Automatica 77 (2017) 336–343

Contents lists available at ScienceDirect

Automatica

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

Brief paper

Noise covariance identification for nonlinear systems using expectation maximization and moving horizon estimation ☆

Ming Ge [a], Eric C. Kerrigan [b]

[a] Department of Electrical & Electronic Engineering, Imperial College London, London SW7 2AZ, UK

[b] Department of Electrical & Electronic Engineering and Department of Aeronautics, Imperial College London, London SW7 2AZ, UK

ARTICLE INFO

Article history:
Received 6 August 2015
Received in revised form
14 July 2016
Accepted 6 October 2016

Keywords:
Noise covariance estimation
Nonlinear system
Expectation maximization
State estimation
Full information estimation
Moving horizon estimation
Extended Kalman filter

ABSTRACT

In order to estimate states from a noise-driven state space system, the state estimator requires a priori knowledge of both process and output noise covariances. Unfortunately, noise statistics are usually unknown and have to be determined from output measurements. Current expectation maximization (EM) based algorithms for estimating noise covariances for nonlinear systems assume the number of additive process and output noise signals are the same as the number of states and outputs, respectively. However, in some applications, the number of additive process noises could be less than the number of states. In this paper, a more general nonlinear system is considered by allowing the number of process and output noises to be smaller or equal to the number of states and outputs, respectively. In order to estimate noise covariances, a semi-definite programming solver is applied, since an analytical solution is no longer easy to obtain. The expectation step in current EM algorithms rely on state estimates from the extended Kalman filter (EKF) or smoother. However, the instability and divergence problems of the EKF could cause the EM algorithm to converge to a local optimum that is far away from true values. We use moving horizon estimation instead of the EKF/smoother so that the accuracy of the covariance estimation in nonlinear systems can be significantly improved.

1. Introduction

For noise-driven state space systems, the performance of state estimation depends on properly defined noise statistics. Any guessed or inappropriately determined noise statistics may result in inaccuracies or even a divergence of state estimates. The pioneering work in Mehra (1972) introduced the innovation correlation and the maximum likelihood estimation (MLE) methods for identifying the noise covariance.

The innovation correlation method is based on establishing an explicit relation between the noise covariance and the correlation of the innovation sequence. One of the algorithms for linear time invariant (LTI) systems is the auto-covariance least squares (ALS) method introduced in Rajamani and Rawlings (2009), in which the noise covariances are estimated by solving one semi-definite linear least squares problem. The ALS method was then extended to linear time-varying (LTV) as well as nonlinear systems in Ge and Kerrigan (2014) and Rajamani and Rawlings (2007).

As an alternative approach, MLE aims to maximize the likelihood of the noise covariance given the output measurement sequence. An MLE-based "one-off" and an EM-based recursive algorithm for estimating noise covariances in LTI and LTV systems are presented in Zagrobelny and Rawlings (2015) and Shumway and Stoffer (1982), respectively. The EM method was first introduced in Dempster, Laird, and Rubin (1977) and was extended to nonlinear systems in Bavdekar, Deshpande, and Patwardhan (2011) by using the extended Kalman filter (EKF) and Kalman smoother combined with the EM method.

In this paper, we consider more general nonlinear systems than (Bavdekar et al., 2011) by allowing the number of process and output noises to be smaller or equal to the number of states and outputs, respectively. In our method, the noise covariance matrices are estimated using a semi-definite programming (SDP) solver, in order to ensure estimated noise covariances are positive definite. The EM algorithm is combined with moving horizon estimation (MHE) and full information estimation (FIE) (Rawlings & Mayne, 2009) to estimate noise covariances for nonlinear systems. The

☆ The material in this paper was not presented at any conference. This paper was recommended for publication in revised form by Associate Editor Erik Weyer under the direction of Editor Torsten Söderström.

E-mail addresses: ming.ge08@imperial.ac.uk (M. Ge), e.kerrigan@imperial.ac.uk (E.C. Kerrigan).

http://dx.doi.org/10.1016/j.automatica.2016.11.011
0005-1098/

first reason for using MHE/FIE, rather than other Kalman-based filters, is to reduce the possibility of instability and divergence of estimated states. The second reason is that MHE/FIE allows one to add constraints, which could improve the accuracy of state estimation, if there exists physical limits or algebraic constraints on system states.

This paper is organized as follows: Sections 2 and 3 introduce the theory of covariance estimation using the EM algorithm combined with MHE/FIE for nonlinear systems. Two numerical examples are given in Section 4. We draw conclusions in Section 5.

$\mathbb{E}\{X|Y\}$ and $\mathbb{C}\{X|Y\}$ denote the conditional expected value and covariance of a random variable X given Y, respectively. $\mathcal{L}(\cdot)$ and $\mathcal{L}_\ell(\cdot)$ denote the likelihood and log-likelihood function of parameters of a statistical model, respectively. $(\cdot)^\dagger$ denotes the Moore–Penrose generalized inverse of a matrix. $\delta(x-m)$ is the Dirac delta function with delay m. $\mathscr{M}_l^{r,c}$ denotes an $r\times c$ auxiliary matrix that

$$\mathscr{M}_l^{r,c} := \begin{bmatrix}\mathbf{0}_{r\times(l-1)} & I_r & \mathbf{0}_{r\times(c-r-l+1)}\end{bmatrix}.$$

$|\cdot|$ and $\log|\cdot|$ denotes the determinant and log-determinant of a square matrix, respectively. $\mathscr{J}(\cdot)$ denotes the Jacobian matrix of a vector-valued function. $\|\cdot\|_F$ is the Frobenius norm of a matrix. $P\succ 0$ denotes that P is a positive-definite symmetric matrix; $\|x\|_W^2$ is weighted least squares of vector x, which equals to $x^\top Wx$. $\mathrm{tr}(A)$ represents the trace of a square matrix A. The symbol $:=$ is to be read as 'is defined as', while $=:$ is to be read as 'defines'.

2. Noise covariance estimation using the EM algorithm

Consider a discrete-time nonlinear state space model:

$$\begin{aligned} x_{k+1} &:= f(x_k) + G_k w_k \\ y_k &:= h(x_k) + H_k v_k \end{aligned} \tag{1}$$

where $x_k\in\mathbb{X}_k$ is the unknown state, $y_k\in\Re^p$ is the output measurement. $G_k\in\Re^{n\times r}$ and $H_k\in\Re^{p\times q}$ are two full column rank time-varying matrices in order to ensure the uniqueness of the conditional densities $p(x_{k+1}|x_k)$ and $p(y_k|x_k)$ to be defined later. w_k and v_k are two unknown noise terms, which affect the state and output, respectively.

Assumption 1. The noise sequences $(w_k)_{k=1}^M$ and $(v_k)_{k=1}^M$ are two random variables having Gaussian (or normal) distributions $\mathcal{N}(0,Q)$ and $\mathcal{N}(0,R)$, respectively, with zero mean and unknown positive-definite covariance matrices Q and R.

Assumption 2. The probability distribution $p(x_1)$ of the initial state has Gaussian distribution $p(x_1)\sim\mathcal{N}\left(\tilde{x}_1,\ P_1\right)$, where $\tilde{x}_1$ is the *a priori* most likely value of x_1 and P_1 is the corresponding error covariance.

Assumption 3. Functions $f(\cdot)$ and $h(\cdot)$ are twice differentiable. The discrete-time nonlinear model (1) is uniformly observable (Moraal & Grizzle, 1995) and there exists a stable state observer for (1) with nonempty feasible region.

Define the full output and state sequence to be $Y_M := (y_k)_{k=1}^M$ and $X_M := (x_k)_{k=1}^M$, respectively. If Y_M and initial guesses of $\tilde{x}_1$, P_1, Q and R are all given, the set of true system parameters $\mathscr{O} := \{\tilde{x}_1, P_1, Q, R\}$ can be recursively estimated using the expectation maximization (EM) method (Shumway & Stoffer, 1982).

Let the estimate of $\tilde{x}_1$ and covariance matrices P_1, Q and R at the ith iteration be

$$\mathscr{O}_i := \left(\tilde{x}_{1,i}, P_{1,i}, Q_i, R_i\right), \quad i=1,\dots,N,$$

where $N\gg 1$ is the maximum number of iterations. The EM method recursively maximizes the expectation of the log-likelihood function $\mathscr{O}_i\mapsto\mathcal{L}_\ell(\mathscr{O}_i|Y_M)$, until the log-likelihood function converges to its maximum value (Dempster et al., 1977).

The expression of $\mathscr{O}\mapsto\mathcal{L}_\ell(\mathscr{O}|Y_M)$ is given by

$$\begin{aligned}\mathcal{L}_\ell(\mathscr{O}|Y_M) = \log(p(Y_M|\mathscr{O})) &= \log\left(\frac{p(X_M,Y_M|\mathscr{O})}{p(X_M|Y_M,\mathscr{O})}\right)\\ &= \log(p(X_M,Y_M|\mathscr{O})) - \log(p(X_M|Y_M,\mathscr{O})).\end{aligned}\tag{2}$$

Taking the conditional expectation on both sides of (2) given Y_M and $\mathscr{O}_{i-1}$, we get the expectation of the log-likelihood function $\mathcal{L}_\ell(\mathscr{O}|Y_M)$, i.e.

$$\mathbb{E}\{\log(p(Y_M|\mathscr{O}))|Y_M,\mathscr{O}_{i-1}\} = \mathcal{Q}(\mathscr{O}|\mathscr{O}_{i-1}) - \mathcal{H}(\mathscr{O}|\mathscr{O}_{i-1}),$$

where $\mathcal{Q}$ and $\mathcal{H}$ are given by

$$\mathcal{Q}(\mathscr{O}|\mathscr{O}_{i-1}) := \mathbb{E}\{\log(p(X_M,Y_M|\mathscr{O}))|Y_M,\mathscr{O}_{i-1}\}, \tag{3a}$$

$$\mathcal{H}(\mathscr{O}|\mathscr{O}_{i-1}) := \mathbb{E}\{\log(p(X_M|Y_M,\mathscr{O}))|Y_M,\mathscr{O}_{i-1}\}. \tag{3b}$$

Because Y_M is a given measurement sequence,

$$\mathbb{E}\{\log(p(Y_M|\mathscr{O}))|Y_M,\mathscr{O}_{i-1}\} = \mathcal{L}_\ell(\mathscr{O}|Y_M).$$

Theorem 4 (Dempster et al., 1977; Jeff Wu, 1983)**.** *For the EM algorithm, the value of $\mathcal{L}_\ell(\mathscr{O}_i|Y_M)$ will monotonically increase at each iteration and converge to the maximum if*

$$\mathcal{Q}(\mathscr{O}_i|\mathscr{O}_{i-1}) \geq \mathcal{Q}(\mathscr{O}_{i-1}|\mathscr{O}_{i-1}), \quad \forall i, \tag{4}$$

with equality if and only if $\mathscr{O}_i=\mathscr{O}_{i-1}$.

Theorem 4 simplifies calculation, so that we only need to focus on fulfilling (4) by maximizing $\mathcal{Q}(\mathscr{O}|\mathscr{O}_{i-1})$, instead of $\mathcal{L}_\ell(\cdot|Y_M)$. Before giving the expression of $\mathcal{Q}$, we start with the expression of $\log(p(X_M,Y_M|\mathscr{O}))$ in (3a).

Lemma 5. *Consider two random vectors w, x and let $x=m+Gw$, where $m\in\Re^{n,1}$, $G\in\Re^{n,r}$ and $n>r$. If G is a full column rank matrix, $w\sim\mathcal{N}(\mathbf{0},Q)$ and $Q\succ 0$, then the probability distribution of x is given by*

$$p(x) \propto \frac{1}{\sqrt{(2\pi)^r|W|}} e^{-\frac{1}{2}\|S(x-m)\|^2_{W^{-1}}},$$

where $W\succ 0$ and S is a constant matrix.

Proof. Because G is a full column rank matrix, random vector x has a singular joint normal distribution, such that $p(x)\sim\mathcal{N}\left(\mathbf{0},GQG^\top\right)$, where $GQG^\top\succeq 0$. By using the singular value decomposition, we have

$$GQG^\top = UPU^\top = U\begin{bmatrix}\tilde{Q} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}U^\top,$$

where $\tilde{Q}\in\Re^r$ is a nonsingular matrix, U is a unitary matrix, such that $U^{-1}=U^\top$. Define a new random vector z, such that

$$z := \begin{bmatrix}z_1^\top & z_2^\top\end{bmatrix}^\top = U^{-1}x = U^{-1}Gw,$$

where $z_1\in\Re^{r\times 1}$. Since $z\sim\mathcal{N}(U^{-1}m,P)$, by definition of the singular joint normal distribution (Graham & Rawlings, 2013, pp. 376–377),

$$p(z) = \frac{\delta(z_2-\mathscr{M}_{r+1}^{n-r,n}U^{-1}m)}{\sqrt{(2\pi)^r|\tilde{Q}|}} e^{-\frac{1}{2}\|z_1-\mathscr{M}_1^{r,n}U^{-1}m\|^2_{\tilde{Q}^{-1}}}.$$

Because $\delta(z_2-\mathscr{M}_{r+1}^{n-r,n}U^{-1}m)$ is the probability mass function of the degenerate variable z_2, we have

$$p(z) \propto \frac{1}{\sqrt{(2\pi)^r|\tilde{Q}|}} e^{-\frac{1}{2}\|z_1-\mathscr{M}_1^{r,n}U^{-1}m\|^2_{\tilde{Q}^{-1}}}.$$