Proceedings of the 29th Conference
on Decision and Control
Honolulu, Hawaii • December 1990

Simultaneous ML Estimation of State and Parameters for Hyperbolic Systems with Noisy Boundary Conditions

Arunabha Bagchi and Paul ten Brummelhuis

Department of Applied Mathematics
University of Twente
P.O. Box 217, 7500 AE Enschede, the Netherlands

Abstract

In this paper, we present a method to estimate simultaneously states and parameters of a discrete-time hyperbolic system with noisy boundary conditions. This method is based on the maximization of a likelihood function. Although this technique is described here for linear systems, possible extensions to non-linear systems are also briefly discussed.

1 Introduction

The original idea of using the ML method for estimating the state of a lumped dynamical system goes back to the sixties (Cox [1], Detchmendy and Sridhar [2] and Mortensen [3]). It has been shown in [1] that the recursive version of state estimation for linear systems is precisely the Kalman Filter which verifies the important result that the mean and the mode coincide for Gaussian distributions. But the restriction to linear systems is a severe one. For example, if the unknown parameters are included in the augmented state space the resulting estimation problem already becomes non-linear. In [3] approximate ML equations are given for non-linear, continuous-time systems which, in general, will differ from the extended Kalman Filter. Any error analysis of such approximations, both for the ML estimation and for the extended Kalman Filter, is not available yet. A more direct method of ML estimation of states and parameters has been suggested by Bar-Shalom [4]. The numerical scheme suggested there is, however, not very realistic for fast implementation in systems with reasonable large dimension of the state vector whereas the situation becomes even more involved if the system is non-linear. For systems without state noise, Chavent [5] in a seminal contribution proposed a fast method for calculating the gradient of the least-squares error (the same as the likelihood function in this situation) with respect to the unknown parameters. Here, we extend the method of Chavent to the more general case of incomplete state estimation.

The maximization of the likelihood function leads to a two-point boundary value problem of considerable complexity. If we restrict to discrete-time problems, the large dimension of the state vector and the direct solution of the two-point boundary value problem may lead to a huge computational load. We therefore propose in this paper an alternative computational method which is much faster and makes use of specific features of the hyperbolic system.

2 The mathematical model

Although the proposed method is quite general, the particular hyperbolic system we are interested in, concerns the linear shallow water equations governing the flow in a two-dimensional basin :

$$\frac{\partial \xi}{\partial t} + \frac{\partial (Du)}{\partial x} + \frac{\partial (Dv)}{\partial y} = 0 \qquad (1)$$

$$\frac{\partial u}{\partial t} + g\frac{\partial \xi}{\partial x} - fv + \frac{\lambda}{D}u - \frac{\gamma}{D}V^2 cos(\phi) = 0 \qquad (2)$$

$$\frac{\partial v}{\partial t} + g\frac{\partial \xi}{\partial y} + fu + \frac{\lambda}{D}v - \frac{\gamma}{D}V^2 sin(\phi) = 0 \qquad (3)$$

where t stands for time, (x,y) for the spatial coordinates, and

$$\begin{aligned}
\xi(x,y,t) &= \text{water height above some reference level}\\
D(x,y) &= \text{water depth below this reference level}\\
u(x,y,t) &= \text{velocity in the x-direction}\\
v(x,y,t) &= \text{velocity in the y-direction}\\
g &= \text{acceleration due to gravity}\\
f &= \text{Coriolis constant}\\
\lambda(x,y) &= \text{bottom friction parameter}\\
\gamma &= \text{wind stress parameter}\\
V &= \text{wind velocity}\\
\phi &= \text{wind direction}
\end{aligned}$$

Since these equations have a hyperbolic character, the equilibrium solution $\xi(x,y,0) = u(x,y,0) = v(x,y,0) = 0$ can always be chosen as the initial condition, because the solution is, after some time, completely determined by the boundary conditions (and the meteorological input).

The eqs. (1)-(3) can be used to model the water height and water velocity in the North Sea in the region between the English and the Dutch coasts [6, 7]. The purpose is now to estimate the bottom friction - and the wind stress parameters and to use these estimates to predict the water heights in the coastal areas. As an example of a two-dimensional model, we may reasonably consider a rectangular basin with three boundaries assumed fixed (expressed by setting the velocity component perpendicular to the boundary equal to zero), while it is impossible to specify precisely the boundary condition on the fourth boundary ($y = 0$ and $0 \leq x \leq x_{max}$), which is connected to the open sea. One possibility is to use, see [8],

$$v(x,0,t) = \tilde{b}(x,t) + B(x,t) \qquad (4)$$

where $\tilde{b}$ (x,t) is an approximation of the true boundary condition (calculated on the basis of atmospheric equations) and B(x,t) is the unknown part of the boundary condition. The following noisy model has been used in [6] to specify this unknown part :

$$\begin{aligned}
B(x,t) &= (1-\beta_x)V_1(t) + \beta_x V_2(t), \quad 0 \leq \beta_x \leq 1 &(5)\\
dV_i(t) &= -\alpha V_i dt + q dW_i(t), \quad i = 1,2, \ \alpha \leq 1 &(6)
\end{aligned}$$

with V_i (t), $i = 1,2$ representing two stochastic processes indicating boundary conditions at the edges ($y = 0$ and $x = 0$ or $x = x_{max}$) and $W_i(t)$, $i = 1,2$ are independent Brownian motion processes. Using an appropriate discretization scheme [7] the dynamical system is converted into a discrete-time form, where the state

$$\underline{\tilde{X}}^k = [..., u^k_{m,n}, v^k_{m,n}, \xi^k_{m,n}, ...]^T$$

evolves in time according to :

$$\begin{aligned}
\underline{\tilde{X}}^{k+1} &= \tilde{F}(\lambda,\gamma)\underline{\tilde{X}}^k + \underline{\tilde{b}}^{k+1} + \tilde{G}\underline{V}^{k+1} &(7)\\
V_i^{k+1} &= (1-\alpha\Delta t)V_i^k + q(W_i^{k+1} - W_i^k)\Delta t, \ i = 1,2 &(8)
\end{aligned}$$

Because it is more convenient in the sequel to deal with only one system equation, we introduce the augmented state

$$\underline{X}^k = [\tilde{\underline{X}}^k, V_1^k, V_2^k] \quad (9)$$

and the white noise processes

$$w_i^k = W_i^k - W_i^{k-1}, \;\; i = 1,2 \quad (10)$$

which satisfy equation (9) :

$$\underline{X}^{k+1} = F(\lambda,\gamma)\underline{X}^k + \underline{b}^{k+1} + G\underline{w}^k \quad (11)$$

3 The estimation procedure

We want to estimate the states and the parameters, based on the measurements $\{\underline{Z}^k, k = N_1, ..., N_2\}$, where $\underline{Z}^k$ is related to $\underline{X}^k$ by

$$\underline{Z}^k = H\underline{X}^k + \underline{n}^k \quad (12)$$

assuming that $\{\underline{n}^k\}$ is a Gaussian white noise sequence. The basic idea of simultaneous ML estimation of the state and the parameters is the minimization, with respect to λ, γ and $\underline{X}^k, k = N_0, ..., N_2$, of the negative of the log-likelihood function :

$$\begin{aligned} J_0(\lambda,\gamma,\underline{\chi}) &= \sum_{k=N_1}^{N_2} \| \underline{Z}^k - H\underline{X}^k \|^2_{R^{-1}} + \\ & \sum_{k=N_0}^{N_2} \| \underline{X}^{k+1} - F(\lambda,\gamma)\underline{X}^k - \underline{b}^{k+1} \|^2_{(GQG^T)^{-1}} \quad (13) \end{aligned}$$

where the time-indices satisfy $N_0 \leq N_1 \leq N_2$, $\underline{\chi}$ denotes the state trajectory $\underline{X}^k, N_0 \leq k \leq N_2\}$ and R and Q are the covariances of the observation and state noises respectively. For a certain positive matrix M, the square of the weighted norm $\| x \|^2_M$ stands for $< x, Mx >$, with $< .,. >$ denoting the inner-product in the appropriate Euclidean space. For given λ and γ, let $\underline{\chi}^{k*}(\lambda,\gamma) = \{\underline{X}^{k*}(\lambda,\gamma); k = N_0, .., N_2\}$ be the trajectory that minimizes $J_0(\lambda,\gamma;\underline{\chi})$ with respect to the state. Then the global minimum of $J_0(.)$ can be found by minimizing $J_0(\lambda,\gamma,\underline{\chi}^*(\lambda,\gamma))$ with respect to $(\lambda,\gamma) \in P$, the set of the admissible parameters.

The minimization problem may be reconverted into a discrete-time optimal control problem to determine $\underline{\chi}^*(\lambda,\gamma)$ as follows :

$$\min \sum_{k=N_1}^{N_2} \| \underline{Z}^k - H\underline{X}^k \|^2_{R^{-1}} + \sum_{k=N_0}^{N_2} \| G\underline{w}^k \|^2_{(GQG^T)^{-1}} \quad (14)$$

subject to the dynamical constraint

$$\underline{X}^{k+1} = F(\lambda,\gamma)\underline{X}^k + \underline{b}^{k+1} + G\underline{w}^k \quad (15)$$

The discrete-time version of Pontryagin's minimum principle gives us the following set of necessary conditions to determine $\underline{\chi}^*(\lambda,\gamma)$:

$$\begin{aligned} \underline{X}^{k+1} &= F(\lambda,\gamma)\underline{X}^k + \underline{b}^{k+1} + \tfrac{1}{2}GQG^T\underline{\mu}^{k+1}, k = N_0, ..., N_2 \quad (16) \\ \underline{\mu}^{k-1} &= F(\lambda,\gamma)^T\underline{\mu}^k, k = N_0, ..., N_1 \quad (17) \\ \underline{\mu}^{k-1} &= F(\lambda,\gamma)^T\underline{\mu}^k - 2H^TR^{-1}[\underline{Z}^{k-1} - H\underline{X}^{k-1}], \\ & \quad k = N_1 + 1, ..., N_2 + 1 \quad (18) \\ \underline{\mu}^{N_2+1} &= 0 \quad (19) \end{aligned}$$

Here, $\{\underline{\mu}^k\}$ are the adjoint variables. The next step is to minimize $J_0(.)$ with respect to λ and γ under the constraint that $\underline{\chi} = \{\underline{X}^k, N_0 \leq k \leq N_2\}$ satisfies the equations (16)-(19). We use the method of Lagrange multipliers to solve this problem. Thus we convert the constrained optimization problem to the following unconstrained one:

min $J_1(\lambda,\gamma,\underline{\chi})$ with respect to λ and γ with

$$\begin{aligned} & J_1(\lambda,\gamma,\underline{\chi}) = J_0(\lambda,\gamma,\underline{\chi}) + \\ & \sum_{k=N_0}^{N_2} (\underline{\nu}^{k+1})^T[\underline{X}^{k+1} - F(\lambda,\gamma)\underline{X}^k - \underline{b}^{k+1} - \tfrac{1}{2}GQG^TQ\underline{\mu}^{k+1}] \\ & + \sum_{k=N_0}^{N_1} (\underline{\sigma}^k)^T[\underline{\mu}^{k-1} - F(\lambda,\gamma)^T\underline{\mu}^k] + \quad (20) \\ & \sum_{k=N_1+1}^{N_2+1} (\underline{\sigma}^k)^T\left[\underline{\mu}^{k-1} - F(\lambda,\gamma)^T\underline{\mu}^k + 2H^TR^{-1}[\underline{Z}^{k-1} - H\underline{X}^{k-1}]\right] \end{aligned}$$

In this equation $\{\underline{\nu}^k, k = N_0 + 1, ..., N_2 + 1\}$ and $\{\underline{\sigma}^k, k = N_1, ..., N_2\}$ are the Lagrange multipliers. This is a purely deterministic parameter estimation problem and one can use a gradient algorithm to obtain the minimum of $J_1(.)$. In order to determine ∇J_1 the method of Chavent can be used [5, 9]. The advantages of this method are that the exact gradient is determined, whereas the computational load is independent of the number of parameters that have to be estimated. To illustrate the idea, we consider only λ to be unknown, and furthermore assume that λ is constant. To consider the incremental change $\Delta\lambda$ and $\Delta J_1, \Delta\underline{X}^k, \Delta\underline{\mu}^k$, we linearize (20) and rearrange the terms to get :

$$\begin{aligned} & \Delta J_1 + \sum_{k=N_0}^{N_1} \Delta\lambda[(\underline{\nu}^{k+1})^T\frac{\partial F}{\partial\lambda}\underline{X}^k + (\underline{\sigma}^k)^T\frac{\partial F^T}{\partial\lambda}\underline{X}^k] = \\ & \sum_{k=N_0+1}^{N_1-1} (\Delta\underline{X}^k)^T[\underline{\nu}^k - F^T\underline{\nu}^{k+1}] + \\ & \sum_{N_1}^{N_2} (\Delta\underline{X}^k)^T\left[\underline{\nu}^k - F^T\underline{\nu}^{k+1} - 2HR^{-1}[\underline{Z}^k - H\underline{X}^k - \underline{\sigma}^{k+1}]\right] + \\ & (\Delta\underline{X}^{N_2+1})^T\underline{\nu}^{N_2+1} + (\Delta\underline{\mu}^{N_0+1})^T\underline{\sigma}^{N_0} + \\ & (\Delta\underline{\mu}^{N_0})^T[\underline{\sigma}^{N_0+1} - F\underline{\sigma}^{N_0}] + \\ & \sum_{k=N_0}^{N_2-1} (\Delta\underline{\mu}^{k+1})^T\left[\underline{\sigma}^{k+2} - F\underline{\sigma}^{k+1} + \tfrac{1}{2}GQG^T[\underline{\mu}^{k+1} - \underline{\nu}^{k+1}]\right] + \\ & (\Delta\underline{\mu}^{N_2+1})^T\left[-F\underline{\sigma}^{N_2+1} + \tfrac{1}{2}[\underline{\mu}^{N_2+1} - \underline{\nu}^{N_2+1}]\right] \quad (21) \end{aligned}$$

Now, this equation must hold for all infinitesimal changes $\Delta\lambda, \Delta J_1$ and $\Delta\underline{X}^k$, $\Delta\underline{\mu}^{k+1}$ for $k = N_0, ..., N_2$. This leads to the equations satisfied by the Lagrange multipliers $\underline{\nu}^{k+1}$ and $\underline{\sigma}^k$ for $k = N_0, ..., N_2$, and a simple expression for the partial derivative of J_1 with respect to λ. The solution of the equations is

$$\underline{\sigma}^k = 0 \text{ and } \underline{\nu}^k = \underline{\mu}^k \text{ for all k} \quad (22)$$

while the component of the gradient is given by

$$\frac{\partial J_1}{\partial\lambda} = - \sum_{k=N_0}^{N_2} (\underline{\mu}^{k+1})^T\frac{\partial F}{\partial\lambda}\underline{X}^{k+1}. \quad (23)$$

Obviously, the expression for $\frac{\partial J_1}{\partial\gamma}$ is analogous. The set of equations (16)-(19), combined with the boundary conditions $\underline{x}^{N_0} = 0$ and $\underline{\mu}^{N_2+1}$, determines the ML estimates of the states $\underline{X}^k, k = N_0, ..., N_2$. These are the smoothed estimates, based on the entire set of available data $\{\underline{Z}^k, k = N_1, .., N_2\}$.

4 Results and Discussion

In the last section we derived the result that the gradient of some error functional J_I with respect to the unknown parameters immediately follows from the solution of the two-point-boundary-value-problem, that determines the smoothed states of a system. So, the major issue to be dealt with now is how we can efficiently calculate the solution of eqs. (16)-(19). We will distinguish three cases :

- If the numerical and physical damping are absent in eq. (11), implying that $\lambda = 0$ and $\alpha = 0$, this equation can be integrated backward in time once an initial condition $\underline{X}_{N_2}$ is established. This can be done by processing the data$\{\underline{Z}^k, k = N_1, ..., N_2\}$ by a Kalman Filter, because at the final time the smoothed state equals the filtered state. Now with both initial conditions given at the time with index N_2, the solution of the eqs. (16)-(19) can be determined. The restrictions that are imposed here can be met in situations where the bottom friction effect is neglected and, for example, the wind stress must be estimated.

- However, in most cases of parameter estimation problems the backward integration is excluded because of the numerical problems. But if we take a closer look at eq. (16) it is clear that the adjoint variables $\underline{\mu}^k$ only directly influence the components V_1^k and V_2^k of $\underline{X}^k$. This means that the problem of determining the smoothed states is essentially a problem of determining the smoothed estimates of V_1^k and V_2^k for $k = N_0, ..., N_2$, which, again, can be done by the Kalman Filter for linear systems. Therefore, we need the augment the state to

$$\underline{Y}^k = [\underline{X}^k, V_1^{k-1}, V_2^{k-1}, ..., V_1^{k+(N_0-N_2)}, V_2^{k+(N_0-N_2)}] \quad (24)$$

  and process all the data sequentially by the Kalman Filter in order to determine the smoothed boundary. The smoothed states then follow from eq. (7) with $\underline{Y}^k$ representing the smoothed boundary. This idea is being further developed at present. It exploits the hyperbolic nature of the problem in an essential way.

- If one has to deal with non-linear systems, the approach mentioned in the second item may also be applicable when the data assimilation is based on an approximated linear system. The legitimacy of this approximation depends on the influence of the non-linearity in the dynamics and the availability of data that are registered in the neigborhood of the particular boundary.

To illustrate the concepts mentioned here, we will show some representative numerical examples at the conference.

References

[1] H. Cox, *On the estimation of state variables and parameters for noisy dynamical systems*, IEEE Transactions on Automatic Control, AC-9, pp. 5-12, 1964

[2] D.M. Detchmendy and R. Sridhar, *Sequential estimation of states and parameters in noisy nonlinear dynamical systems*, Proceedings of the Joint Automatic Control Conference, Troy, New York, 1965

[3] R.E. Mortensen, *Maximum likelihood recursive nonlinear filtering*, Journal of Optimization Theory and Applications, vol. 2, pp. 386-394, 1968

[4] Y. Bar-Shalom, *Optimal simultaneous state estimation and parameter identification in linear discrete-time systems*, IEEE Transactions on Automatic Control, AC-17, pp. 308-319, 1972

[5] G. Chavent, *Identification of distributed parameter systems : about the least square method, its implementation and identifiability*, Identification and System Parameter Estimation, Proceedings of the Fifth IFAC Symposium held at Darmstadt, Pergamon Press, Oxford, 1979

[6] A.W. Heemink, *Two-dimensional shallow water flow identification*, Applied Mathematical Modelling, vol. 12, 1988

[7] P.G.J. ten Brummelhuis, A.W. Heemink and B. de Jong, *Stochastic dynamic approach to predict water levels in estuaries*, Journal of Hydraulic Engineering, vol, 114, New York, 1988

[8] P.G.J. ten Brummelhuis and A.W. Heemink, *Parameter identification in tidal models with uncertain boundary conditions*, Stochastic Hydrology and Hydraulics (to appear)

[9] G.A. van Zee and O.H. Bosgra, *Gradient computation in predictor error identification of linear discrete-time systems*, IEEE Transactions on Automatic Control, vol. AC-27, pp. 738-739, 1982