SYSTEM IDENTIFICATION FROM MULTI-RATE DATA

R. Bhushan Gopaluni * **Harigopal Raghavan** *
Sirish L. Shah *

* *Department of Chemical & Materials Engineering, University of Alberta, Edmonton, AB, CANADA - T6G 2G6*

Abstract: In this paper, we provide a novel iterative identification algorithm for multi-rate sampled data systems. The procedure involves, as a first step, identifying a simple initial model from multi-rate data. Based on this model, the "missing" data points in the slow sampled measurements are estimated following the expectation maximization approach. Using the estimated missing data points and the original data set, a new model is obtained and this procedure is repeated until the models converge. An attractive feature of the proposed method lies in its applicability to irregularly sampled data. An application of the proposed method to an industrial data set is also included.

Keywords: identification, multi-rate processes, expectation maximization algorithm

1. INTRODUCTION

Traditional identification methods assume that the data are sampled at uniformly spaced sample instants. There is extensive literature on identification of processes from such data (Ljung (1999)). However, in many chemical processes it is either not physically possible to measure certain variables at regular intervals or it is impractical to have frequent or rapid lab assays done. For instance, estimating the composition of the distillate in a distillation column generally takes a few minutes while control moves are implemented at much smaller sample intervals. Such processes with differing sample times for the measured variables are termed multi-rate processes in the rest of this paper. In particular, identification of models for multi-rate processes at the fastest sample rate is the subject of this paper. We refer to the fastest sample rate as the base sample rate and the unavailable data points in the slowly sampled measurements as *missing data*. This allows us to accommodate systems in which outputs are irregularly sampled within the same identification scheme.

A commonly used approach for the identification of processes from multi-rate data is to interpolate between available sampled data. Techniques such as linear or quadratic interpolation are used. Interpolations of these types do not take into account the variation in the input during the period over which the interpolations are made. There have also been attempts at solving the multi-rate identification problem using lifting techniques (Li et al. (2001)). The lifting operator is used to convert the multi-rate identification problem into a multivariable identification problem. However, applying these techniques towards process identification in chemical industries is not easy because the difference in the sampling rates is generally large. For example, (input to output) sampling ratios of $1:15$ are common and estimation techniques to identify a 15-input 'lifted' process would have to be considered. Hence, for processes with a number of inputs and/or large output to input sampling ratios, the identification problem using lifting techniques can become unmanageable. In addition, these techniques are incapable of handling irregularly sampled data.

In this paper we present a method which uses an initial crude model to estimate the unavailable data points in the slowly sampled variables. The estimated unavailable data points are then used with original data set to identify a new model. From the new model,

the missing data is again estimated and this process is repeated until the models converge. This approach reduces to the *Expectation Maximization* (EM) algorithm if optimal estimates of the missing data points are used in the estimation stage. The advantage of this method lies in the methodical manner in which the missing data points are estimated instead of using the traditional interpolation methods. Instead of interpolating, the missing data are estimated based on the current estimate of the process model at each iteration. Use of the EM algorithm guarantees convergence and consistency of the identified models (Dempster et al. (1977)).

The rest of this paper is organized as follows: section 2 lists the assumptions and the notation. In section 3, the EM algorithm is presented and in section 4 a method for identification of linear dynamical systems using the EM approach is developed. In section 5 the EM based identification method is extended to the missing data case. An industrial example is presented in section 6 followed by concluding remarks in section 7.

2. ASSUMPTIONS AND NOTATION

Let us assume that the true process is of the form

$$\begin{aligned} x_{t+1} &= Ax_t + Bu_t + w_t \\ y_t &= Cx_t + v_t \end{aligned} \tag{1}$$

where A, B, C are the system matrices and $x_t \in \mathbb{R}^n$ is the n-dimensional state vector. Assume that $u(t) \in \mathbb{R}^m$ and $y(t) \in \mathbb{R}^p$. w_t and v_t are uncorrelated white noise sequences *i*.e.,

$$\begin{aligned} E[w_t w_t^T] = Q\,; E[w_t] = 0 \qquad \forall\, t \\ E[v_t v_t^T] = R\,; E[v_t] = 0 \qquad \forall\, t \\ E[w_t v_t] = 0 \qquad \forall\, t \end{aligned} \tag{2}$$

Let us represent the time series data from $t = 1$ to $t = N$ of any variable by $(.)_{1:N}$. Through out this paper, we will use the following notation for the expected values of various variables,

$$x_t^s := E(x_t | Y_{1:s}) \tag{3}$$

and

$$\begin{aligned} P_t^s &:= E(x_t - x_t^s)(x_t - x_t^s)^T \\ P_{t,t-1}^s &:= E(x_t - x_t^s)(x_{t-1} - x_{t-1}^s)^T \end{aligned}$$

In addition, the following assumptions are made:

Assumptions

A1. Inputs are sampled uniformly every T units of time.

A2. Outputs are sampled at $T_1, \cdots, T_n$ respectively.

A3. The input sampling time, T, is assumed to be the smallest sampling time *i.e.,*

$$T \leq T_i \qquad \forall\ i \tag{4}$$

A4. Assume that the initial state is zero *i.e.,* $x_0 = 0$.

3. THE EM ALGORITHM

The central idea behind the algorithm presented in this paper is to pose the multi-rate identification problem in the maximum likelihood framework and solve for the system matrices. The iterative algorithm presented in this section is based on the popular *Expectation Maximization algorithm*(EM algorithm) developed in Dempster et al. (1977). Before utilizing this algorithm in identifying multi-rate processes, a brief summary of the algorithm is presented below.

The EM algorithm addresses the problem of estimating model parameters under the maximum likelihood framework. More often than not, the maximum likelihood function is a complicated nonlinear function of the unknown parameters. Hence, one of the earliest methods proposed for solving for the optimal parameters was to use the Newton-Raphson method (Gupta and Mehra (1974)). A simpler method based on the EM algorithm was proposed in Shumway and Stoffer (1982).

The EM algorithm can be summarized in the following few steps :

- Obtain an initial estimate of the parameter vector, Θ^0.
- Carry out the following steps at each iteration, k, until convergence:
  - **Expectation (E-step)**: Find the expected value of the complete data log likelihood function(Q-function) given the observed data set and the previously estimated parameter vector, Θ^k.
  - **Maximization (M-step)**: Maximize the Q-function with respect to the parameter vector

The above steps ensure that the log likelihood function of the observed data increases at every iteration. Therefore, the EM algorithm is guaranteed to converge to a local maximum of the likelihood function. This is an important feature of the EM algorithm. However, there are a few drawbacks associated with any iterative algorithm. The EM algorithm can be sensitive to the initial guess and also the rate of convergence can sometimes be extremely slow. In order to avoid problems with a bad initial parameter guess, we identify an initial unbiased FIR model of the process. An example illustrating the use of EM algorithm in estimating models from multi rate data is presented below.

Example 1. Consider an ARX model

$$y(k) = 0.8y(k-1) + 0.3u(k-1) + e(k) \qquad (5)$$

where $e(k)$ is normally distributed white noise with variance $\sigma_e^2 = 0.01$. Let us assume that the output is sampled at every alternate sample instant and that $y(1)$ is known. Then the following identification objective function based on squared prediction errors can be used

$$\begin{aligned} V_N(\theta) &:= \frac{1}{N}\sum_{k=1}^{N} \varepsilon(t,\theta)^2 \\ &= \frac{1}{N}\sum_{k=1}^{N} [y(k) - \theta_2 y(k-1) - \theta_1 u(k-1)]^2 \end{aligned}$$

where N is the data length and $\theta = [\theta_1 \quad \theta_2]^T$. Since only alternate data points are available, the above objective function can not be evaluated. Instead, it is possible to estimate the expected value of the above objective function given the estimate of θ from the previous iteration, $\hat{\theta}^{(j-1)}$ *i.e.,*

$$\begin{aligned} E\left[V_N(\theta)|\hat{\theta}^{(j-1)}, Z_N\right] = E[\frac{1}{N}\sum_{k=1}^{N} [y(k) - \theta_2 y(k-1) \\ -\theta_1 u(k-1)]^2] \end{aligned} \qquad (6)$$

where Z_N denotes all the available data. Now let us consider two cases:
Case I: $y(k)$ is known, then

$$\begin{aligned} &E\,[y(k) - \theta_2 y(k-1) - \theta_1 u(k-1)]^2 = \\ &(y(k) - \theta_1 u(k-1))^2 + \theta_2^2(\hat{\theta}_1^{(j-1)} u(k-2) \\ &+\hat{\theta}_2^{(j-1)} y(k-2))^2 + \theta_2^2\sigma_e^2 \\ &-2(y(k) - \theta_1 u(k-1))\theta_2(\hat{\theta}_1^{(j-1)} u(k-2) \\ &+\hat{\theta}_2^{(j-1)} y(k-2)) \end{aligned} \qquad (7)$$

Case II: $y(k)$ is unknown, then

$$\begin{aligned} &E\,[y(k) - \theta_2 y(k-1) - \theta_1 u(k-1)]^2 = \\ &(\hat{\theta}_1^{(j-1)} u(k-1) + \hat{\theta}_2^{(j-1)} y(k-1))^2 \\ &+\sigma_e^2 + (\theta_1 u(k-1) + \theta_2 y(k-1))^2 \\ &-2(\theta_1 u(k-1) + \theta_2 y(k-1))(\hat{\theta}_1^{(j-1)} u(k-1) \\ &+\hat{\theta}_2^{(j-1)} y(k-1)) \end{aligned} \qquad (8)$$

Using (7) and (8) in (6) it is possible to find the model parameters at the current iteration, j,

$$\theta^{(j)} = \min_{\theta} E\left[V_N(\theta)|\hat{\theta}^{(j-1)}, Z_N\right] \qquad (9)$$

The iterations are performed until the parameters converge. A plot showing the two parameters in this example and the number of iterations is shown in fig.1. The estimated model parameters converge to the true parameters despite missing data. In general, the estimates using the EM algorithm need not converge to the true parameters with finite data sets. However, the estimated parameters converge to the true parameters asymptotically as the data length increases. On the other hand, the least squares model obtained by interpolating the data is $\hat{\theta}_1 = 0.83$ and $\hat{\theta}_2 = 0.24$, which is clearly not the true model. There is a small amount of bias in the estimated model using the interpolated data. In general, the estimated models are biased if arbitrary interpolation methods are used to fill the missing data points. ■

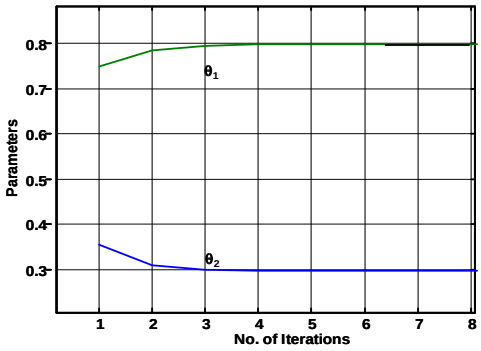

Fig. 1. *Plot of θ_1 and θ_2 as a function of number of iterations*

Now it is possible to use this algorithm to estimate the state matrices of a linear dynamical system described in (1). As a first step we provide an algorithm to estimate the model from the complete data set *i.e.,* there are no missing data. Even though the data set is complete the states are unknown/unobserved and hence the EM algorithm can be utilized. Once a method for identification of single rate systems is developed, it can be extended to multi-rate systems.

4. ESTIMATION OF LINEAR DYNAMICAL SYSTEMS

A maximum likelihood framework is adopted in this section to identify the system matrices of (1). Two independent sequences of noise enter the dynamical system in (1). Hence, the joint log likelihood function of the complete data set can be expressed as

$$\begin{aligned} &\log\mathscr{L}(y_{1:N}, x_{1:N}, \Theta) = \log\mathscr{L}(w_{1:N}, v_{1:N}, \Theta) \\ &= -\frac{N}{2}\log|Q| - \frac{N}{2}\log|R| \\ &-\frac{1}{2}\sum_{t=1}^{N}(x_t - Ax_{t-1} - Bu_{t-1})^T Q^{-1}(x_t - Ax_{t-1} - Bu_{t-1}) \\ &-\frac{1}{2}\sum_{t=1}^{N}(y_t - Cx_t)^T R^{-1}(y_t - Cx_t) \end{aligned}$$

where the parameter vector $\Theta = \{A, B, C, Q, R\}$. The Q-function can then be evaluated by taking the expectation of the log likelihood function given the observed data and parameters from the previous iterate (say Θ^k). Let us define the conditional expectation operator E_k as follows

$$E_k(.) = E(.|y_{1:N}, u_{1:N}, \Theta^k) \tag{10}$$

Now using the above expectation operator the Q-function can be evaluated,

$$\begin{aligned} Q(y_{1:N}, \Theta^k, \Theta) &= -\frac{N}{2}\log|Q| - \frac{N}{2}\log|R| \\ &-\frac{1}{2}\sum_{t=1}^{N} tr\{Q^{-1}E_k(x_t - Ax_{t-1} - Bu_{t-1})(x_t - Ax_{t-1} - Bu_{t-1})^T\} \\ &-\frac{1}{2}\sum_{t=1}^{N} tr\{R^{-1}E_k(y_t - Cx_t)(y_t - Cx_t)^T\} \end{aligned}$$

where $tr(.)$ denotes the trace of a matrix. At each iteration in the EM algorithm a new estimate of the model is obtained by maximizing $Q(y_{1:N}, \Theta^k, \Theta)$ with respect to Θ *i.e.*,

$$\Theta^{k+1} = \max_{\Theta} Q(y_{1:N}, \Theta^k, \Theta) \tag{11}$$

Complete details on obtaining the new estimate, Θ^{k+1} are given in the appendix.

***Example** 2.* Consider the following state space model

$$A = \begin{bmatrix} 0.3688 & 0.4767 & 0.0114 \\ -0.5976 & 0.6095 & -0.5408 \\ -0.0156 & -0.0686 & 0.0422 \end{bmatrix} \quad B = \begin{bmatrix} 0.34 \\ 0.56 \\ 0.78 \end{bmatrix}$$
$$C = [1.2 \quad 0.96 \quad 1.5] \tag{12}$$

with the true covariance matrices

$$Q = \begin{bmatrix} 0.0407 & 0.0001 & 0.0015 \\ 0.0001 & 0.0407 & -0.0020 \\ 0.0015 & -0.0020 & 0.0428 \end{bmatrix}; R = 0.398$$

Using the method proposed in this section one can estimate the model parameters. A plot showing the step responses of the true model, a model obtained using the subspace identification method - N4SID and the model obtained using the EM algorithm are shown in fig.2. The EM algorithm performs as well as the subspace method. The EM algorithm presented in this section, theoretically, will provide asymptotic unbiased estimates. However, in practice the algorithm may not converge fast enough or if a bad initial guess is given, it may converge to a local maximum. Hence, a good initial guess for the EM algorithm is needed. An unbiased least squares model can be used as the initial guess.

■

5. ESTIMATION WITH MISSING DATA

The strength of EM algorithm lies in being able to estimate asymptotically unbiased models even if a portion of the data is missing. As shown in the previous section, it is possible to use the EM algorithm

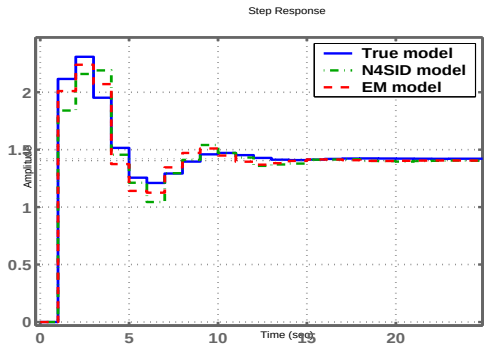

Fig. 2. *Step responses of - the true model, the N4SID model and EM model*

for identification of models from single rate data sets. However, the computational effort involved in using the EM algorithm is too heavy to warrant this method for single rate identification problems. Moreover, traditional identification methods can provide asymptotically unbiased estimates for single rate data sets. On the other hand, in general, identification methods involving arbitrary interpolations to substitute for missing data result in biased estimates; thus, necessitating the development of new methods for identification of models from multi-rate data.

It is interesting to note that the EM algorithm presented in the previous section for identification of linear dynamical systems from single rate data, treats the states as unknown/missing data. Hence, it is possible to extend the same algorithm to include the case of missing data in the outputs by making appropriate changes to the Kalman filter and the Kalman Smoother presented in the Appendix. Full details regarding these modifications can easily be derived along the lines of the arguments given in Shumway and Stoffer (2000).

The procedure can be summarized as follows:

Step 1: Obtain an initial estimate of the model. For instance, it is easy to obtain an FIR model.

Step 2: Estimate the missing data points using the initial estimate of the model. This can be done using the Kalman Filter and the Kalman Smoother.

Step 3: Predict all missing data points using the current model.

Step 4: Using the true and the estimated missing data points identify a new model by minimizing the Q-function.

Step 5: Repeat the above steps until convergence.

***Example** 3.* The process in example 2 is used to generate multi-rate data. The input is sampled every second and the output is sampled every four seconds. Initially, a model is identified using linearly interpolated data and the N4SID algorithm. Then the proposed method is used on the same data set without interpolating the data. The step responses of both the models are shown in fig.3. Clearly, the EM based method outperforms the N4SID method. ■

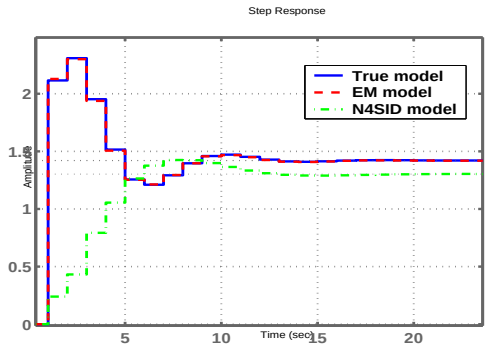

Fig. 3. *Step responses of - the true model, the N4SID model and EM model; Output sampled every four seconds*

6. INDUSTRIAL APPLICATION

In this application, modelling of a mechanical pulp bleaching process at Millar Western, Whitecourt, AB, Canada is shown. The system has four manipulated inputs, two measured disturbance variables and one output. The output, pulp brightness, is an irregularly measured quality variable (distribution of sampling intervals are provided in fig.4). The manipulated in-

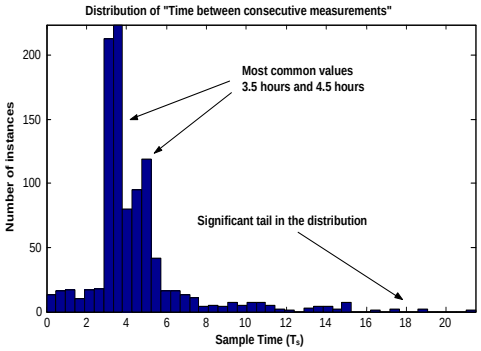

Fig. 4. *Distribution of "time between consecutive measurements"*

puts are chemical add-rates (Peroxide and Caustic) to two towers. The measured disturbances are two wood quality variables (Aspen and Freeness). All inputs are sampled every 10 minutes). The process is known to be a time-delay dominant recycle process. The step responses of the true model have large delay, fast dynamics and recycle characteristics. In general, the presence of a recycle stream can significantly alter the dynamics of a process (Morud and Skogestad (1994),Kwok et al. (2001)). This is especially evident when the process dynamics are faster than the time-delay effects in the process. For example, a step change in one of the inputs in a time-delay dominant recycle process, can cause a staircase-like structure in the output as shown in the fig.5.

When the time-delay in the system is greater than the settling time, including lagged inputs (the extra lags being equal to the sum of the delays in the forward path and the recycle path) as predictors can give a better model. In this particular modelling exercise, there was no provision for performing dynamic tests

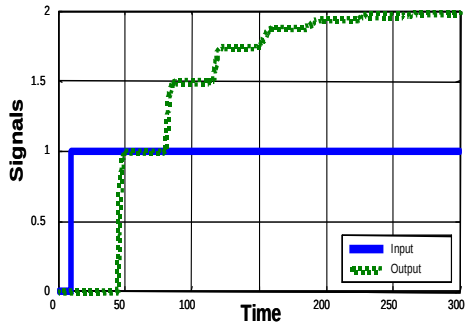

Fig. 5. *Step response of a Delay-dominant recycle process*

to aid model development. Hence we used routine operating data to perform model identification. The routine operating data has enough excitation in the form of grade changes to justify the exercise of model identification using this data.

We used the proposed method based on the EM algorithm for identifying the model. The predictions based on the EM model (without interpolation) and N4SID model (with zero order hold interpolation) are presented in fig.6. The models shown have been adjusted taking the recycle characteristics into account. Hence, only the forward path dynamics are shown. Though it appears that the EM model and the N4SID model perform comparably well for the given data set, it is clear from the step responses (fig.7) that the EM model is representative of the true process dynamics (fast dynamics).

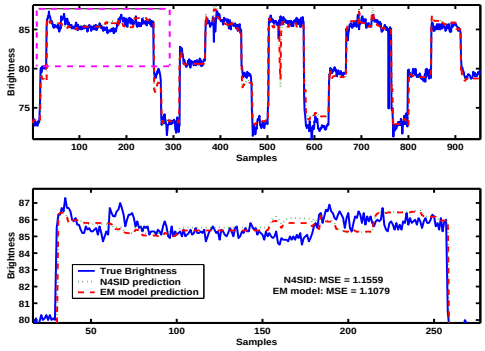

Fig. 6. *Comparison of N4SID and EM model predictions with actual brightness*

7. CONCLUSIONS

An identification approach for multi-rate data, based on the Expectation Maximization approach is presented. Unlike, traditional identification methods for multi-rate data, the proposed method does not use interpolation. An attractive feature of the algorithm is that it can easily handle irregularly sampled data. It leads to asymptotically unbiased estimates of the true model. However, the proposed method is sensitive to the initial guess and is computationally intensive.

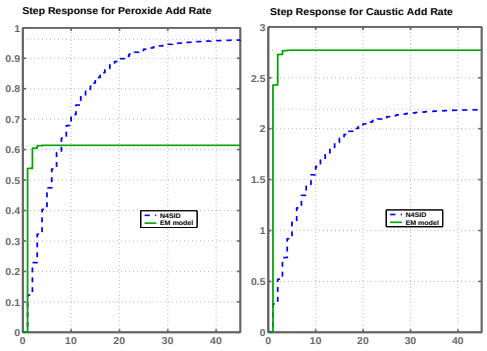

Fig. 7. *Step responses of the N4SID and EM models*

APPENDIX : ESTIMATES OF SYSTEM MATRICES AT EACH ITERATION

Let us first evaluate the second term in the Q-function

$$\begin{aligned} T_2 &:= -\frac{1}{2}\sum_{t=1}^{N} tr\{Q^{-1}E_k(x_t - Ax_{t-1} - Bu_{t-1}) \\ &\quad (x_t - Ax_{t-1} - Bu_{t-1})^T\} \\ &= -\frac{1}{2}tr\{Q^{-1}\left[\Phi_1 + A\Phi_2 A^T + B\Phi_4 B^T + 2\Phi_3 A^T\right. \\ &\quad \left. + 2\Phi_5 B^T - 2A\Phi_6 B^T\right]\} \end{aligned} \tag{A-1}$$

where

$$\begin{aligned} \Phi_1 &:= \sum_{t=1}^{N} E[x_t x_t^T] \quad \Phi_2 := \sum_{t=1}^{N} E[x_{t-1}x_{t-1}^T] \\ \Phi_3 &:= \sum_{t=1}^{N} E[x_t x_{t-1}^T] \quad \Phi_4 := \sum_{t=1}^{N} u_{t-1}u_{t-1}^T \\ \Phi_5 &:= \sum_{t=1}^{N} x_t^N u_{t-1}^T \quad \Phi_6 := \sum_{t=1}^{N} x_{t-1}^N u_{t-1}^T \end{aligned} \tag{A-2}$$

All the expectations are evaluated using the previous model estimate *i.e.*, using $\Theta^k = \{A^k, B^k, C^k, Q^k, R^k\}$. Observe that this is the only term in the Q-function that depends on the system matrices A and B. Now it is straightforward to differentiate the above expression to obtain the optimal estimates of the system matrices at the $(k+1)$th iteration.

$$\begin{aligned} A^{k+1} &= [\Phi_3 - \Phi_5\Phi^{-1}\Phi_6^T][\Phi_2 - \Phi_6\Phi_4^{-1}\Phi_6^T]^{-1} \\ B^{k+1} &= [\Phi_5 - A\Phi_6]\Phi_4^{-1} \end{aligned} \tag{A-3}$$

Similarly we can differentiate the first two terms to obtain the optimal new estimate of the covariance matrix, Q^{k+1}

$$\begin{aligned} Q^{k+1} = \frac{1}{N}\Big[&\Phi_1 + A^k\Phi_2 A^{kT} + B^k\Phi_4 B^{kT} + 2\Phi_3 A^{kT} \\ &+ 2\Phi_5 B^{kT} - 2A^k\Phi_6 B^{kT}\Big] \end{aligned} \tag{A-4}$$

The fourth term in the Q-function can similarly be evaluated,

$$\begin{aligned} T_4 &:= -\frac{1}{2}\sum_{t=1}^{N} tr\{R^{-1}E_k(y_t - Cx_t)(y_t - Cx_t)^T\} \\ &= -\frac{1}{2}tr\{R^{-1}\left[\sum_{t=1}^{N} y_t y_t^T + C\Phi_1 C^T - 2C\sum_{t=1}^{N} x_t^N y_t^T\right]\} \\ &:= -\frac{1}{2}tr\{R^{-1}\left[\Phi_7 + C\Phi_1 C^T - 2C\Phi_8\right]\} \end{aligned} \tag{A-5}$$

Since the matrix C appears only in T_4, the new estimate of C, C^{k+1} can be obtained by differentiating T_4 and equating it to zero.

$$C^{k+1} = \Phi_8^T\Phi_1^{-1} \tag{A-6}$$

In order to obtain an expression for the optimal value of R at the current iteration, we must differentiate the third and fourth term in the Q-function. The new estimate of R can be shown to be

$$R^{k+1} = \frac{1}{N}\left[\Phi_7 - \Phi_8^T\Phi_1^{-1}\Phi_8\right] \tag{A-7}$$

Now we must evaluate all Φ_is. In order to do so, we need to use a Kalman filter and a Kalman smoother. Expressions for the Kalman filter and Kalman smoother provided in Shumway and Stoffer (2000) can be modified to suit the current problem.

REFERENCES

A.P. Dempster, N.M. Laird, and D.B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *J. R. Stat. Soc. B*, 39:1–38, 1977.

N.K Gupta and R.K Mehra. Computational aspects of maximum likelihood estimation and reduction in sensitivity function calculations. *IEEE Trans. Automat. Contr.*, 19:774–783, 1974.

K.E Kwok, M Chong-Ping, and G.A Dumont. Seasonal model based control of processes with recycle dynamics. *Ind. Eng. Chem. Res.*, 40:1633–1640, 2001.

D. Li, S.L Shah, and T Chen. Identification of fast rate models from multi-rate data. *International Journal of Control*, pages 680–689, 2001.

L. Ljung. *System Identification: Theory for the user*. Prentice Hall, 1999.

J Morud and S Skogestad. Effects of recycle on dynamics and control of chemical-processing plants. *comput chem eng*, 18:529–534, 1994.

R.H Shumway and D.S. Stoffer. An approach to time series smooting and forecasting using the em algorithm. *J. Time Series Anal.*,, 3:253–264, 1982.

R.H. Shumway and D.S. Stoffer. *Time Series Analysis and Its Applications*. Springer, 2000.