5.7 Adaptive Problems

A major point of concern in estimation procedures based upon Gauss-Markov type methods lies in specification of the various covariance matrices, especially those describing the model error (here included in $\mathbf{Q}(t)$). The reader will probably have concluded that there is, however, nothing precluding deduction of the covariance matrices from the model and observations, given that adequate numbers of observations are available. For example, it is straightforward to show that if a Kalman filter is operating properly, then the so-called innovation, $\mathbf{y}(t) - \mathbf{E}\tilde{\mathbf{x}}(t,-)$, should be uncorrelated with all previous measurements:

{66013}
$$\langle \mathbf{y}(t') \left[\mathbf{y}(t) - \mathbf{E}\tilde{\mathbf{x}}(t,-) \right] \rangle = 0\,, \quad t' < t \tag{5.29}$$

(recall Ch. 2, Eq. (2.431)). To the extent that (5.29) is not satisfied, the covariances need to be modified, and algorithms can be formulated for driving the system toward this condition. The possibilities for such procedures are known under the title "adaptive estimation."[177]

The major issues here are that accurate determination of a covariance matrix of a field, $\langle \mathbf{z}(t)\, \mathbf{z}(t') \rangle$, requires a vast volume of data. Note in particular that if the mean of the field $\mathbf{z}(t) \neq 0$, and it is inaccurately removed from the estimates, then major errors can creep into the estimated second moments. This bias problem is a very serious one in adaptive methods.

In practical use of adaptive methods, it is common to reduce the problem dimensionality by *modelling the error covariance matrices,* that is by assuming a particular, simplified structure described by only a number of parameters much less than the number of matrix elements (accounting for the matrix symmetry). We must leave this subject to the references.

Appendix to Chapter. Doubling

We wish to make the doubling algorithm plausible.[178] Consider the matrix equation,

$$\mathbf{B}_{k+1}= \mathbf{F}\mathbf{B}_k\mathbf{F} + \mathbf{C}, \tag{5.30}$$ {doubling1}

and we seek to time-step it. Starting with $\mathbf{B}_1$, one has, time stepping as far as $k = 3$,

$$\begin{aligned}
\mathbf{B}_2 &= \mathbf{F}\mathbf{B}_1\mathbf{F}^T+\mathbf{C},\\
\mathbf{B}_3 &= \mathbf{F}\mathbf{B}_2\mathbf{F}^T + \mathbf{C} = \mathbf{F}^2\mathbf{B}_1\mathbf{F}^{2T}+\mathbf{F}\mathbf{Q}\mathbf{F}^T+\mathbf{C},\\
\mathbf{B}_4 &= \mathbf{F}\mathbf{B}_3\mathbf{F}^T+\mathbf{C}\\
&= \mathbf{F}^2\mathbf{B}_2\mathbf{F}^{2T} + \mathbf{F}\mathbf{C}\mathbf{F}^T + \mathbf{C}\\
&= \mathbf{F}^2\mathbf{B}_2\mathbf{F}^{2T}+\mathbf{B}_2,
\end{aligned}$$

that is, $\mathbf{B}_4$ is given in terms of $\mathbf{B}_2$. More generally, putting $\mathbf{M}_{k+1} = \mathbf{M}_k^2$, $\mathbf{N}_{k+1} = \mathbf{M}_k\mathbf{N}_k\mathbf{M}_k^T+\mathbf{N}_k$, with $\mathbf{M}_1= \mathbf{F}, \mathbf{N}_1= \mathbf{Q}$, then $\mathbf{M}_{2k} = \mathbf{F}^{2^k}$, $\mathbf{N}_{k+1} = \mathbf{B}_{2^k}$ and one is solving Eq. (5.30) so that the time step doubles at each iteration. An extension of this idea underlies the doubling algorithm used for the Riccati equation.

Notes

[152]See for example, Kitagawa and Sato (2001) for references.

[153]See, e.g., Arulampalam et al. (2002). Their development relies on a straightforward Bayesian approach.

[154]See Evensen (1996) and the references there for a more complete discussion.

[155]See Press et al. (1992) for detailed help concerning generating values from known probability distributions.

[156]Kalnay (2003).

[157]See Gardiner (1985) for a complete discussion

[158]Evensen (1994, 1996) are good starting points for practical applications, insofar as problem dimension have permitted. See Kalnay (2003) for a broad discussion of the specific numerical weather forecasting problem.

[159]See the reviews by Lorenc (1986), Daley, (1991); or Ghil & Malanotte-Rizzoli, 1991).

[160]Usually called "3D-VAR", by meteorologists, although like "4D-VAR" it is neither variational nor restricted to three dimensions.

[161]Anthes (1974)

[162]Gelb (1974, Chs. 7,8) has a general discussion of the computation reduction problem, primarily in the continuous time context, but the principles are identical.

[163]Kalman's (1960) filter derivation was specifically directed at extending the Wiener theory to the transient situation, and it evidently reduces back to the Wiener theory when a steady-state is appropriate.)

[164]Anderson and Moore (1979) should be consulted for a complete discussion.

[165]Fukumori et al. (1993) discuss this problem in greater generality for a fluid flow.

[166]Fukumori (1995), who interchanges the roles of $\mathbf{D}, \mathbf{D}^{+}$.

[167]A general discussion of various options for carrying out the transformations between fine and coarse states is provided by Fieguth et al. (2003).

[168]Used for example, by Cane et al. (1996).

[169]E.g., Brogan (1985).

[170]Thacker (1989) and Marotzke and Wunsch (1993).

[171]Tziperman et al. (1992b) grapple with ill-conditioning in their results; the ill-conditioning is interpretable as arising from a nullspace in the Hessian.

[172]This potential confusion is the essence of the conclusions drawn by Farrell (1989), and Farrell and Moore (1993) and leads to the discussion by Trefethen (1997, 1999) of pseudo-spectra.

[173]Bracewell (1978)

[174]Trefethen (1997)

[175]Hasselmann (1988); von Storch et al., (1988, 1993).

[176]The meteorological literature, e.g., Farrell and Moore (1993), renamed this singular vector as the "optimal" vector.

[177]Among textbooks that discuss this subject are those of Haykin (1986), Goodwin and Sin (1984), and Ljung (1987).

[178]Following Anderson and Moore (1979, p. 67).