Automatica 83 (2017) 317–330

Contents lists available at ScienceDirect

Automatica

journal homepage: www.elsevier.com/locate/automatica

# On the efficient low cost procedure for estimation of high-dimensional prediction error covariance matrices☆

Hong Son Hoang [1], Remy Baraille

*SHOM/HOM/REC, 42 av Gaspard Coriolis 31057 Toulouse, France*

ARTICLE INFO

*Article history:*
Received 5 December 2014
Received in revised form 14 March 2017
Accepted 23 May 2017

*Keywords:*
Covariance matrices
Adaptive filters
Filter stability
Prediction error sampling
Separation of vertical and horizontal structures
Nearest Kronecker problem
Parameter estimation
Stochastic approximation

ABSTRACT

A simple, efficient algorithm is proposed for estimating the prediction error covariance matrix which plays the key role for successful state estimation in very high dimensional systems. The main results are obtained by introducing the hypothesis on the separability of vertical and horizontal structure of the error covariance matrix and its parameterization. A new parameter optimization problem is formulated which is closely related to the Nearest Kronecker Problem (NKP). This allows to estimate optimally the unknown parameters of the structured parametrized ECM as well as to approach numerically the solution of the traditional NKP in a simple and efficient way. The algorithm for the state estimation will be detailed. The results from experiments on parameter and state estimation problems, for both moderate and high dimensional numerical models, demonstrate a high effectiveness of the proposed filtering approach.


## 1. Introduction

Partial Differential Equations are used practically in all scientific areas, from financial markets to mathematical biology — not to say on the well-known domains like quantum mechanics, electrodynamics, oceanography and meteorology etc. As a mathematical model is only a simplification or abstraction of a (complex) real world, the measurements (observations) constitute the most important source of information which should be used promptly in order to improve the model solution for practical problems. This task can be excellently accomplished by filtering algorithms.

The difficulties encountered in the design of an optimal filter are numerous. These concern system nonlinearities, uncertainties in specification of system parameters and noise statistics etc. But the most insurmountable difficulty in application of optimal filtering algorithms lies on very high dimension of the system state. By the very dimensional system we mean a system whose state dimension is of order $10^6$–$10^7$. For example, a typical dimension of the 2d image vector is $10^4$–$10^5$. As for the resulting system state of oceanic numerical models, its dimension is of order $10^6$–$10^7$.

One of the most advantages of the Kalman filter (KF) (Kalman, 1960) is that it allows, under mild conditions, to produce the best estimate for the system state along with providing error covariance matrices (ECM) of the filtered and prediction errors (PE) during the estimation process. However, for very high dimensional systems, the KF will fail to be applied simply due to the fact that it is impossible to handle and store the ECM with $10^{12}$–$10^{14}$ elements in the most modern and powerful computers at the present (and at least in the near future).

First mention that, covariance matrices of high dimension play a fundamental role in the analysis of multivariate data collected from a variety of fields like business and economics, health care, engineering etc. Formally, we are interested in the question of how to approximate the actual covariance matrix on the basis of a sample of the multivariate distribution in high-dimensional settings (Bai & Shi, 2011). The two major difficulties in estimation of the ECM are high dimensionality and positive-definiteness (Pourahmadi, 2013). High dimensional covariance estimation focuses on the methodologies based on shrinkage, banding, tapering, thresholding, penalized likelihood etc. (Pourahmadi, 2013; Touloumis, 2015)

This paper addresses the problem of how one can well estimate, and at a low cost, the ECM of high dimension which constitutes an essential element in the design of a filter for data assimilation problems in geophysical systems (DA-Geos). Various approaches

☆ The material in this paper was presented at the 19th World Congress of the International Federation of Automatic Control (IFAC 2014), August 24–29, 2014, Cape Town, South Africa. This paper was recommended for publication in revised form by Associate Editor Brett Ninness under the direction of Editor Torsten Söderström.

*E-mail addresses:* hhoang@shom.fr (H.S. Hoang), remy.baraille@shom.fr (R. Baraille).

[1] Fax: +33 5 61 40 52 58.

http://dx.doi.org/10.1016/j.automatica.2017.06.018
0005-1098/

have been pursued to overcome the difficulties in estimating high dimensional ECMs. This includes methods without specification of ECM like nudging procedure in which a relaxation term is added to the equations of the model to force the observations to the model (Auroux & Blum, 2008) or dynamically based method (Cooper & Haines, 1996) which projects altimeter surface height data in the vertical by lifting or lowering water columns. In the multivariate optimal interpolation (Srinivasan et al., 1999) covariances are expressed as a product of the correlation matrix and a diagonal matrix with variances at the diagonal. The correlations are further separated into horizontal and vertical components. All scalar auto correlations between values at locations separated by scaled horizontal distances, and scaled vertical distances, are modeled as products of (Second Order) Auto Regressive functions. Background variances are computed from a time series of differences of samples of system state and their average generated in a free running (Srinivasan et al., 1999; Stein, 1999).

The more advanced methods are developed in the form of variational methods (Talagrand & Courtier, 1987) with the ECM specified a priori on the basis of physical consideration or simulation. For example, in meteorology one widely used and very crude PE-ECM (Prediction Error ECM) is estimated using the difference between a 48 hours (h) and 24 h forecasts valid at the same time as a proxy for 6 h forecasts (Parrish & Derber, 1992). Another method is based on producing an ensemble of forecasts by integration of different filtered estimates perturbed by noisy observations (Houtekamer & Mitchell, 1998). The balance operator method assumes that the most obvious correlation in the PE is the balance between mass errors and wind errors (Derber & Bouttier, 1999). As to oceanic assimilation, in Oke, Brassington, and Griffin et al. (2008) deviations of the state about 3 month running average are considered as a proxy for forecast error. In this context, we mention a fast minimum norm filtering algorithm in Feng, Ma, Fu, and Yang (2015) which is proposed to overcome the difficulty in the specification of the process and measurement noises and it can be considered as a successive application of the variational method (Talagrand & Courtier, 1987). For more details on physical consideration, see Gaspari, Cohn, Guo, and Pawson (2006).

The idea on an adaptive filter (AF) in Hoang, De Mey, Talagrand, and Baraille (1997) is closely related to Dullerud and Paganini (2000) and Ge, Yang, Dai, Jiao, and Lee (2009) on a robust adaptive control. It is supposed hence to design an AF in such a way that the overall filtering system should be stable in the presence of bounded parameter uncertainties of the filter's gain with corresponding ECMs. An optimal filter is next found by adjusting the parameters of stabilizing gain to minimize the PE of the system outputs. Here the parameters vary in the intervals of admissible values guaranteeing a filter stability. This will prevent the filter to be divergent.

Recently, one class of filters known as ensemble based filters is widely used for DA-GeoS. This class includes different filters like ensemble KF (Evensen, 2007), ensemble square-root KF (Furrer & Bengtsson, 2007), singular evolutive extended KF (Pham, Verron, & Roubaud, 1998) etc. where the PE-ECM is replaced by a sample low-rank approximation which evolves in time during data assimilation according to the KF formalism. All such filters can be considered as belonging to the class of Sequential Monte Carlo methods, also known as particle filters (Doucet & Godsill, 2000). One of the disadvantages of the EnBFs is that for high dimensional systems, as the sample size is too small compared to the large dimension of the system state, the empirical estimators of covariance and correlation are very unstable. The estimation procedure, developed in this paper, is aimed at overcoming the difficulties related to considerable variability in the state estimate, primarily through prior and posterior sample covariance matrices due to rank deficiency. The structure of covariance proposed in multivariate optimal interpolation (Srinivasan et al., 1999) with priori specified parameters can be considered as a particular case of the present method. Comparing with the classical shrinkage technique where a priori specification of a target diagonal matrix is used, here the target matrix is estimated by fitting the sample covariance with a matrix of given Separation of Vertical and Horizontal Structures (SeVHS). For a review of data assimilation methods, see Reichle (2008).

The paper is organized as follows. In Section 2 a stabilizing structure of the filter gain is described based on Hoang, Baraille, and Talagrand (2009). Section 3 outlines the sampling procedure for simulating PE samples developed in Hoang, Baraille, and Talagrand (2001). The hypothesis on separation of vertical and horizontal structure for the ECM is introduced in Section 4. Mention that such separability of the vertical and horizontal structure is a widely used approach to the modeling of the ECM for meteorological and oceanic assimilation (Daley, 1991; Furrer & Bengtsson, 2007; Srinivasan et al., 1999). The parameter estimation problem is formulated here along with the algorithms for estimating the unknown parameters. It will be shown that under certain conditions the SeVHS of the ECM will imply the SeVHS for the filter gain (Theorem 5.1 and Corollary 5.1). The proof of convergence of the estimation procedure based on an ensemble of samples for ECM will be given. Simplicity and efficiency of the procedure for estimating the ECM are verified in Section 6. Section 7 is devoted to the experiment on assimilation of simulated data into the high dimensional ocean Miami Isopycnic Coordinate Ocean Model.

## 2. Structure of prediction error covariance matrix

### 2.1. Filter structure

Consider a standard filtering problem for linear time-invariant system

$$x(t+1) = \Phi x(t) + w(t),\ t = 0, 1, 2, \dots,$$
$$z(t+1) = Hx(t+1) + \epsilon(t+1),\ t = 0, 1, 2, \dots \quad (1)$$

here $x(t)$ is the $n$-dimensional system state at the $t$ assimilation instant, $\Phi$ is the $(nxn)$ fundamental matrix, $z(t)$ is the $p$-dimensional observation vector, $H$ is the $(pxn)$ observation matrix, $w$, $\epsilon$ are the model and observation noises. We assume $w(t)$, $\epsilon(t)$ are uncorrelated sequences of zero mean and time-invariant covariance $Q$ and $R$ respectively. Mention that in the DA-Geos, the system (1) is a state-space representation of the numerical model derived from a set of equations discretized at some spatial grid.

The idea on an AF is to introduce first a non-adaptive filter (Hoang et al., 1997)

$$\hat{x}(t+1) = \hat{x}(t+1/t) + K\zeta(t+1), \quad (2)$$

$\hat{x}(t+1/t) = \Phi\hat{x}(t)$, $\zeta(t+1) = z(t+1) - H\hat{x}(t+1/t)$ is the innovation vector, $\hat{x}(t+1)$ is the filtered (or analysis) estimate, $\hat{x}(t+1/t)$ is the prediction for $x(t+1)$. In (2) the gain $K := K(\theta)$ is assumed to be given a priori and parametrized by some vector of unknown parameters $\theta$. The AF is obtained by tuning $\theta$ to minimize the PE of the system output.

### 2.2. Structure of filter gain

Different parametrized stabilizing structures of $K$ are obtained in Hoang et al. (2001). The AF is obtained by

$$J[\theta] = E[\|\zeta(t)\|^2] \to \min \theta, \quad (3)$$

where $E(.)$ denotes the mathematical expectation.

For example, one of the possible stabilizing gain structures is given by

$$K = MH^T[HMH^T + R]^{-1},\ M = P_r M_e P_r^T, \quad (4)$$