Available online at www.sciencedirect.com
ScienceDirect
Mathematics and Computers in Simulation 119 (2016) 128–141
MATHEMATICS AND COMPUTERS IN SIMULATION
www.elsevier.com/locate/matcom

Original articles

A unified square-root approach for the score and Fisher information matrix computation in linear dynamic systems

M.V. Kulikova[a,*], J.V. Tsyganova[b]

[a] CEMAT, Instituto Superior Técnico, Universidade de Lisboa, Av. Rovisco Pais 1, 1049-001 Lisbon, Portugal
[b] Ulyanovsk State University, Str. L. Tolstoy 42, 432017 Ulyanovsk, Russian Federation

Received 23 June 2011; received in revised form 18 June 2015; accepted 29 July 2015
Available online 18 August 2015

Abstract

One of the most frequently encountered problems in practice is to combine a priori knowledge about a physical system with experimental data to provide on-line estimation of an unknown dynamic state and system parameters. The classical way for solving this problem is to use adaptive filtering techniques. The adaptive schemes for the maximum likelihood estimation based on gradient-based optimization methods are, in general, preferable. They require the likelihood function and its gradient evaluation (score), and might demand the Fisher information matrix (FIM) computation. All techniques for the score and the FIM calculation in linear dynamic systems yield the implementation of the Kalman filter (KF) and its derivatives (with respect to unknown system parameters), which is known to be numerically unstable. An alternative solution can be found among algorithms developed in the KF community for solving ill conditioned problems: the square-root algorithms, the UD-based factorization methods and the fast SR Chandrasekhar–Kailath–Morf–Sidhu techniques. Recently, these advanced KF implementations have been extended on the filter derivatives computation. However there is no systematic way of designing the robust "differentiated" methods. In this paper, we develop a unified square-root methodology of generating the computational techniques for the filter/smoother derivatives evaluation required in gradient-based adaptive schemes for the score and the FIM computation.

Keywords: Kalman filtering; Maximum likelihood estimation; Gradient-based methods; Fisher information matrix; Square-root algorithms

1. Introduction

Application of the Kalman filter (KF) for estimating an unknown state vector of linear dynamic system assumes a complete *a priori* knowledge of the process and measurement noise statistics. In most practical situations, these statistics are unknown. The classical way of determining the uncertain information is to use adaptive filtering (AF). A number of approaches exist for performing the AF, i.e. for making the filtering method as a learning algorithm; see the detailed discussion in [20]. One of the most frequently used AF techniques is to combine a priori knowledge about a physical system with experimental data to provide on-line estimation of an unknown dynamic state and system

* Corresponding author.
E-mail addresses: maria.kulikova@ist.utl.pt (M.V. Kulikova), tsyganovajv@gmail.com (J.V. Tsyganova).

http://dx.doi.org/10.1016/j.matcom.2015.07.007

parameters simultaneously. Such AF schemes consist of a recursive optimization method to identify uncertain system parameters by minimizing a properly defined performance index (PI) and the filtering method, see [2,6]. The filter inside the AF is used to estimate the states and to calculate the PI. If the maximum likelihood principal is implemented, then the negative log likelihood function (LF) represents the PI. It is then used in the recursive optimization method to improve the estimates of the process parameters. Among optimization methods, the gradient-based algorithms have a high convergence rate and, in general, are preferable. They require a determination of the log LF gradient (known as a "score") and might demand the Fisher information matrix (FIM) calculation.

The classical approach for the score and the FIM computation is based on the explicit differentiation of the KF equations (with respect to unknown system parameters). It leads to a set of p vector equations, known as the *filter sensitivity equations*, and a set of p matrix equations, known as the *Riccati-type sensitivity equations*, where p is the number of unknown system parameters to be estimated. This *forward filtering method* requires roughly the implementation of the p equivalent KF's all running recursively in the forward time direction. The alternative approach is the *backward filtering approach*. The method was first designed by Sandell and Yared in [23] and then improved in [32]. The backward filtering technique is based on the theory of adjoint equations. They run backward in time and compute the adjoint variables, which together with the forward (KF) variables perform the evaluation of the score. In contrast to the direct (forward) approach, this method requires roughly two KFs. The backward filtering technique was used for the reduced gradient evaluation in the prediction error identification as well; see [10,29]. The most recently designed and the most advanced approach for the score evaluation requires only smoothing algorithm [14,25,26]. This method enables the score computation exactly in a single pass of the KF and a smoother.

This paper addresses the numerical aspects of the log LG and the FIM computation. All techniques mentioned above require the implementation of the KF/smoother and its derivatives (with respect to unknown system parameters). It is usually carried out by the conventional implementation (see, for instance, the methodology in [5,7]), which is known to be numerically unstable with respect to roundoff errors [30,31]. This yields to a failure of the filter and, hence, to incorrect PI evaluation. Consequently, it destroys the entire AF computational scheme for the system parameters and states estimation. The solution can be found among algorithms developed in the KF community for solving ill conditioned problems. These are the array square-root (SR) algorithms [3,13,22], the UD-based factorization methods [1,4] and the fast SR Chandrasekhar–Kailath–Morf–Sidhu KF techniques [11,21,24]. Such methods perform the robust filtering having the property of better conditioning and reduced dynamical range. Apart from numerical advantages, array SR algorithms appear to be better suited to parallel and to very large scale integration (VLSI) implementations [12, Chapter 12]. Recently, these advanced KF algorithms have been extended on the derivatives computation required in gradient-based AF schemes for the score and the FIM evaluation. More precisely, such "differentiated" techniques have been designed for the SR schemes in [2,16,27], for the UD-based factorization methods in [28] and for the fast KF implementations in [17,19]. However there is no systematic way of designing the "differentiated" algorithms. The goal of this paper is to develop a unified methodology of generating the robust SR methods for the KF/smoother derivatives evaluation required in the gradient-based AF for the score and the FIM computation. Similar result has been obtained recently in [18], however, it does not cover the fast SR Chandrasekhar–Kailath–Morf–Sidhu filters. The fast algorithms employ J-orthogonal rotations instead of usual orthogonal transformations at each iteration step of the filter/smoother. The unified methodology developed in this paper is applicable for any SR method with a J-orthogonal or usual orthogonal transformation.

2. Maximum likelihood estimation of linear discrete-time stochastic systems

Consider the discrete-time linear stochastic system

$$x_{k+1} = F_\theta x_k + G_\theta w_k, \quad k = 1, \dots, N, \tag{1}$$

$$z_k = H_\theta x_k + v_k \tag{2}$$

where $x_k \in \mathbb{R}^n$ and $z_k \in \mathbb{R}^m$ are, respectively, the state and the measurement vectors; k is a discrete time, i.e. x_k means $x(t_k)$. The process noise, $\{w_k\}$, and the measurement noise, $\{v_k\}$, are Gaussian white-noise processes, with covariance matrices $Q_\theta \geq 0$ and $R_\theta > 0$, respectively. All random variables have known mean values, which we can take without