The Relationship between State Space Subspace Identification Methods and the EM Method

Stuart Gibson Brett Ninness

Department of Electrical and Computer Engineering,
University of Newcastle, Australia.
Web: http://www.ee.newcastle.edu.au
Email: {shgibson, brett}@ee.newcastle.edu.au

Abstract

This paper exposes a close connection between subspace-type estimates and single iterates of the Expectation Maximisation (EM) method for Maximum Likelihood (ML) estimation. A key implication of this is that it suggests a means by which ML estimates for MIMO systems may be computed in a numerically robust and straightforward manner.

1 Introduction

Suppose that one is presented with observations $\{y_t\}$ of a stationary stochastic process and is faced with the task of estimating a state-space representation of this process in *innovations* form

$$x_{t+1} = Ax_t + Be_t, \quad (1)$$
$$y_t = Cx_t + De_t \quad (2)$$

where $\{e_t\}$ is an i.i.d. zero mean and unit variance white noise process.

One popular approach to this problem is to apply a subspace-based method [5] to estimate A, B, C and D.

A competing solution is the Maximum Likelihood (ML) method; the statistical superiority of which has been thoroughly examined [3].

As shown here, if the ML estimate is found via the EM scheme [2], then both this and the subspace-based approach employ identical projections on an approximate state-space in order to form their parameter estimates.

Thus, the only difference between a subspace-based estimate and one EM iteration seeking an ML estimate is the approximation used for the state space.

2 An Overview of Subspace Identification Methods

This overview is drawn from that in [1]. Define $Y^-_{t,p} \in \mathbf{R}^{k\times N}$ and $Y^+_{t,f} \in \mathbf{R}^{k\times N}$ as $[Y^-_{t,p}]_{m,n} \triangleq y_{t-m+n-1}$ and $[Y^+_{t,f}]_{m,n} \triangleq y_{t+m+n}$, where $y_s = 0$ when $s \leq 0$ or $s > N$.

Then a so-called subspace method for estimating A, B, C and D is as follows:

1. Derive a matrix $\widehat{\beta} = R_{fp}(R_{pp})^{-1}$ where
$$R_{fp} \triangleq \frac{1}{N} Y^+_{t,f}(Y^-_{t,p})^T \text{ and } R_{pp} \triangleq \frac{1}{N} Y^-_{t,p}(Y^-_{t,p})^T.$$
2. Choose weighting matrices W_f, W_p and perform the SVD factorisation
$$W_f \widehat{\beta} W_p = USV^T = [U_n \mid u] \begin{bmatrix} S_n & 0 \\ 0 & s \end{bmatrix} \begin{bmatrix} V_n^T \\ v^T \end{bmatrix}$$
3. Define further matrices
$$\widehat{\mathcal{O}}_f = W_f^{-1} U_n S_n^{\frac{1}{2}} \qquad \widehat{\mathcal{K}}_p = S_n^{\frac{1}{2}} V_n^T W_p^{-1}.$$
4. Define the state estimate as $\widehat{X}^-_{t,p} = \widehat{\mathcal{K}}_p Y^-_{t,p}$.
5. Estimate C as
$$\widehat{C} = \left(\frac{1}{N}\sum y_t \widehat{x}_t^T\right)\left(\frac{1}{N}\sum \widehat{x}_t \widehat{x}_t^T\right)^{-1}. \quad (3)$$
6. Define a residual estimate as
$$\widehat{De_t} = [y_t^T, y_{t+1}^T, \ldots, y_N^T] - \widehat{C}\widehat{X}^-_{t,p}. \quad (4)$$
7. Estimate A and BD^{-1} as the least-squares solution of
$$\widehat{X}^-_{t+1,p} = [A,\ BD^{-1}] \begin{bmatrix} \widehat{X}^-_{t,p} \\ \widehat{De_t} \end{bmatrix}. \quad (5)$$
8. Estimate D as the lower triangular Cholesky factor of the sample covariance of $\widehat{De_t}$.

3 The EM Algorithm for Likelihood Maximisation

An essential feature of the EM algorithm is the postulate of an unobserved 'complete data set' $Z_N \triangleq (X_N, Y_N)$ that contains what is actually observed $Y_N \triangleq \{y_1, \ldots, y_N\}$, plus other observations X_N which one might wish were available, but in fact are not. Then approximating $\log p_\theta(Z_N \mid Y_N)$ as (subscripts denote conditional dependence)

$$\log p_\theta(Z_N \mid Y_N) \approx Q(\theta, \theta') \triangleq \mathsf{E}_{\theta'}\{\log p_\theta(Z_N \mid Y_N)|Y_N\}$$

leads to the EM algorithm:

1. **E Step** : Calculate $Q(\theta, \widehat{\theta}_n)$.
2. **M Step** : Maximise $\widehat{\theta}_{n+1} = \arg\max_\theta Q(\theta, \widehat{\theta}_n)$.

4 Application of the EM Algorithm for State-Space Estimation

The most obvious choice for the incomplete data set X_N is the state sequence, i.e. $X_N = \{x_0, x_1, \ldots, x_N\}$ (see [4]). Via repeated applications of Bayes' Rule

$$p_\theta(Z_N) = p_\theta(x_0) \prod_{k=0}^{N} p_\theta(x_k \mid x_{k-1}) \prod_{k=0}^{N} p_\theta(y_k \mid x_k).$$

Furthermore, suppose that the data is generated by (2) and $x_{t+1} = Ax_t + Be_t + w_t$, where $e_t \sim \mathcal{N}(0, I_m)$ is independent of $w_t \sim \mathcal{N}(0, \epsilon I)$ and $\epsilon > 0$ is arbitrarily small. This system will be 'close' to (1), (2) yet avoid the mathematical complication of singular distributions.

For this new system the M-step with $\theta = (A, B, C, D)$ becomes the closed form expressions (all expectations in (6)-(9) are to be understood as conditional on the observed data Y_N)

$$\widehat{A} = \left[\sum_{k=1}^{N} \mathbf{E}_{\theta'} \left\{x_k x_{k-1}^T\right\}\right] \left[\sum_{k=1}^{N} \mathbf{E}_{\theta'} \left\{x_{k-1} x_{k-1}^T\right\}\right]^{-1}, \quad (6)$$

$$\widehat{B}\widehat{B}^T = \frac{1}{N}\left[\sum_{k=1}^{N} \mathbf{E}_{\theta'}\{x_k x_k^T\} - \left[\sum_{k=1}^{N} \mathbf{E}_{\theta'} \left\{x_k x_{k-1}^T\right\}\right] \times \right.$$
$$\left. \left[\sum_{k=1}^{N} \mathbf{E}_{\theta'} \left\{x_{k-1} x_{k-1}^T\right\}\right]^{-1} \left[\sum_{k=1}^{N} \mathbf{E}_{\theta'} \left\{x_k x_{k-1}^T\right\}\right]^T\right] - \epsilon I, \quad (7)$$

$$\widehat{C} = \overline{Y}_N \widehat{X}_N^T \left[\sum_{k=0}^{N} \mathbf{E}_{\theta'}\{x_k x_k^T\}\right]^{-1}, \quad (8)$$

$$\widehat{D}\widehat{D}^T = \frac{1}{N}\overline{Y}_N \left[I - \widehat{X}_N^T \left[\sum_{k=0}^{N} \mathbf{E}_{\theta'}\{x_k x_k^T\}\right]^{-1} \widehat{X}_N\right] \overline{Y}_N^T, \quad (9)$$

where $\widehat{x}_k$ and $\mathbf{E}_{\theta'} \left\{x_k x_k^T \mid Y_N\right\}$ may be extracted from a Kalman smoother and $\widehat{X}_N \triangleq [\widehat{x}_0, \ldots, \widehat{x}_N]$, $\overline{Y}_N \triangleq [y_1, \ldots, y_N]$.

5 Relationship between EM and Subspace Methods

Both the EM and the subspace methods estimate the state sequence. In the case of subspace methods the estimate $\widehat{X}_{t,p}^-$ can be thought of as an approximate Kalman filtered estimate [5]. For the EM algorithm a Kalman smoother is employed to estimate the state sequence. Both the subspace and EM method then calculate parameter estimates using a series of projections which involve the state sequence estimate.

In particular, given an estimate of the state sequence the subspace method calculates A, $\overline{B} \triangleq BD^{-1}$ as in (5), that is,

$$\widehat{X}_{t+1,p}^- = \left[A,\ \overline{B}\right] \begin{bmatrix} \widehat{X}_{t,p}^- \\ \\ \overline{Y}_N \widehat{X}_{t,p}^\perp \end{bmatrix}, \quad (10)$$

where $\widehat{X}_{t,p}^\perp = I - (\widehat{X}_{t,p}^-)^T \left[\widehat{X}_{t,p}^- (\widehat{X}_{t,p}^-)^T\right]^{-1} \widehat{X}_{t,p}^-$. Solving this leads to

$$\widehat{A} = \widehat{X}_{t+1,p}^- (\widehat{X}_{t,p}^-)^T \left[\widehat{X}_{t,p}^- (\widehat{X}_{t,p}^-)^T\right]^{-1}, \quad (11)$$

$$\widehat{\overline{B}} = \widehat{X}_{t+1,p}^- \widehat{X}_{t,p}^\perp \overline{Y}_N^T \left[\overline{Y}_N \widehat{X}_{t,p}^\perp \overline{Y}_N^T\right]^{-1}. \quad (12)$$

Comparing (11) and (6) indicates that $\widehat{A}$ is calculated by both the subspace and EM methods using an identical projection. Similarly, equations (3) and (8) immediately indicate that both methods employ the same projection to find $\widehat{C}$.

The subspace method's equation for $\widehat{De_t}$, (4), yields

$$\widehat{D}\widehat{D}^T = \frac{1}{N}\overline{Y}_N \left[I - \left(\widehat{X}_{t,p}^-\right)^T \left[\widehat{X}_{t,p}^- \left(\widehat{X}_{t,p}^-\right)^T\right]^{-1} \widehat{X}_{t,p}^-\right] \overline{Y}_N^T,$$

which is identical to the EM algorithm projection (9). Lastly, the subspace method estimates $\widehat{\overline{B}} = \widehat{B}\widehat{D}^{-1}$ as per (12), implying that

$$\widehat{B}\widehat{B}^T = \frac{1}{N}\widehat{X}_{t+1,p}^- \widehat{X}_{t,p}^\perp \left(\widehat{X}_{t+1,p}^-\right)^T,$$

which, again, is identical to the EM scheme (7) except for the particular method of state estimation.

6 Conclusion

Subspace methods have attracted interest because of their numerical robustness and (because of their lack of imposed state space parameterisation) their suitability for MIMO systems. The fact shown here, that a subspace method is essentially an iteration towards an ML estimate (once Kalman smoother state estimates are used), means that ML estimates can be obtained with the same numerical robustness and with the same parameterisation free advantages. Preliminary results towards this goal are provided in an extended report available via http://www.ee.newcastle.edu.au.

References

[1] M. Deistler, K. Peternell, and W. Scherrer, *Consistency and relative efficiency of subspace methods*, Automatica, 31 (1995).

[2] A. Dempster, N. Laird, and D. Rubin, *Maximum likelihood from incomplete data via the em algorithm*, J. R. Stat. Soc., B, 39 (1977), pp. 1–38.

[3] L. Ljung, *System Identification: Theory for the User*, Prentice-Hall, Inc., New Jersey, 1987.

[4] R. Shumway and D. Stoffer, *An approach to time series smoothing and forecasting using the em algorithm*, J. Time Ser. Anal., 3 (1982).

[5] P. van Overschee and B. de Moor, *Subspace Identification for Linear Systems*, Kluwer Academic Publishers, 1996.