\top} \widehat{\boldsymbol{s}_r(t)w(t)}^{(\ell)} + \widehat{\boldsymbol{b}}^{(\ell)\top} \widehat{\boldsymbol{s}_r(t)\boldsymbol{s}_r^{\top}(t)}\,\widehat{\boldsymbol{b}}^{(\ell)} \Big] \tag{233}$$

The signal estimation step is identical to the E-step in the EM algorithm, and it is performed by applying the Kalman smoothing equations. The difference between the two algorithms is in the parameter updating. We note that $\tilde{\delta}_s$, $\tilde{\delta}_w$, $\tilde{\delta}_1$, and $\tilde{\delta}_2$ are dimension-free normalized coefficients.

5.4 Two-Sensor Case: Sequential/Adaptive Algorithm

To convert the algorithm into a sequential/adaptive scheme, we replace the iteration index by the time index, and replace the Kalman smoother by the Kalman filter whose parameters are continuously updated. The formulas for the signal (state) estimation are identical to those presented in conjunction with the EM algorithm.

To update the parameter estimates, once again we replace the iteration index by the time index, and substitute the cumulative averages $\frac{1}{N}\sum(\cdot)$ by their most current term, to obtain:

$$\hat{\boldsymbol{\alpha}}(t+1) = \hat{\boldsymbol{\alpha}}(t) - \frac{\delta_\alpha}{\hat{g}_s(t)} \cdot \left[\widehat{\boldsymbol{s}_{p-1}(t-1|t)s(t|t)} + \widehat{\boldsymbol{s}_{p-1}(t-1|t)\boldsymbol{s}_{p-1}^{\mathrm{T}}(t-1|t)}\,\hat{\boldsymbol{\alpha}}(t) \right] \tag{234}$$

$$\hat{g}_s(t+1) = \left(1 - \frac{\bar{\delta}_s}{2}\right)\hat{g}_s(t) + \frac{\bar{\delta}_s}{2} \cdot \left[\widehat{s^2}(t|t) + 2\hat{\boldsymbol{\alpha}}^{\mathrm{T}}(t)\widehat{\boldsymbol{s}_{p-1}(t-1|t)s(t|t)} \right.$$
$$\left. \hat{\boldsymbol{\alpha}}^{\mathrm{T}}(t)\widehat{\boldsymbol{s}_{p-1}(t-1|t)\boldsymbol{s}_{p-1}^{\mathrm{T}}(t-1|t)}\,\hat{\boldsymbol{\alpha}}(t) \right] \tag{235}$$

$$\hat{g}_w(t+1) = \left(1 - \frac{\bar{\delta}_w}{2}\right)\hat{g}_w(t) + \frac{\bar{\delta}_w}{2}\widehat{w^2}(t|t) \tag{236}$$

$$\hat{\boldsymbol{a}}(t+1) = \hat{\boldsymbol{a}}(t) + \frac{\delta_a}{\hat{g}_1(t)} \left[\hat{\boldsymbol{w}}_q(t|t)z_1(t) - \widehat{\boldsymbol{w}_q(t|t)s(t|t)} \right.$$
$$\left. - \widehat{\boldsymbol{w}_q(t|t)\boldsymbol{w}_q^{\mathrm{T}}(t|t)}\,\hat{\boldsymbol{a}}(t) \right] \tag{237}$$

$$\hat{g}_1(t+1) = \left(1 - \frac{\bar{\delta}_1}{2}\right)\hat{g}_1(t) + \frac{\bar{\delta}_1}{2}\left[z_1^2(t) - 2z_1(t)\hat{s}(t|t) + \widehat{s^2}(t|t) \right.$$
$$\left. - 2\hat{\boldsymbol{a}}^{\mathrm{T}}(t)\hat{\boldsymbol{w}}_q(t|t)z_1(t) + 2\hat{\boldsymbol{a}}^{\mathrm{T}}(t)\widehat{\boldsymbol{w}_q(t|t)s(t)} + \hat{\boldsymbol{a}}^{\mathrm{T}}(t)\widehat{\boldsymbol{w}_q(t|t)\boldsymbol{w}_q^{\mathrm{T}}(t|t)}\,\hat{\boldsymbol{a}}(t) \right] \tag{238}$$

$$\hat{\boldsymbol{b}}(t+1) = \hat{\boldsymbol{b}}(t) + \frac{\delta_b}{\hat{g}_2(t)} \left[\boldsymbol{s}_r(t|t)z_2(t) - \widehat{\boldsymbol{s}_r(t|t)w(t|t)} - \widehat{\boldsymbol{s}_r(t|t)\boldsymbol{s}_r^{\mathrm{T}}(t|t)}\,\hat{\boldsymbol{b}}(t) \right] \tag{239}$$

$$\hat{g}_2(t+1) = \left(1 - \frac{\bar{\delta}_2}{2}\right)\hat{g}_2(t) + \frac{\bar{\delta}_2}{2}\left[z_2^2(t) - 2z_2(t)\hat{w}(t|t) + \widehat{w^2}(t|t) \right.$$
$$\left. - 2\hat{\boldsymbol{b}}^{\mathrm{T}}(t)\hat{\boldsymbol{s}}_r(t|t)z_2(t) + 2\hat{\boldsymbol{b}}^{\mathrm{T}}(t)\widehat{\boldsymbol{s}_r(t|t)w(t)} + \hat{\boldsymbol{b}}^{\mathrm{T}}(t)\widehat{\boldsymbol{s}_r(t|t)\boldsymbol{s}_r^{\mathrm{T}}(t|t)}\,\hat{\boldsymbol{b}}(t) \right] \tag{240}$$

Note the nice intuitive form of Eq. (236). Specifically, the new estimate $\hat{g}_w(t+1)$ is obtained by weighting the previous estimate $\hat{g}_w(t)$ with the current estimate of the average power of $w(t)$. All other gain update equations have the same meaning.

We finally note that similarly to the single sensor case, the structure of the equations for the signal (state) estimate can be exploited to further simplify the form of the parameter update equations.

Appendix A: Development of the Efficient Form of the Kalman Filtering Equations for the Two-Sensor Case

Substituting (125) and (148) into (143)

$$\mu_{t|t-1} = \left[\begin{array}{c|c} \Phi_s & 0 \\ \hline 0 & \Phi_w \end{array}\right] \left[\begin{array}{c} \mu_s \\ \hline \mu_w \end{array}\right] = \left[\begin{array}{c} \Phi_s \mu_s \\ \hline \Phi_w \mu_w \end{array}\right] \tag{A.1}$$

where

$$\Phi_s \mu_s = \left[\begin{array}{c|c|c} \mathbf{0} & \multicolumn{2}{c}{I} \\ \hline 0 & \mathbf{0}^\top & -\boldsymbol{\alpha}^\top \end{array}\right] \left[\begin{array}{c} \mu_1 \\ \hline \mu_2 \end{array}\right] = \left[\begin{array}{c} \mu_2 \\ \hline -\boldsymbol{\alpha}^\top \mu_p \end{array}\right] \tag{A.2}$$

and

$$\Phi_w \mu_w = \left[\begin{array}{c|c} \mathbf{0} & I \\ \hline 0 & \mathbf{0}^\top \end{array}\right] \left[\begin{array}{c} \mu_3 \\ \hline \mu_4 \end{array}\right] = \left[\begin{array}{c} \mu_4 \\ \hline 0 \end{array}\right] \tag{A.3}$$

Substituting (A.2) and (A.3) into (A.1), we obtain (156).

Substituting (125), (150), and (123) into (144)

$$\begin{aligned} P_{t|t-1} &= \left[\begin{array}{c|c} \Phi_s & 0 \\ \hline 0 & \Phi_w \end{array}\right] \left[\begin{array}{c|c} P_{ss} & P_{sw} \\ \hline P_{sw}^\top & P_{ww} \end{array}\right] \left[\begin{array}{c|c} \Phi_s^\top & 0 \\ \hline 0 & \Phi_w^\top \end{array}\right] + GG^\top \\ &= \left[\begin{array}{c|c} \Phi_s P_{ss} \Phi_s^\top & \Phi_s P_{sw} \Phi_w^\top \\ \hline \Phi_w P_{sw}^\top \Phi_s^\top & \Phi_w P_{ww} \Phi_w^\top \end{array}\right] + GG^\top \end{aligned} \tag{A.4}$$

where

$$\begin{aligned} \Phi_s P_{ss} \Phi_s^\top &= \left[\begin{array}{c|c|c} \mathbf{0} & \multicolumn{2}{c}{I} \\ \hline 0 & \mathbf{0}^\top & -\boldsymbol{\alpha}^\top \end{array}\right] \left[\begin{array}{c|c} P_{11} & P_{12} \\ \hline P_{12}^\top & P_{22} \end{array}\right] \left[\begin{array}{c|c} \mathbf{0}^\top & 0 \\ \hline I & \begin{array}{c} \mathbf{0} \\ \hline -\boldsymbol{\alpha} \end{array} \end{array}\right] \\ &= \left[\begin{array}{c|c} P_{22} & -\Gamma_p \boldsymbol{\alpha} \\ \hline -\boldsymbol{\alpha}^\top \Gamma_p^\top & \boldsymbol{\alpha}^\top \Gamma_{pp} \boldsymbol{\alpha} \end{array}\right] \end{aligned} \tag{A.5}$$

$$\begin{aligned} \Phi_s P_{sw} \Phi_w^\top &= \left[\begin{array}{c|c|c} \mathbf{0} & \multicolumn{2}{c}{I} \\ \hline 0 & \mathbf{0}^\top & -\boldsymbol{\alpha}^\top \end{array}\right] \left[\begin{array}{c|c} P_{13} & P_{14} \\ \hline P_{23} & P_{24} \end{array}\right] \left[\begin{array}{c|c} \mathbf{0}^\top & 0 \\ \hline I & \mathbf{0} \end{array}\right] \\ &= \left[\begin{array}{c|c} P_{24} & \mathbf{0} \\ \hline -\boldsymbol{\alpha}^\top \Lambda_p & 0 \end{array}\right] \end{aligned} \tag{A.6}$$

$$\begin{aligned} \Phi_w P_{ww} \Phi_w^\top &= \left[\begin{array}{c|c} \mathbf{0} & I \\ \hline 0 & \mathbf{0}^\top \end{array}\right] \left[\begin{array}{c|c} P_{33} & P_{34} \\ \hline P_{34}^\top & P_{44} \end{array}\right] \left[\begin{array}{c|c} \mathbf{0}^\top & 0 \\ \hline I & \mathbf{0} \end{array}\right] \\ &= \left[\begin{array}{c|c} P_{44} & \mathbf{0} \\ \hline \mathbf{0}^\top & 0 \end{array}\right] \end{aligned} \tag{A.7}$$

and $GG^\top$ is the matrix having only two non-zero components:

$$GG^\top = \begin{array}{c} \left[\begin{array}{ccccc} & & & & \\ & g_s & & & \\ & & & g_w & \\ & & & & \end{array}\right] \begin{array}{l} \\ \leftarrow r+1 \\ \leftarrow r+q+2 \\ \\ \end{array} \\ \begin{array}{cc} \uparrow & \uparrow \\ r+1 & r+q+2 \end{array} \end{array} \tag{A.8}$$

Substituting (A.5)–(A.8) into (A.4), we obtain (157).

Now, let the matrix H in (124) be written in the form:

$$H = \left[\begin{array}{c|c|c|c} \mathbf{0}^\top & 1 & \boldsymbol{a}_1^\top & a_0 \\ \hline \boldsymbol{b}_1^\top & b_0 & \mathbf{0}^\top & 1 \end{array}\right] \tag{A.9}$$

Define the matrices:

$$D_t = P_{t|t-1}H^\top \tag{A.10}$$

$$F_t = HD_t + R \tag{A.11}$$

Substituting (157) and (A.9) into (A.10) and carrying out the matrix multiplication, we find that D_t is given by (161). Substituting (161), (A.9), and (128) into (A.11) and performing the indicated matrix multiplications, we find that F_t is the matrix in (162), whose elements are given by (163)–(165).

Finally, substituting (A.10) and (A.11) into (146) and (147), we obtain Eqs. (159) and (160), respectively.

Appendix B: Development of Eqs. (214)–(216)

We start from Eq. (211). Decomposing $\widehat{s_p(t|t)s_p^\top(t|t)}$ into a product of means plus a covariance term (see (51)), and invoking (65)–(67), we obtain:

$$\begin{aligned}
\widehat{\alpha}(t+1) &= \widehat{\alpha}(t) - \delta_\alpha \cdot \frac{1}{\widehat{g}_s(t)}\Bigg\{\widehat{s}_{p-1}(t-1|t)\left[\widehat{s}(t|t) + \widehat{\alpha}^\top(t)\widehat{s}_{p-1}(t-1|t)\right] \\
&\qquad - \frac{\widehat{g}_\epsilon(t)}{\eta(t)}\Lambda_{t-1}\widehat{\alpha}(t) + \left[\Lambda_{t-1} - \frac{1}{\eta(t)}\Lambda_{t-1}\widehat{\alpha}(t)\widehat{\alpha}^\top(t)\Lambda_{t-1}\right]\widehat{\alpha}(t)\Bigg\} \\
&= \widehat{\alpha}(t) - \delta_\alpha\frac{1}{\widehat{g}_s(t)}\Bigg\{\widehat{s}_{p-1}(t-1|t)\cdot\frac{z(t) + \widehat{\alpha}^\top(t)s_{p-1}(t-1|t-1)}{\eta(t)}\widehat{g}_s(t) \\
&\qquad + \Lambda_{t-1}\widehat{\alpha}(t)\underbrace{\left[-\frac{\widehat{g}_\epsilon(t)}{\eta(t)} + 1 - \frac{\widehat{\alpha}^\top(t)\Lambda_{t-1}\widehat{\alpha}(t)}{\eta(t)}\right]}_{\widehat{g}_s(t)/\eta(t)}\Bigg\} \\
&= \widehat{\alpha}(t) - \frac{\delta_\alpha}{\eta(t)}\{\widehat{s}_{p-1}(t-1|t)[z(t) + \widehat{\alpha}^\top(t)s_{p-1}(t-1|t-1)] + \Lambda_{t-1}\widehat{\alpha}(t)\} \qquad \text{(B.1)}
\end{aligned}$$

Substituting (217) into (B.1) immediately yields (214).

Next, we manipulate Eq. (212):

$$\begin{aligned}
\widehat{g}_s(t+1) &= \widehat{g}_s(t) + \frac{\tilde{\delta}_s}{2}\Bigg\{-\widehat{g}_s(t) + \left[\widehat{s}(t|t) + \widehat{\alpha}^\top(t)\widehat{s}_{p-1}(t-1|t)\right]^2 \\
&\qquad + \frac{\widehat{g}_\epsilon(t)}{\eta(t)}\left[\widehat{\alpha}^\top(t)\Lambda_{t-1}\widehat{\alpha}(t) + \widehat{g}_s(t)\right] - 2\cdot\frac{\widehat{g}_\epsilon(t)}{\eta(t)}\widehat{\alpha}^\top(t)\Lambda_{t-1}\widehat{\alpha}(t) \\
&\qquad + \widehat{\alpha}^\top(t)\left[\Lambda_{t-1} - \frac{1}{\eta(t)}\Lambda_{t-1}\widehat{\alpha}(t)\widehat{\alpha}^\top(t)\Lambda_{t-1}\right]\widehat{\alpha}(t)\Bigg\} \\
&= \widehat{g}_s(t) + \frac{\tilde{\delta}_s}{2}\Bigg\{-\widehat{g}_s(t) + \left[\frac{z(t) + \widehat{\alpha}^\top(t)\widehat{s}_{p-1}(t-1|t-1)}{\eta(t)}\widehat{g}_s(t)\right]^2 \\
&\qquad - \frac{\widehat{g}_\epsilon(t)}{\eta(t)}\widehat{\alpha}^\top(t)\Lambda_{t-1}\widehat{\alpha}(t) + \frac{\widehat{g}_\epsilon(t)\widehat{g}_s(t)}{\eta(t)} + \widehat{\alpha}^\top(t)\Lambda_{t-1}\widehat{\alpha}(t) \\
&\qquad - \frac{1}{\eta(t)}\left[\widehat{\alpha}^\top(t)\Lambda_{t-1}\widehat{\alpha}(t)\right]^2\Bigg\} \\
&= \widehat{g}_s(t) + \frac{\tilde{\delta}_s}{2}\Bigg\{-\widehat{g}_s(t) + \frac{\widehat{g}_s^2(t)}{\eta^2(t)}\left[z(t) + \widehat{\alpha}^\top(t)s_{p-1}(t-1|t-1)\right]^2 \\
&\qquad + \widehat{\alpha}^\top(t)\Lambda_{t-1}\widehat{\alpha}(t)\underbrace{\left[-\frac{\widehat{g}_\epsilon(t)}{\eta(t)} + 1 - \frac{\widehat{\alpha}^\top(t)\Lambda_{t-1}\widehat{\alpha}(t)}{\eta(t)}\right]}_{\widehat{g}_s(t)/\eta(t)} + \frac{\widehat{g}_\epsilon(t)\widehat{g}_s(t)}{\eta(t)}\Bigg\} \\
&= \widehat{g}_s(t) + \frac{\tilde{\delta}_s}{2}\Bigg\{\frac{\widehat{g}_s^2(t)}{\eta^2(t)}\left[z(t) + \widehat{\alpha}^\top(t)s_{p-1}(t-1|t-1)\right]^2
\end{aligned}$$

$$+\frac{\hat{g}_s(t)}{\eta(t)}\underbrace{\left[-\eta(t)+\hat{\boldsymbol{\alpha}}^\top(t)\Lambda_{t-1}\hat{\boldsymbol{\alpha}}(t)+\hat{g}_\epsilon(t)\right]}_{-\hat{g}_s(t)}\Bigg\}$$

$$= \quad \hat{g}_s(t)+\frac{\tilde{\delta}_s}{2}\cdot\frac{\hat{g}_s^2(t)}{\eta^2(t)}\left\{\left[z(t)+\hat{\boldsymbol{\alpha}}^\top(t)\hat{\mathbf{s}}_{p-1}(t-1|t-1)\right]^2-\eta(t)\right\} \qquad \text{(B.2)}$$

Substituting (207) and (217) into (B.2) immediately yields (215).

Finally, we manipulated Eq. (213) as follows:

$$\begin{aligned}
\hat{g}_\epsilon(t+1) &= \hat{g}_\epsilon(t)+\frac{\tilde{\delta}_\epsilon}{2}\left\{-\hat{g}_\epsilon(t)+[\hat{s}(t|t)-z(t)]^2+\frac{\hat{g}_\epsilon(t)}{\eta(t)}[\hat{\boldsymbol{\alpha}}^\top(t)\Lambda_{t-1}\hat{\boldsymbol{\alpha}}(t)+\hat{g}_s(t)]\right\} \\
&= \hat{g}_\epsilon(t)+\frac{\tilde{\delta}_\epsilon}{2}\Bigg\{-\hat{g}_\epsilon(t)+\Bigg[-\hat{\boldsymbol{\alpha}}^\top(t)\hat{\mathbf{s}}_{p-1}(t-1|t-1) \\
&\qquad +\frac{z(t)+\hat{\boldsymbol{\alpha}}^\top(t)\hat{\mathbf{s}}_{p-1}(t-1|t-1)}{\eta(t)}[\hat{\boldsymbol{\alpha}}^\top(t)\Lambda_{t-1}\hat{\boldsymbol{\alpha}}(t)+\hat{g}_s(t)] \\
&\qquad -z(t)\Bigg]^2+\frac{\hat{g}_\epsilon(t)}{\eta(t)}[\hat{\boldsymbol{\alpha}}^\top(t)\Lambda_{t-1}\hat{\boldsymbol{\alpha}}(t)+\hat{g}_s(t)]\Bigg\} \\
&= \hat{g}_\epsilon(t)+\frac{\tilde{\delta}_\epsilon}{2}\Bigg\{-\hat{g}_\epsilon(t)+\left[-\frac{\hat{g}_\epsilon(t)}{\eta(t)}[z(t)+\hat{\boldsymbol{\alpha}}^\top(t)\hat{\mathbf{s}}_{p-1}(t-1|t-1)]\right]^2 \\
&\qquad +\frac{\hat{g}_\epsilon(t)}{\eta(t)}[\eta(t)-\hat{g}_\epsilon(t)]\Bigg\} \\
&= \hat{g}_\epsilon(t)+\frac{\tilde{\delta}_\epsilon}{2}\frac{\hat{g}_\epsilon^2(t)}{\eta^2(t)}\left\{[z(t)+\hat{\boldsymbol{\alpha}}^\top(t)\hat{\mathbf{s}}_{p-1}(t-1|t-1)]^2-\eta(t)\right\} \qquad \text{(B.3)}
\end{aligned}$$

Substituting (208) and (217) into (B.3) immediately yields (216).

Bibliography

[1] A.P. Dempster, N.M. Laird, and D.B. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm," J. Roy. Stat. Soc., Ser. 3g, pp. 1–38, 1977.

[2] M. Feder, A.V. Oppenheim, E. Weinstein, "Maximum Likelihood Noise Cancellation Using the EM Algorithm," IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. ASSP-37, No. 2, pp. 204–216, February 1989.

[3] J.S. Lim and A.V. Oppenheim, "All Pole Modelling of Degraded Speech," IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. ASSP-26, pp. 197–210, 1978.

[4] R.A. Fisher, "Theory of Statistical Estimation," Proc. Cambridge Phil. Soc., Vol. 22, pp. 700–725, 1925.

[5] T.A. Louis, "Finding the Observed Information Matrix when using the EM Algorithm." Jour. Roy. Stat. Soc. Ser. B, Vol. 44, No. 2, pp. 226–233, 1982.

[6] J. Meilijson, "A Fast Improvement to the EM Algorithm On Its Own Terms," to be published in Jour. Roy. Stat. Soc. Ser. B.

[7] M. Segal and E. Weinstein, "A New Method for Evaluationg the Log-Likelihood Gradient (Score) of Linear Dynamic Systems," IEEE Trans. on Automatic Control, Vol. 33, No. 8, pp. 763–766, August 1988.

[8] M. Segal and E. Weinstein, "A New Method for Evaluating the Score, Heissian, and the Fisher's Information Matrix of Discrete-Time Dynamic Systems," IEEE Trans. on Information Theory, Vol. 35, No. 3, pp. 682–687, May 1989.