MULTIRATE LINEAR PREDICTION

Charles W. Therrien

Department of Electrical and Computer Engineering, Naval Postgraduate School
Monterey, CA 93943, USA
phone: + 831.656.3347, fax: + 831.656.2760, email: therrien@nps.edu
http://web.nps.navy.mil/˜therrien

ABSTRACT

The problem of linear prediction for multiple channels of data sampled at different rates is considered. The prediction error filters in this case are periodically time-varying. A Levinson-type recursion is developed for the linear prediction parameters. A sliding window formulation is also developed.

1. INTRODUCTION

In multirate optimal filtering problems [1], we are concerned with the estimation of a random process $d[n]$ from one or more observations processes $x_1, x_2, \ldots$ where all of the observation sequences and the desired process d may be sampled at different rates. A typical example is shown in Fig. 1, where the dots represent samples of the sequences in time.

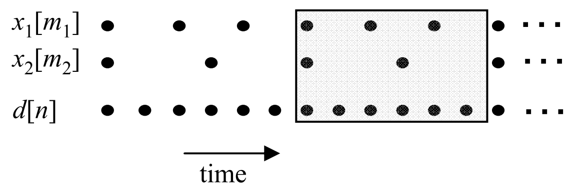

Figure 1: Multirate optimal filtering.

We are generally concerned with *causal* filtering, i.e., estimation of d using samples of the observations occuring only up to the current time. The problem is therefore both *multirate* and *multichannel* with respect to the input. The problem of multiple outputs could also be considered, but this is a trivial extension.

If the input and output sampling rates are integer-valued, then the structure of the observations is periodic. Further, if the input sequences are jointly cyclostationary (see [2] for a definition) then the optimal linear filter is *periodically time-varying*. In this case the set of observations can be grouped into blocks as shown in Fig. 1 so the problem has structure *within* blocks as well as *between blocks*, which we can exploit. In this paper, we consider the problem of joint *linear prediction* of the inputs. Since linear prediction results in orthogonalization of the input sequences, it underlies efficient implementation for more general optimal linear filtering problems.

2. LINEAR PREDICTION

For the linear prediction problem, we are concerned with only the input channels $x_1, x_2, \ldots$ shown in Fig. 1. The goal is to predict the next input observation from previous observations in all channels. In most cases this is a single observation, but in some cases observations from more than one channel can occur simultaneously. As we have noticed, there is a block stucture asssociated with this problem. Without considering the specific stucture of observations within one block, let P be the number of observations within a block and Q denote the number of full blocks that are to be used in the prediction.

We first consider the prediction of an entire block of observations at once. Let $\mathbf{v}[m]$ denote the vector of observations within the m^{th} block. The components of $\mathbf{v}[m]$ are assumed to be ordered in time from latest to earliest, where an arbitrary choice is made when more than one observation occurs simultaneously. The prediction equation for the blocks is then given by

$$\hat{\mathbf{v}}[m] = -\mathbf{A}_1'^T\mathbf{v}[m-1] - \mathbf{A}_2'^T\mathbf{v}[m-2]\cdots - \mathbf{A}_Q'^T\mathbf{v}[m-Q]$$

where ^{T} denotes transpose and the $\mathbf{A}_j'$ satisfy the block Normal equations[1]

$$\begin{bmatrix} \mathbf{R}_0 & \mathbf{R}_1 & \cdots & \mathbf{R}_Q \\ \mathbf{R}_{-1} & \mathbf{R}_0 & \cdots & \mathbf{R}_{Q-1} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{R}_{-Q} & \mathbf{R}_{1-Q} & \cdots & \mathbf{R}_0 \end{bmatrix} \begin{bmatrix} \mathbf{I} \\ \mathbf{A}_1' \\ \vdots \\ \mathbf{A}_Q' \end{bmatrix} = \begin{bmatrix} \mathbf{E}' \\ \mathbf{O} \\ \vdots \\ \mathbf{O} \end{bmatrix} \quad (1)$$

where $\mathbf{R}_j = E\{\mathbf{v}[m]\mathbf{v}^T[m-j]\}$ and $\mathbf{E}'$ is the prediction error covariance matrix. We denote the columns of $\mathbf{A}_j'$ by $\mathbf{a}_{i,j}'$, ordered as follows:

$$\mathbf{A}_j' = \begin{bmatrix} \mathbf{a}_{P-1,j}' & \mathbf{a}_{P-2,j}' & \cdots & \mathbf{a}_{0,j}' \end{bmatrix}, \quad j = 1, 2, \ldots, Q$$

We next consider the problem of predicting each of the points within a block, using the points within that block that occur earlier in time as well as *all* of the points in the Q previous blocks. We are then led to the following set of equations

$$\mathbf{R} \begin{bmatrix} \mathbf{A}_0 \\ \mathbf{A}_1 \\ \vdots \\ \mathbf{A}_Q \end{bmatrix} = \begin{bmatrix} \mathbf{E} \\ \mathbf{O} \\ \vdots \\ \mathbf{O} \end{bmatrix} \quad (2)$$

where $\mathbf{R}$ is the correlation matrix that occurs in (1) above, and

$$\mathbf{A}_j = \begin{bmatrix} \mathbf{a}_{P-1,j} & \mathbf{a}_{P-2,j} & \cdots & \mathbf{a}_{0,j} \end{bmatrix}, \quad j = 0, 1, \ldots, Q$$

[1] We assume that the observation processes are jointly stationary; therefore the correlation matrix in (1) is block Toeplitz.

The matrix $\mathbf{A}_0$ has the unit lower triangular form

$$\mathbf{A}_0 = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ \alpha_{P-1,1} & 1 & \cdots & 0 \\ \alpha_{P-1,2} & \alpha_{P-2,1} & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ \alpha_{P-1,P-1} & \alpha_{P-2,P-2} & \cdots & 1 \end{bmatrix}$$

while $\mathbf{E}$ is upper triangular

$$\mathbf{E} = \begin{bmatrix} \sigma^2_{P-1} & \times & \cdots & \times \\ 0 & \sigma^2_{P-2} & \cdots & \times \\ \vdots & \vdots & \ddots & \times \\ 0 & 0 & \cdots & \sigma^2_0 \end{bmatrix}$$

with off-diagonal values $\times$ that are typically non-zero, but do not concern us.

Now, let the vector of nonzero terms in the i^{th} column of $\mathbf{A}_0$ be denoted by α_{P-i} Then the group of terms, which we denote by a vector with a single index

$$\mathbf{a}_i = \begin{bmatrix} \alpha^T_{i-1} & \mathbf{a}^T_{i-1,1} & \mathbf{a}^T_{i-1,2} & \cdots & \mathbf{a}^T_{i-1,Q} \end{bmatrix}^T, \quad i = 1,\dots,P \tag{3}$$

is the set of coefficients needed to generate the prediction error for the i^{th} point in the block from the previous data as we have described it. Note that $\mathbf{a}_i$ corresponds to column $P-i+1$ on the left of (2) (without the zeros), and σ^2_{i-1} is the prediction error variance.

3. SOLVING FOR THE MULTIRATE LINEAR PREDICTION PARAMETERS

To find an efficient solution to the equations (2) for the linear prediction parameters, multiply (1) by $\mathbf{A}_0$ to obtain:

$$\mathbf{R} \begin{bmatrix} \mathbf{A}_0 \\ \mathbf{A}'_1\mathbf{A}_0 \\ \vdots \\ \mathbf{A}'_Q\mathbf{A}_0 \end{bmatrix} = \begin{bmatrix} \mathbf{E}'\mathbf{A}_0 \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix}$$

This equation is of the form (2); thus (2) has the solution

$$\mathbf{A}_j = \mathbf{A}'_j\mathbf{A}_0, \quad j = 1,\dots,Q \tag{4}$$

if

$$\mathbf{E}'\mathbf{A}_0 = \mathbf{E} \tag{5}$$

By considering the forms of $\mathbf{A}_0$ and $\mathbf{E}$ (above), we can recognize (5) as the set of Normal equations for another linear prediction problem. Specifically, if $\mathbf{e}[m]$ is the prediction error vector corresponding to the block prediction of $\mathbf{v}[m]$, then the solution of (5) provides the terms necessary for prediction of the elements of the $\mathbf{e}[m]$ *within* the block. Algebraically, (5) corresponds to the triangular decomposition problem for the positive semi-definite error covariance matrix $\mathbf{E}'$.

An efficient solution for the multirate linear prediction coefficients $\mathbf{a}_i$ and the prediction error variances σ^2_{i-1} can be obtained by using the multichannel Levinson recursion (see appendix), solving (5) by any method, and then computing the terms $\mathbf{a}_{i,j}$ in (3) from the relation

$$\mathbf{a}_{i,j} = \mathbf{A}'_j\mathbf{a}_{i,0}, \quad i = 0,1,\dots,P-1, \quad j = 1,\dots,Q \tag{6}$$

The prediction error variances are just the diagonal elements of $\mathbf{E}$.

4. CASE OF SIMULTANEOUS OBSERVATIONS

A slight modification needs to be made when two or more observations occur simultaneously. Consider the example shown in Fig. 1. The first observations of x_1 and x_2 within the block occur simultaneously. While one could arbitarily delay one of the observations by a small amount δ and use the preceeding theory, it is better to consider prediction of the two observations at once. In this case, the number of points P in the block is 5, and the matrices $\mathbf{A}_0$ and $\mathbf{E}$ become

$$\mathbf{A}_0 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ \alpha_{4,1} & 1 & 0 & 0 & 0 \\ \alpha_{4,2} & \alpha_{3,1} & 1 & 0 & 0 \\ \alpha_{4,3} & \alpha_{3,2} & \alpha_{2,1} & 1 & 0 \\ \alpha_{4,4} & \alpha_{3,3} & \alpha_{2,2} & 0 & 1 \end{bmatrix}$$

and

$$\mathbf{E} = \begin{bmatrix} \sigma^2_4 & \times & \times & \times & \times \\ 0 & \sigma^2_3 & \times & \times & \times \\ 0 & 0 & \sigma^2_2 & \times & \times \\ 0 & 0 & 0 & \sigma^2_1 & \rho_{1,0} \\ 0 & 0 & 0 & \rho_{0,1} & \sigma^2_0 \end{bmatrix}$$

In general, a block of $\mathbf{A}_0$ is replaced by the identity matrix, and the corresponding block of $\mathbf{E}$ is replaced by a small covariance matrix $\mathbf{\Sigma}$.

5. STRUCTURE OF THE MULTIRATE PREDICTION ERROR FILTER

While (6) provides and explicit solution for the linear prediction parameters, an interesting structure for the multirate prediction error filter arises if the inter-block and intra-block computations are performed separately. Figure 2 shows this

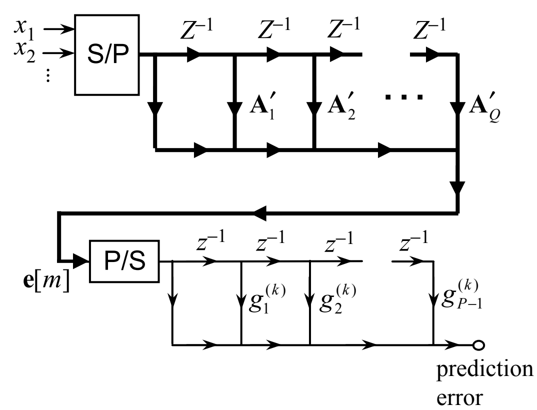

Figure 2: Form of the multirate prediction error filter.

structure (assuming no simultaneous observatons). Conceptually, the inputs $x_1, x_2, \dots$ for one block are brought into a register to form the vector $\mathbf{v}[m]$. This is labeled in the figure as serial/parallel (S/P) conversion. The blocks are then processed through the block prediction error filter to form the block error $\mathbf{e}[m]$. (Note that this filter could be implemented in block lattice form instead of the direct form shown here.) The block errors are then converted to serial form (P/S) and processed through a within-block prediction error filter. This filter, which is also shown in direct form, has time-varying gains $g_i^{(k)}$ whose values change according to the table below:

k	$g_1^{(k)}$	$g_2^{(k)}$	$\cdots$	$g_{P-1}^{(k)}$
0	0	0	$\cdots$	0
1	$\alpha_{1,1}$	0	$\cdots$	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$P-1$	$\alpha_{P-1,1}$	$\alpha_{P-1,2}$	$\cdots$	$\alpha_{P-1,P-1}$

The advantage of the form in Fig. 2 is that only the *within*-block parameters need to be time-varying. Thus one can save on storage requirements for the most of the parameters at the expense of a few additional multiplications and additions.

Although the filter structure is shown with explicit blocks representing the S/P and P/S conversions, such explicit blocks are not in fact necessary. Upon careful examination of the structure one can observe that computations can occur as soon as each new observation x is available; so the data can be pipe-lined through the filter. This provides an additonal advantage.

6. SLIDING WINDOW PREDICTOR

The prediction error filter just described has a variable prediction order as one proceeds through the block because we always use the full number of points in the last (Q^{th}) block. Although this does not usually present any problem in practice, it is still of interest to explore a linear prediction problem where the order remains fixed, although the filter is time-varying. The situation is depicted in Fig. 3. Here, as we pre-

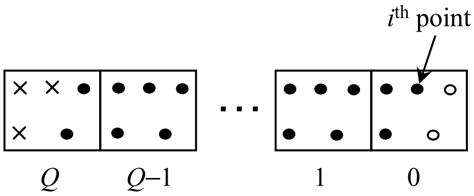

Figure 3: Sliding window prediction error filter.

dict the i^{th} point of the current block, a corresponding number of i older points in the last block are dropped out. These points are indicated by $\times$'s in Fig. 3.

Now notice that if we have the solution for the block linear prediction problem (1), where there is a *reduced* number of points in the last (Q^{th}) block, then the procedure of section 3 can be applied to obtain the desired sliding window parameters. To formulate this block linear prediction problem, let us replace the unknowns in (1) by $\mathbf{A}''_j$, $j = 1, 2, \ldots, Q$ and $\mathbf{E}''$, where $\mathbf{A}''_Q$ is a "short" block, to obtain

$$\mathbf{R}\begin{bmatrix} \mathbf{I} \\ \mathbf{A}''_1 \\ \vdots \\ \mathbf{A}''_{Q-1} \\ \mathbf{A}''_Q \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} \mathbf{E}'' \\ \mathbf{O} \\ \vdots \\ \mathbf{O} \\ \mathbf{0} \\ \times \end{bmatrix} \qquad (7)$$

Note that the last i rows of the $\mathbf{A}$ matrix are constrained to be zero and that the corresponding rows of the matrix on the right have values that do not concern us.

The efficient solution to (7) requires the consideration of the *backward* block preciction problem of order $Q-1$ where the points in block Q are predicted from the points in blocks 1 through $Q-1$. The Normal equations for this backward block prediction problem can be writtten as

$$\mathbf{R}^{(Q-1)}\begin{bmatrix} \mathbf{B}'^{(Q-1)}_{Q-1} \\ \vdots \\ \mathbf{B}'^{(Q-1)}_1 \\ \mathbf{I} \end{bmatrix} = \begin{bmatrix} \mathbf{O} \\ \vdots \\ \mathbf{O} \\ \mathbf{E}'^{(Q-1)}_b \end{bmatrix}$$

where the superscript denotes the order. Observe that the matrix $\mathbf{R}^{(Q-1)}$ can be obtained from the matrix $\mathbf{R}$ appearing in (1) and (2) by striking out the first row and column, and that the solution to this equation gives the parameters for estimating the *full* set of points in the Q^{th} block.

Now, suppose i points are removed from the last block and we wish to estimate only the $k = P - i$ most recently occuring points using all points in blocks Q through 1 that occur *after* the point to be estimated. According to the procedures developed in sections 2 through 4, we are led to consider the problem

$$\mathbf{E}'^{(Q-1)}_b \begin{bmatrix} \mathbf{I} & \beta \\ \mathbf{0} & \mathbf{I} \end{bmatrix} = \begin{bmatrix} \mathbf{E}_0 & \mathbf{0} \\ \mathbf{E}_1 & \mathbf{\Sigma}_{bb} \end{bmatrix} \qquad (8)$$

where β is a $k \times i$ matrix and $\mathbf{\Sigma}_{bb}$ is the $i \times i$ prediction covariance matrix for the i removed points. The terms $\mathbf{E}_0$ and $\mathbf{E}_1$ are simply appropriate partitions of $\mathbf{E}'^{(Q-1)}_b$. By combining these last two equations, we arrive at the desired Normal equations for backward prediction:

$$\mathbf{R}^{(Q-1)}\begin{bmatrix} \mathbf{B}_{Q-1} \\ \vdots \\ \mathbf{B}_1 \\ \mathbf{B}_0 \end{bmatrix} = \begin{bmatrix} \mathbf{O} \\ \vdots \\ \mathbf{O} \\ \mathbf{E}_b \end{bmatrix}$$

where

$$\begin{aligned} \mathbf{B}_j &= \mathbf{B}'^{(Q-1)}_j \mathbf{B}_0, \quad j = 1, \ldots, Q-1 \qquad (9) \\ \mathbf{E}_b &= \mathbf{E}'^{(Q-1)}_b \mathbf{B}_0 = \begin{bmatrix} \mathbf{E}_0 & \mathbf{0} \\ \mathbf{E}_1 & \mathbf{\Sigma}_{bb} \end{bmatrix}, \quad \mathbf{B}_0 = \begin{bmatrix} \mathbf{I} & \beta \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \end{aligned}$$

In order to find the $\mathbf{A}''_j$ in (7) we write them as a linear combination

$$\begin{bmatrix} \mathbf{I} \\ \mathbf{A}''_1 \\ \vdots \\ \mathbf{A}''_{Q-1} \\ \mathbf{A}''_Q \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} \mathbf{I} \\ \mathbf{A}'^{(Q-1)}_1 \\ \vdots \\ \mathbf{A}'^{(Q-1)}_{Q-1} \\ \mathbf{O} \end{bmatrix} - \begin{bmatrix} \mathbf{O} \\ \mathbf{B}_{Q-1} \\ \vdots \\ \mathbf{B}_1 \\ \mathbf{B}_0 \end{bmatrix} \mathbf{G} \qquad (10)$$

where the $\mathbf{A}'^{(Q-1)}_j$ are the solution to the block Normal equations (1) when the order is $Q-1$, and $\mathbf{G}$ is a square matrix to be determined. The right hand side of (7) is then written as

$$\begin{bmatrix} \mathbf{E}'' \\ \mathbf{O} \\ \vdots \\ \mathbf{O} \\ \mathbf{0} \\ \times \end{bmatrix} = \begin{bmatrix} \mathbf{E}'^{(Q-1)} \\ \mathbf{O} \\ \vdots \\ \mathbf{O} \\ \\ \mathbf{\Delta} \end{bmatrix} - \begin{bmatrix} \mathbf{D} \\ \mathbf{O} \\ \vdots \\ \mathbf{O} \\ \\ \mathbf{E}_b \end{bmatrix} \mathbf{G} \qquad (11)$$

where $\mathbf{E}'^{(Q-1)}$ is the prediction error covariance matrix for block prediction of order $Q-1$ and

$$\mathbf{\Delta} = \sum_{j=0}^{Q-1} \mathbf{R}_{-Q+j}\mathbf{A}_j'^{(Q-1)}, \qquad \mathbf{D} = \sum_{j=1}^{Q} \mathbf{R}_j \mathbf{B}_{Q-j} \qquad (12)$$

(It can also be shown that $\mathbf{D} = \mathbf{\Delta}^T \mathbf{B}_0$.)

Now, let us partition $\mathbf{G}$ and $\mathbf{\Delta}$ into blocks of k and i rows:

$$\mathbf{G} = \begin{bmatrix} \mathbf{G}_0 \\ \mathbf{G}_1 \end{bmatrix}, \quad \mathbf{\Delta} = \begin{bmatrix} \mathbf{\Delta}_0 \\ \mathbf{\Delta}_1 \end{bmatrix}$$

The condition that the solution to (7) be in the form (10) requires that $\mathbf{G}_1 = \mathbf{0}$. The condition that the right hand side of (7) be of the form (11) can be satisfied if we choose

$$\mathbf{G}_0 = \mathbf{E}_0^{-1}\mathbf{\Delta}_0 \qquad (13)$$

This completes the solution of (7) using (10) and (11).

7. SUMMARY OF SLIDING WINDOW ALGORITHM

The following summarizes the necessary steps to compute the sliding window parameters.

1. Find the forward and backward block linear prediction parameters up to order Q using the multichannel Levinson recursion (see appendix).
2. For $i = 1, 2, \dots P$:
   (a) Solve (8) and apply (9) to obtain $\mathbf{B}_j$ and $\mathbf{E}_b$. (This can be done recursively on i.)
   (b) Compute $\mathbf{D}$ from (12) and $\mathbf{G}_0$ from (13).
   (c) Compute $\mathbf{A}_j''$ and $\mathbf{E}''$ from (10) and (11).
   (d) Solve for the sliding window parameters by substituting $\mathbf{A}_j''$ for $\mathbf{A}_j'$ and $\mathbf{E}''$ for $\mathbf{E}'$ in section 3 and using (4), (5) and (6).

8. CONCLUSIONS

In previous work we have considered the problem of optimal filtering for sets of observation sequences sampled at different rates and observed that the optimal linear filter is periodically time-varying. In this paper a first attempt is made to separate the filter into time-varying and non-time-varying parts for efficient implementation and estimation of the filter parameters. In this paper we have chosen to examine the linear prediction (self-orthogonalization) problem for the input sequences since this problem is fundamental to more general optimal filtering problems.

It is shown that if we allow the prediction to be based on a fixed number of full data blocks and a variable number of data points within the most recent block, then the problem separates into a inter-block linear prediction problem followed by an intra-block linear prediction problem and the multichannel Levinson recursion can be used to find the filter parameters. The filter realization is correspondingly separable into time-varying and non-time-varying parts as shown in Fig. 2.

While the variable order of the filter is probably not objectionable in most practical problems, the requirement to use a filter of fixed order with a sliding window of observations seems to be at least of theoretical interest. We have examined this latter problem and found that the solution involves a certain combination of forward and backward linear prediction problems both between blocks as well as within blocks. The algorithm derived for computing the sliding window linear prediction parameters also involves use of the multichannel Levinson algorithm.

APPENDIX

Multichannel Levinson Recursion [3]

$$\mathbf{\Delta}^{(q)} = \sum_{j=0}^{q-1} \mathbf{R}_{-q+j}\mathbf{A}_j'^{(q-1)} \qquad \mathbf{\Delta}_b^{(q)} = \sum_{j=1}^{q} \mathbf{R}_j \mathbf{B}_{q-j}'^{(q-1)}$$

$$\mathbf{\Gamma}^{(q)} = \left(\mathbf{E}_b'^{(q-1)}\right)^{-1} \mathbf{\Delta}^{(q)} \qquad \mathbf{\Gamma}_b^{(q)} = \left(\mathbf{E}'^{(q-1)}\right)^{-1} \mathbf{\Delta}_b^{(q)}$$

$$\begin{bmatrix} \mathbf{I} \\ \mathbf{A}_1'^{(q)} \\ \vdots \\ \\ \mathbf{A}_q'^{(q)} \end{bmatrix} = \begin{bmatrix} \mathbf{I} \\ \mathbf{A}_1'^{(q-1)} \\ \vdots \\ \mathbf{A}_{q-1}'^{(q-1)} \\ \mathbf{O} \end{bmatrix} - \begin{bmatrix} \mathbf{O} \\ \mathbf{B}_{q-1}'^{(q-1)} \\ \vdots \\ \mathbf{B}_1'^{(q-1)} \\ \mathbf{I} \end{bmatrix} \mathbf{\Gamma}^{(q)}$$

$$\begin{bmatrix} \mathbf{I} \\ \mathbf{B}_1'^{(q)} \\ \vdots \\ \\ \mathbf{B}_q'^{(q)} \end{bmatrix} = \begin{bmatrix} \mathbf{I} \\ \mathbf{B}_1'^{(q-1)} \\ \vdots \\ \mathbf{B}_{q-1}'^{(q-1)} \\ \mathbf{O} \end{bmatrix} - \begin{bmatrix} \mathbf{O} \\ \mathbf{A}_{q-1}'^{(q-1)} \\ \vdots \\ \mathbf{A}_1'^{(q-1)} \\ \mathbf{I} \end{bmatrix} \mathbf{\Gamma}_b^{(q)}$$

$$\mathbf{E}'^{(q)} = \mathbf{E}'^{(q-1)} \left(\mathbf{I} - \mathbf{\Gamma}_b^{(q)}\mathbf{\Gamma}^{(q)}\right)$$

$$\mathbf{E}_b'^{(q)} = \mathbf{E}_b'^{(q-1)} \left(\mathbf{I} - \mathbf{\Gamma}^{(q)}\mathbf{\Gamma}_b^{(q)}\right)$$

Note: $\mathbf{\Delta}_b^{(q)} = (\mathbf{\Delta}^{(q)})^T$

REFERENCES

[1] R. J. Kuchler and C. W. Therrien, "Optimal filtering with multirate observations," in *Proceedings of the* 37^{th} *Asilomar Conference on Signals, Systems, and Computers*, November 2003, pp. 1208–1212, (Pacific Grove, CA).

[2] C. W. Therrien, "Defining correlation functions and power spectra for multirate random processes," in *Proceedings of the IEEE International Symposium on Circuits and Systems*. IEEE Circuits and Systems Society, May 2005, Kobe, Japan.

[3] A. Lindquist, "A new algorithm for optimal filtering of discrete-time stationary processes," *SIAM Journal, Control*, vol. 12, no. 4, pp. 736–746, November 1974.