DH

MET O 11 TECHNICAL NOTE No 112

The Kalman filter in relation to autoregressive time series

R. DIXON

1. Introduction

In Met O 11 we are currently exploring the possibility that the forecast errors of our operational numerical model may themselves be forecast to some worthwhile extent by the use of an autoregressive-moving average(ARMA) model.

An error sequence $\{Z_t\}$ can be represented in the form

$$Z_t = \alpha_1 Z_{t-1} + \alpha_2 Z_{t-2} + ----- + \alpha_p Z_{t-p}$$

$$- \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - ------ \theta_q \varepsilon_{t-q} + \varepsilon_t \qquad (1)$$

In this representation the "Z" terms are the numerical forecast errors themselves and the "ε" terms are the errors which we make in attempting to predict the numerical forecast errors. With p terms of the "Z" type and q terms of the "ε" type, in the established terminology we have an ARMA (p, q) process.

Our immediate hope is that a fairly low order ARMA process, say p and $q \leq 3$, will be found to be adequate and that in each analysis and forecast cycle we will be able to find the best ARMA model by a brute force real time search. Should this prove to be impractical we will then have to devise some recursive updating procedure to obtain the best ARMA model at each cycle. The models most suitable for this treatment are the pure AR models $(q=0)$. This note deals with the application of Kalman filtering to the AR process.

Met O 11 (Forecasting Research Branch)
Meteorological Office
London Road
Bracknell
Berks
UK.

2. The State-Space representation

Consider the time-dependent system

$$\underline{a}_t = M_{t-1} \cdot \underline{a}_{t-1} + \underline{\lambda}_{t-1} \qquad (2)$$

$$\underline{h}_t = F_t \cdot \underline{a}_t + \underline{\tau}_t \qquad (3)$$

$\underline{h}_t$ is an $m \times 1$ vector of observations on m dependent variables

F_t is a known $m \times n$ matrix of fixed regressors

$\underline{\tau}_t$ is an $m \times 1$ vector of residuals which is assumed to have a zero mean and a known R_t covariance matrix

$\underline{a}_t$ is an $n \times 1$ vector of regression coefficients

M_{t-1} is a known $n \times n$ transition matrix

$\underline{\lambda}_{t-1}$ is an $n \times 1$ disturbance vector with zero mean and known $n \times n$ covariance matrix Λ_{t-1}

Equation (3) is known as the observation equation, whilst (2) is variously known as the model equation, the plant equation, or the process equation. The term process equation will be preferred here.

The components of the vector $\underline{a}$ are to be regarded as variables, the values of which give a measure of the "state" of some physical/dynamical system. The essence of the problem is that this "state vector" is not observed or measured directly but only indirectly through the regression relationship (3) . Nevertheless it may be necessary to obtain some good idea of the latest value of $\underline{a}$ and to use it to obtain a forecast value of at the next time step. It is thus basically an estimation problem. The true vector $\underline{a}_{t-1}$ is not known. An estimate of it, $\hat{\underline{a}}_{t-1}$, is obtained. A common use to which this estimate is then put is to obtain a forecast value $\overset{*}{\underline{a}}_t$ by using (2) in the approximate form $\overset{*}{\underline{a}}_t = M_{t-1} \cdot \hat{\underline{a}}_{t-1}$

3. The basic Kalman filter

A recursive solution to this problem is provided by the Kalman filter equations -

$$\overset{*}{S}_t = M_{t-1} \cdot \hat{S}_{t-1} \cdot \tilde{M}_{t-1} + \Lambda_{t-1} \qquad (4)$$

$$\hat{S}_t^{-1} = \overset{*}{S}_t^{-1} + \tilde{F}_t \cdot R_t^{-1} \cdot F_t \qquad (5)$$

$$K_t = \hat{S}_t \cdot \tilde{F}_t \cdot R_t^{-1} \qquad (6)$$

$$\overset{*}{\underline{a}}_t = M_{t-1} \cdot \hat{\underline{a}}_{t-1} \qquad (7)$$

$$\hat{\underline{a}}_t = \overset{*}{\underline{a}}_t + K_t \cdot (\underline{h}_t - F_t \cdot \overset{*}{\underline{a}}_t) \qquad (8)$$

where S_{t-1} is the covariance matrix

$$S_{t-1} = E(\hat{\underline{a}}_{t-1} - \underline{a}_{t-1})(\hat{\underline{a}}_{t-1} - \underline{a}_{t-1})^{\sim} \qquad (9)$$

The action of the filter is that having got a forecast value $\overset{*}{\underline{a}}_t$ from (7) the next estimate $\hat{\underline{a}}_t$ for the next cycle is acquired from (8) by an optimal correction of $\overset{*}{\underline{a}}_t$ obtained by weighting the observation error $(\underline{h}_t - F_t \cdot \overset{*}{\underline{a}}_t)$ with a "gain matrix" K_t. The gain matrix itself is kept updated from cycle to cycle by equations (4), (5) and (6). This sequence of operations provides minimum mean square linear unbiased estimates $\hat{\underline{a}}_{t-1}$ of $\underline{a}_{t-1}$ at each cycle. Naturally, the process has to start somewhere so it is assumed that reasonable first guesses or estimates of the covariance matrices mentioned above and also of $\underline{a}_0$ are available.

The Kalman filter was first developed in the literature on control theory in relation to engineering, particularly electrical engineering, and is now widely used. A faily complete account of the derivation of the filter equations, the properties of the filter, and its applications is given in NATO Agardograph Nos 139 and 131, copies of which are in the Met O Library.

More general sets of equations are sometimes used in the literature, but the above set has sufficient generality for the purpose of this note. By the use of matrix identities and Lemmas the filter equations can be thrown into various forms. A strictly equivalent set is

$$\overset{*}{S}_t = M_{t-1} \cdot \hat{S}_{t-1} \cdot \tilde{M}_{t-1} + \Lambda_{t-1} \tag{10}$$

$$\hat{S}_t = \overset{*}{S}_t - K_t \cdot F_t \cdot \overset{*}{S}_t \tag{11}$$

$$K_t = \overset{*}{S}_t \cdot \tilde{F}_t \cdot \left[F_t \cdot \overset{*}{S}_t \cdot \tilde{F}_t + R_t\right]^{-1} \tag{12}$$

$$\overset{*}{\underline{a}}_t = M_{t-1} \cdot \hat{\underline{a}}_{t-1} \tag{13}$$

$$\hat{\underline{a}}_t = \overset{*}{\underline{a}}_t + K_t \cdot (\underline{h}_t - F_t \cdot \overset{*}{\underline{a}}_t) \tag{14}$$

(10) is the same as (4). (13) and (14) are the same as (7) and (8). (11) and (12) are equivalent to (5) and (6), though this is not obvious.

4. <u>The autoregressive series case</u>

The problem is, given a sequence $\{z_t\}$ and the representation,

$$z_t = \alpha_1 z_{t-1} + \alpha_2 z_{t-2} + \cdots + \alpha_p z_{t-p} + \varepsilon_t \tag{15}$$

to obtain the best least-squares estimate $\hat{\underline{\alpha}}_t$ of $\underline{\alpha}_t = (\alpha_1, \alpha_2, \cdots, \alpha_p)$ at each time step.

The state space/Kalman filter approach is to regard this as being the specialization of (2) and (3) to the case where the matrix F_t in (3) is the $p \times 1$ vector $\underline{f}_t$, where

$$\underline{f}_t \triangleq (z_{t-1}, z_{t-2}, z_{t-3}, \cdots, z_{t-p})^{\sim}$$

M_{t-1} is the $p \times p$ unit matrix I

$\underline{\lambda}_{t-1}$ is the zero vector

$\underline{h}_t$ becomes the scalar z_t

$\underline{I}_t$ becomes the scalar error ε_t

The system is thus

$$\underline{\alpha}_t = I \cdot \underline{\alpha}_{t-1} \qquad (16)$$

$$z_t = \tilde{\underline{f}}_t \cdot \underline{\alpha}_t + \varepsilon_t \qquad (17)$$

and the Kalman filter equations become

$$\overset{*}{S}_t = \hat{S}_{t-1} \qquad (18)$$

$$\hat{S}_t = \overset{*}{S}_t - \underline{k}_t \cdot \tilde{\underline{f}}_t \cdot \overset{*}{S}_t \qquad (19)$$

$$\underline{k}_t = \overset{*}{S}_t \cdot \underline{f}_t \cdot \left[\tilde{\underline{f}}_t \cdot \overset{*}{S}_t \cdot \underline{f}_t + \sigma_\varepsilon^2\right]^{-1} \qquad (20)$$

$$\overset{*}{\underline{\alpha}}_t = I \cdot \hat{\underline{\alpha}}_{t-1} \qquad (21)$$

$$\hat{\underline{\alpha}}_t = \overset{*}{\underline{\alpha}}_t + (z_t - \tilde{\underline{f}}_t \cdot \overset{*}{\underline{\alpha}}_t)\underline{k}_t \qquad (22)$$

where the gain matrix K_t of (12) is now a gain vector $\underline{k}_t$ in (20). When reduced to this simpler form it becomes more obvious that the filter incorporates a predictor-corrector element for (22) shows that the forecasts $\overset{*}{\underline{\alpha}}_t$ are optimally corrected by a vector weighting of the error $(z_t - \tilde{\underline{f}}_t \cdot \underline{\alpha}_t)$ The form of the process equation (16), with $M_{t-1} = I$ and no process error term, means that an exact but unknown fixed "true" process is being envisaged. The "true" coefficient vector does not change, with each iteration but our estimate of it does. In this connection it should be noted that (21) is simply $\overset{*}{\underline{\alpha}}_t = \hat{\underline{\alpha}}_{t-1}$ and so the forecast of z_t is

$$\overset{*}{z}_t = \tilde{\underline{f}}_t \cdot \overset{*}{\underline{\alpha}}_t = \tilde{\underline{f}}_t \cdot \hat{\underline{\alpha}}_{t-1} \qquad (23)$$

Of course a value for σ_ε^2 must be provided and also initial estimates of S_0 and $\underline{\alpha}_0$. It is possible to start with $S_0 = \gamma I$ where γ is some large number, and $\underline{\alpha}_0 = 0$, but usually something more sophisticated can be done.

R Dixon

R Dixon
September 1978

This is ROLS