Brief Paper

Unbiased Minimum-variance Linear State Estimation*

PETER K. KITANIDIS†

Abstract—A method is developed for linear estimation in the presence of unknown or highly non-Gaussian system inputs. The state update is determined so that it is unaffected by the unknown inputs. The filter may not be globally optimum in the mean square error sense. However, it performs well when the unknown inputs take extreme or unexpected values. In many geophysical and environmental applications, it is performance during these periods which counts the most. The application of the filter is illustrated in the real-time estimation of mean areal precipitation.

Introduction

STATE ESTIMATION in the presence of unknown model inputs or parameters, $\theta(t)$, is common in geophysical applications. Perhaps the most common approach is to treat $\theta(t)$ as a (vector) stochastic process with given wide-sense representation. For example, it may be assumed that $\theta(t)$ is a wide-sense stationary or stationary-increment process with known mean and covariance. However, if these assumptions are not valid, they may have a major adverse effect on filter performance. Adaptive filtering and parameter identification methods (e.g. Mehra, 1970) may be applied for the calculation of the first two moments of $\theta(t)$, but this approach must also rely on some assumptions about how $\theta(t)$ varies in time.

In some cases, there may be insufficient evidence that $\theta(t)$ is a stationary or stationary-increment process. A case in point is environmental monitoring in the face of major spills of unknown magnitude. The modeler may find out that filtering is too dependent on poorly justified assumptions about how $\theta(t)$ varies in time and about the numerical values of its mean and covariance matrix. Another line of research is decision-theoretic approaches which have been developed for the detection and estimation of isolated impulse inputs or jumps (e.g. Sanyal and Shen, 1974; Wilsky and Jones, 1976).

The objective of this paper is to show how state estimation given the most recent measurements may be achieved independent of the values of unknown parameters. The developed methodology may also be used in conjunction with detection rules, such as those in Sanyal and Shen (1974) and Wilsky and Jones (1976), which serve to determine whether a "jump" has occurred. However, the applicability of the proposed methodology is not limited to the case of *isolated* impulses or jumps.

The linear-additive unknown input case

Attention will be limited to linear discrete-time systems with the unknown input being added to the state. The system

† Department of Civil Engineering, Terman Engineering Center, Stanford University, Stanford, CA 94305, U.S.A.

equation is

$$x(t+1) = \Phi(t)x(t) + G(t)\theta(t) + w(t) \qquad (1)$$

where t is the time index; $x(t)$ is the $n \times 1$ state vector; $\theta(t)$ is the $p \times 1$ vector of unknown inputs or parameters; $w(t)$ is an $n \times 1$ zero-mean (without loss of generality) white-noise (uncorrelated in time) process with covariance matrix $Q(t)$; and $\Phi(t)$ and $G(t)$ are known $n \times n$ and $n \times p$ matrices, respectively. It will be convenient to assume that the rank of $G(t)$ is p. This assumption is not limiting since one can always parameterize the problem in terms of the minimum number of parameters which effect the state. The measurement equation is

$$z(t) = H(t)x(t) + v(t) \qquad (2)$$

where $z(t)$ is the $m \times 1$ vector of observations; $v(t)$ is an $m \times 1$ vector of observation error, represented as a white-noise series of zero mean and covariance matrix $R(t)$; and $H(t)$ is a known $m \times n$ matrix. We are concerned with applications for which the number of measurements is much larger than the number of unknowns. This is common in geophysical applications involving spatially distributed systems (e.g. see Matheron, 1971; Kitanidis, 1986). In this work, it will be sufficient to assume that $m \geq p$.

If $\theta(t)$ is treated as a white-noise random process with mean $m_\theta(t)$ and covariance matrix $\Sigma_\theta(t)$, then Kalman filtering yields the minimum variance unbiased estimate of the state vector $x(t+1)$ given measurements up to time $t+1$. One problem is that the "optimality" of this estimator can be compromised by a poor choice of $m_\theta(t)$ and $\Sigma_\theta(t)$. Even if acceptable estimates of these quantities can be obtained, an even more serious objection is that they are inadequate to represent the variability of highly non-Gaussian processes. A case in point is when $\theta(t) t = 1, 2, \ldots$, is a sequence of large but seldom occurring "jumps", such as system failures, environmental spills, or major geophysical events. For example, in the measurement of rainfall, long periods of drought may be followed by short periods of torrential rain. The filter is slow in moving the state estimates to the new level created by the major input (Kitanidis and Bras, 1980). One may be willing to trade some of the theoretical optimality, in the mean square error sense, of the estimator for improved resistance to ignorance about $\theta(t)$.

Robust linear filter

In the following, a linear estimator of the state $x(t+1)$ given measurements up to time $t+1$ will be developed so that it is not affected by the values of the unknown deterministic vectors $\theta(t)$. Let

$$E[\hat{x}(t/t) - x(t)] = 0 \qquad (3)$$

and

$$P(t/t) = E[(\hat{x}(t/t) - x(t))(\hat{x}(t/t) - x(t))^{T}]. \qquad (4)$$

Consider the linear filter

$$\hat{x}(t+1/t+1) = \Phi(t)\hat{x}(t/t)$$
$$+ L(t+1)[z(t+1) - H(t+1)\Phi(t)\hat{x}(t/t)] \qquad (5)$$

where $L(t+1)$ is an $n \times m$ matrix, generally different from the Kalman gain, to be selected according to the following criteria.

(a) *Unbiasedness.* The estimator must satisfy

$$E[\hat{x}(t+1/t+1) - x(t+1)] = 0 \qquad (6)$$

or

$$E[\Phi(t)\hat{x}(t/t) + L(t+1)[H(t+1)\Phi(t)x(t)$$
$$+ H(t+1)G(t)\theta(t) + H(t+1)w(t) + v(t+1) - H(t+1)$$
$$\Phi(t)\hat{x}(t/t)] - \Phi(t)x(t) - G(t)\theta(t) - w(t)] = 0. \qquad (7)$$

For this condition to hold for any value of $\theta(t)$, the gain matrix must satisfy the constraint

$$L(t+1)H(t+1)G(t) - G(t) = 0. \qquad (8)$$

In this work, we will concern ourselves with cases for which $H(t+1)G(t)$ has rank p, i.e. its columns are linearly independent. Since $G(t)$ has rank p, this condition is required for the existence of a solution which satisfies constraint (8).

Before proceeding, it may be useful to point out that (5) is the most general form of a linear unbiased estimator. Indeed, if the ostensibly more general form $\hat{x}(t+1 \mid t+1) = A_1(t+1)\hat{x}(t \mid t) + A_2(t+1)z(t+1)$ were assumed, one could easily show that, for unbiasedness, it is required that $A_1(t+1) - \Phi(t) + A_2(t+1)H(t+1)\Phi(t) = 0$. Thus, setting $A_2(t+1) = L(t+1)$ and $A_1(t+1) = \Phi(t) - L(t+1)H(t+1)\Phi(t)$, the form of equation (5) is obtained.

(b) *Minimum-variance.* The estimation error matrix corresponding to estimator (5) is

$$P(t+1/t+1) = [I - L(t+1)H(t+1)][\Phi(t)P(t/t)\Phi^T(t)$$
$$+ Q(t)][I - L(t+1)H(t+1)]^T$$
$$+ L(t+1)R(t+1)L^T(t+1). \qquad (9)$$

Defining

$$P(t+1/t) = \Phi(t)P(t/t)\Phi^T(t) + Q(t) \qquad (10)$$

and

$$C(t+1) = H(t+1)P(t+1/t)H^T(t+1) + R(t+1) \qquad (11)$$

the expression for the estimation error matrix may be simplified to read

$$P(t+1/t+1) = L(t+1)C(t+1)L^T(t+1)$$
$$- P(t+1/t)H^T(t+1)L^T(t+1)$$
$$- L(t+1)H(t+1)P(t+1/t) + P(t+1/t). \qquad (12)$$

For minimum-variance estimation, the problem is equivalent to finding $L(t+1)$ which minimizes the trace of the estimation error matrix of equation (12) subject to the constraint (8). The Langrangian is

$$Tr[L(t+1)C(t+1)L^T(t+1)$$
$$- 2P(t+1/t)H^T(t+1)L^T(t+1) + P(t+1/t)]$$
$$- 2 Tr[(L(t+1)H(t+1)G(t) - G(t))A(t+1)^T] \qquad (13)$$

where $A(t+1)$ is an $n \times p$ matrix of Lagrange multipliers. The coefficient 2 is included for notational convenience. Taking the derivatives with respect to $L(t+1)$ equal to zero yields:

$$2C(t+1)L^T(t+1) - 2H(t+1)P(t+1/t)$$
$$- 2H(t+1)G(t)A(t+1)^T = 0. \qquad (14)$$

Equations (8) and (14) are the first-order necessary conditions for the optimum. Together, they form a system of $(n \times m) + (n \times p)$ linear equations from which $L(t+1)$ and $A(t+1)$ may be determined. Combining them in matrix notation:

$$\begin{bmatrix} C(t+1) & -H(t+1)G(t) \\ G^T(t)H^T(t+1) & 0 \end{bmatrix}$$
$$\times \begin{bmatrix} L^T(t+1) \\ A^T \end{bmatrix} = \begin{bmatrix} H(t+1)P(t+1 \mid t) \\ G^T(t) \end{bmatrix}. \qquad (15)$$

Note that exactly the same conditions would have been obtained if the variance of each element of the state vector had been maximized separately subject to the unbiasedness constraints.

Equation (15) has a unique solution if and only if the $(m+p) \times (m+p)$ matrix of coefficients is nonsingular. Let $L^*(t+1)$ and $A^*(t+1)$ satisfy the first-order necessary conditions, equations (8) and (14). Then, if the solution $L^*(t+1)$ is substituted by another feasible solution $L^*(t+1) + \Delta L$, where ΔL is nonzero and satisfies

$$\Delta L H(t+1)G(t) = 0, \qquad (16)$$

then the value of the objective function is increased by $Tr[\Delta L G(t+1) \Delta L^T]$. Thus, the second-order necessary and sufficient condition for L^* to correspond to a unique minimum of the constrained optimization problem (see Luenberger, 1984, Section 10.5) is that for any nonzero ΔL which satisfies (16),

$$Tr[\Delta L C(t+1) \Delta L^T] > 0. \qquad (17)$$

Summarizing the necessary and sufficient conditions for a unique solution to the optimization of the quadratic objective function subject to linear constraints: the columns of $H(t+1)G(t)$ must be linearly independent and $C(t+1)$ must be such that any nonzero $n \times m$ matrix ΔL which satisfies (16) also satisfies (17). In further developments, these conditions are assumed to hold.

Condition (17) is less strict a requirement than nonsingular $C(t+1)$. For purposes of obtaining a useful expression for the gain matrix and making some points, we will consider the case that $C(t+1)$ is invertible. [If $C(t+1)$ has rank $k, m > k \geq m - p$, one may substitute the measurements by k of their linear combinations which have an invertible $C(t+1)$ matrix.] From (8) and (14)

$$A(t+1)$$
$$= [G(t) - P(t+1/t)H^T(t+1)C^{-1}(t+1)H(t+1)G(t)]$$
$$\times [G^T(t)H^T(t+1)C^{-1}(t+1)H(t+1)G(t)]^{-1} \qquad (18)$$

and

$$L(t+1) = P(t+1/t)H^T(t+1)C^{-1}(t+1) + [G(t)$$
$$- P(t+1/t)H^T(t+1)C^{-1}(t+1)H(t+1)G(t)]$$
$$\times [G^T(t)H^T(t+1)C^{-1}(t+1)H(t+1)G(t)]^{-1}$$
$$\times G^T(t)H^T(t+1)C^{-1}(t+1). \qquad (19)$$

It is noted that this solution is different from the one given in Sanyal and Shen (1974) [equation (15)], where a similar problem is examined. The difference is that in this and other references, the best estimate of the input was first estimated and then used for compensation purposes whereas we bypass the estimation of the unknown input. We concentrate on minimizing the error matrix of the filtered state and the only effect of the unknown input is constraint (8). Also, the objective of this paper is different from that of Glover (1969), which concentrated on the estimation of the unknown input.

The estimation error matrix corresponding to this estimator is

$$P(t+1/t+1) = P(t+1/t) - P(t+1/t)H^T(t+1)$$
$$\times C^{-1}(t+1)H(t+1)P(t+1/t)$$
$$+ A(t+1)G^T(t)H^T(t+1)C^{-1}(t+1)$$
$$\times H(t+1)G(t)A(t+1)^T \qquad (20)$$

or

$$P(t+1/t+1) = P(t+1/t) - P(t+1/t)H^T(t+1)$$
$$\times C^{-1}(t+1)H(t+1)P(t+1/t)$$
$$+ [G(t) - P(t+1/t)H^T(t+1)C^{-1}(t+1)$$
$$\times H(t+1)G(t)][G^T(t)H^T(t+1)$$
$$\times C^{-1}(t+1)H(t+1)G(t)]^{-1}[G(t)$$
$$- P(t+1/t)H^T(t+1)C^{-1}(t+1)$$
$$\times H(t+1)G(t)]^T. \qquad (21)$$

From (19) through (21) it is evident that if $\Lambda(t+1) = 0$ then the developed filter is identical to the discrete-time Kalman filter. However, the constraint (8) is generally binding and $\Lambda(t+1)$ is nonzero. The gain [equation (19)] may be seen as the sum of the Kalman gain [for $\Sigma_\theta(t) = 0$] plus an additional term signifying the additional reliance on measurements to cancel out the effects of the unknown input. The state estimation error matrix is the sum of the Kalman estimation error matrix plus an additional term, quadratic with respect to the matrix of the Lagrange multipliers $\Lambda(t+1)$. The additional term is non-negative definite reflecting the loss of accuracy caused by the unknown input. The increase in the state estimation error matrix also results in larger gain in subsequent steps. Finally, note that for no temporal correlation structure [$\Phi(t) = 0$], the estimator becomes essentially equivalent to the linear minimum-variance unbiased estimator known in geophysics as kriging (e.g. Matheron, 1971).

The same filter may be obtained as follows: consider $\theta(t)$ as a white-noise random process with mean $m_\theta(t)$ and covariance matrix $\Sigma_\theta(t)$; develop the Kalman filter equations; and calculate the limit for $\Sigma_\theta^{-1}(t) \to 0$. The proposed filter equations are thus obtained. It is obvious that the derived filter is not globally optimal in the mean square error sense, except in the special case that there is absolutely no information about $\theta(t)$. However, the main advantage of the proposed filter is that its performance is not affected by the value of $\theta(t)$, an important consideration when $\theta(t)$ is highly non-Gaussian or has unknown properties.

Real-time estimation of mean areal precipitation during a storm

Calculation of mean areal precipitation associated with a given event or period of time is an important practical problem which has been addressed in several publications. The most common conventional (nonstatistical) methods are Thiessen polygons and the method of isoyetals. Methods based on linear estimation have been attracting much attention recently (e.g. Bras and Colon, 1978; Delhomme, 1978; Gandin, 1965; Rodriguez-Iturbe and Mejia, 1974; to mention only a few of the relevant publications). The basic advantage of linear estimation, compared to conventional nonstatistical methods, is that they are based on an explicit model of the correlation structure of rainfall which may be corroborated with data. Furthermore, linear estimation provides the means to evaluate the mean-squared error of areal-average estimates, a very useful feature in real-time flow forecasting and operation of reservoirs, as well as in the optimization of rainfall networks.

Most applications associated with the estimation of mean areal precipitation have examined rainfall as a realization of a stochastic process in space only. However, the way precipitation evolves in time is important in real-time estimation. Generalizing a commonly used model (Bras and Colon, 1978; Rodriguez-Iturbe and Mejia, 1974)

$$f(\mathbf{s}, t) = \sum_{i=1}^{p} F_i(\mathbf{s}, t)\theta_i + \varepsilon(\mathbf{s}, t) \qquad (22)$$

where $f(\mathbf{s}, t)$ is precipitation at a location with spatial coordinates $\mathbf{s} = (s_1, s_2)^T$ at time t; $F_i(\mathbf{s}, t), i = 1, \ldots, n$ are known deterministic functions of spatial coordinates and time; $\theta_i, i = 1, \ldots, p$ use unknown coefficients; and $\varepsilon(\mathbf{s}, t)$ is a zero-mean stochastic process, assumed stationary in both space and time. It is also commonly assumed that the covariance functions can be adequately represented by the product of the covariance in time and space, and that the process is Markovian in time so that its sampled values can be represented in a state-space model such as that of equation (1).

Illustrative example

Consider the problem of estimating mean areal precipitation over a square region A at time t given measurements of precipitation at four locations with spatial coordinates $s_1 = (0.2, 0.2)^T$, $s_2^T = (0.5, 0.4)^T$, $s_3 = (0.3, 0.7)^T$ and $s_4 = (0.8, 0.7)^T$. The four vertices of the square are at $(0, 0)$, $(0, 1)$, $(1, 0)$ and $(1, 1)$. Furthermore, assume that precipita-

tion is a realization of the process

$$f(\mathbf{s}, t) = \theta(t) + \varepsilon(\mathbf{s}, t) \qquad (23)$$

where $\theta(t)$ is an unknown parameter, different for each time period, and $\varepsilon(\mathbf{s}, t)$ is a stochastic process with zero mean and covariance, the one used in Rodriguez-Iturbe and Mejia (1974):

$$E[\varepsilon(\mathbf{s}_i, t)\varepsilon(\mathbf{s}_j, \tau)] = \rho^{|t-\tau|} \cdot K(|\mathbf{s}_i - \mathbf{s}_j|) \qquad (24)$$

where $K(|\mathbf{s}_i - \mathbf{s}_j|)$ is a covariance function. Let $m(t)$ be the mean areal precipitation during period t or, more formally,

$$m(t) = \frac{1}{|A|} \int_A f(\mathbf{s}, t) \, d\mathbf{s} \qquad (25)$$

where $|A|$ is the area of region A and $\int f(\mathbf{s}, t) \, d\mathbf{s}$ denotes the double integral $\int\int f(s_1, s_2, t) \, ds_1 \, ds_2$. Then the expected value of $m(t)$ is $\theta(t)$, and

$$\text{cov}(m(t), m(\tau)) = \frac{\rho |t - \tau|}{|A|^2} \int_A \int_A K(|\mathbf{u} - \mathbf{v}|) \, d\mathbf{u} \, d\mathbf{v} \qquad (26)$$

$$\text{cov}(m(t), f(\mathbf{s}_i, \tau)) = \frac{\rho |t - \tau|}{|A|} \int_A K(|\mathbf{u} - \mathbf{s}_i|) \, d\mathbf{u}. \qquad (27)$$

The quadruple integral of (26) and the double integral of (27) can be calculated numerically for arbitrary geometry or covariance function. For the simple geometry of this problem and the linear covariance function

$$K(|\mathbf{s}_i - \mathbf{s}_j|) = -\sigma^2 |\mathbf{s}_i - \mathbf{s}_j| \qquad (28)$$

these equations have analytical expressions (see Hoeksema and Kitanidis, 1984, p. 1019). It is pointed out that (28) is a generalized covariance function (see Matheron, 1973). It is conditionally positive definite and can be used to calculate the moments of data increments which are invariant to $\theta(t)$. Note that a linear minimum-variance estimator of $m(t)$ involves only covariances of $m(t)$ with measurements up to period t. In general terms, the augmented state vector of the system may be defined by the values of f on a grid and the $m(t) \cdot \Phi$, G, Q and H matrices of the state-space representation may be easily determined. Measurement error is accounted for in R which is taken diagonal with each variance equal to r. Note that in this case, $m = 4$ and $p = 1$.

Some results corresponding to a severe precipitation event are shown in Fig. 1. For $\rho = 0.8$, $\sigma^2 = 100$, $r = 50$, and initial condition $x(0 \mid 0) = 0$ and $P(0 \mid 0) = 30 \, Q$, Fig. 1 depicts the actual mean areal precipitation and the one estimated by the filter. Mean area precipitation is obviously a highly
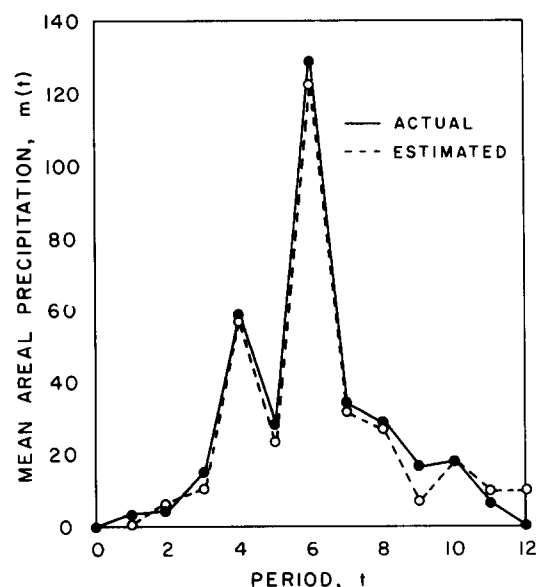

Fig. 1. Actual (simulated) and estimated (using proposed filter) mean areal precipitation.

non-Gaussian process. Nevertheless, the filter performed equally well in periods of high and low precipitation.

Conclusions

This paper has illustrated how linear minimum-variance unbiased state estimation can be used to determine mean areal precipitation in the presence of unknown inputs. Precipitation is represented as a stochastic process in space and time with a mean which is unknown and variable in time. The estimates of space-averaged precipitation are determined so that they have minimum-variance and are unbiased for any value of the unknown mean of the process. Under certain conditions the result is identical to kriging, a linear estimation technique used in geophysical sciences. However, the methodology presented in this paper can account for temporal correlation which may be pronounced when the time interval is short.

References

Bras, R. L. and R. Colon (1978). Time-averaged areal mean of precipitation: estimation and network design. *Wat. Resources Res.*, **14**, 878–888.

Bras, R. L. and I. Rodriguez-Iturbe (1976). Rainfall network design for runoff prediction. *Wat. Resources Res.*, **12**, 1197–1208.

Delhomme, J. P. (1978). Kriging in the hydrosciences. *Adv. Wat. Resources*, **1**, 251–266.

Gandin, L. S. (1965). *Objective Analysis of Meteorological Fields.* Israel Program for Scientific Translation. Jerusalem.

Glover, J. D. (1969). The linear estimation of completely unknown signals. *IEEE Trans. Aut. Control*, **AC-14**, 766–769.

Hoeksema, R. J. and P. K. Kitanidis (1984). An application of the geostatistical approach to the inverse problem in 2-D groundwater modeling. *Wat. Resources Res.*, **20**, 1003–1020.

Kitanidis, P. K. (1986). Parameter uncertainty in estimation of spatial functions: a Bayesian approach. *Wat. Resources Res.*, **22**, 499–507.

Kitanidis, P. K. and R. L. Bras (1980). Adaptive filtering through detection of isolated transient errors in rainfall-runoff models. *Wat. Resources Res.*, **16**, 740–748.

Luenberger, D. G. (1984). *Linear and Nonlinear Programming.* Addison-Wesley, Reading, Massachusetts.

Matheron, G. (1971). *The Theory of Regionalized Variables and Its Applications.* Ecole de Mines, Fountainbleau, France.

Matheron, G. (1973). The intrinsic random functions and their applications. *Adv. Appl. Probability*, **5**, 439–468.

Mehra, R. K. (1970). On the identification of variances and adaptive Kalman filtering. *IEEE Trans. Aut. Control*, **AC-15**, 175–184.

Rodriguez-Iturbe, I. and J. M. Mejia (1974). On the design of rainfall networks in time and space. *Wat. Resources Res.*, **10**, 713–728.

Sanyal, P. and C. N. Shen (1974). Bayes' decision rule for rapid detection and adaptive estimation scheme with space applications. *IEEE Trans. Aut. Control*, **AC-19**, 228–231.

Willsky, A. S. and H. L. Jones (1976). A generalized likelihood ratio approach to the detection and estimation of jumps in linear systems. *IEEE Trans. Aut. Control*, **AC-21**, 108–112.