Automatica 62 (2015) 26–31

Contents lists available at ScienceDirect

Automatica

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

Brief paper

A novel suboptimal method for solving polynomial filtering problems☆

Xue Luo [a], Yang Jiao [b], Wen-Lin Chiou [c], Stephen S.-T. Yau [d,1]

[a] School of Mathematics and Systems Science, Beihang University, Beijing, 100191, PR China

[b] Department of Mathematics, Statistics and Computer Science, University of Illinois at Chicago, Science and Engineering Offices (M/C 249), 851 S. Morgan Street, Chicago, IL 60607-7045, USA

[c] Department of Mathematics, Fu-Jen University, New Taipei City, Taiwan, ROC

[d] Department of Mathematical Sciences, Tsinghua University, Beijing, 100084, PR China

ARTICLE INFO

Article history:
Received 22 April 2014
Received in revised form
31 May 2015
Accepted 19 August 2015

Keywords:
Polynomial filtering problems
Suboptimal filter
Recursive algorithm
Estimation theory

ABSTRACT

In this paper we derive the stochastic differentials of the conditional central moments of the nonlinear filtering problems, especially those of the polynomial filtering problem, and develop a novel suboptimal method by solving this evolution equation. The basic idea is to augment the state of the original nonlinear system by including the original states' conditional central moments such that the augmented states form a so-called bilinear system after truncating. During our derivation, it is clear to see that the stochastic differentials of the conditional central moments of the linear filtering problem (i.e., f, g and h are all at most degree one polynomials) form a closed system automatically without truncation. This gives one reason for the existence of optimal filtering for linear problems. On the contrary, the conditional central moments form an infinite dimensional system, in general. To reduce it to a closed-form, we let all the high enough central moments to be zero, as one did in the Carleman approach (Germani et al., 2007). Consequently, a novel suboptimal method is developed by dealing with the bilinear system. Numerical simulation is performed for the cubic sensor problem to illustrate the accuracy and numerical stability.

1. Introduction

The nonlinear filtering (NLF) problem has been extensively studied since the linear one has been satisfactorily solved by Kalman in 1960s. But until now there exists no universal optimal method for the general nonlinear settings. The main goal of NLF is to get "good" estimation of the conditional expectation, or perhaps even the conditional density function of the state, given the observation history. We refer the readers to the book by Jazwinski (1970) for excellent introduction to NLF.

One possible general method to NLF is the so-called global approaches, see the survey paper (Luo, 2014) for more detailed discussion. All these methods try to solve analytically or numerically for the conditional density function. One of the most recent global approach is Yau–Yau's on- and off-line algorithm, which is first derived in Yau and Yau (2008) and further generalized to nonlinear time-varying setting in Luo and Yau (2013). The Yau–Yau's method works for all NLF problems theoretically, however, there are still some technical works to be done for high-dimensional states' problem, say to overcome "the curse of dimensionality". Therefore, certain suboptimal methods still need to be developed.

Another possible way-out to solve the general NLF problems stems from the local approaches, especially the Kalman filter and its derivatives. The basic idea is to augment the states of the original NLF problem in certain way such that the augmented states satisfy a linear or so-called bilinear system (Carravetta, Germani, & Shuakayev, 2000). Generally speaking, one cannot obtain a closed system unless it is Benes' filter (Benes, 1981) or Yau filter (Yau, 1994). The suboptimal filtering therefore is derived by truncating the infinite-dimensional system in some way. In this direction, Basin (2003) is the first paper where the conditional higher order moments were employed for suboptimal polynomial filtering (PF). Later, a series of papers, say (Basin, 2008; Basin, Shi, & Calderon-Alvarez, 2010) and references therein follow this

☆ The material in this paper was not presented at any conference. This paper was recommended for publication in revised form by Associate Editor Michael V. Basin under the direction of Editor Ian R. Petersen.

E-mail addresses: xluo@buaa.edu.cn (X. Luo), jiaoyang0715@gmail.com (Y. Jiao), chiou@math.fju.edu.tw, kubulinchiou@gmail.com (W.-L. Chiou), yau@uic.edu (S.S.-T. Yau).

[1] Tel.: +86 10 62787874; fax: +86 10 62798033.

http://dx.doi.org/10.1016/j.automatica.2015.09.001
0005-1098/

line. Based on the suboptimal approach introduced for the bilinear system in Carravetta et al. (2000) and Germani, Manes, and Palumbo (2007) developed a Carleman approximation approach for the NLF problems. In their paper, the higher moments are omitted to form a finite-dimensional system. However, it is as early as in 1967 that Kushner (1967) considered the moment sequences. Even in the one-dimensional problem, the moment sequence has to satisfy the following inequalities:

$$m_2 > 0,\ m_4 > m_2^2,\ m_6 > m_4^2/m_2, \ldots,$$

where $m_s,\ s = 1, 2, \ldots$ represent the s-moment of some random variable. In particular, if the random variable is the standard Gaussian, then all the higher moments can be computed explicitly:

$$m_s = \begin{cases} 0, & \text{all odd } s \geq 1 \\ (s-1)!!\, m_2^s, & \text{all even } s \geq 2, \end{cases}$$

where $(s-1)!! = 1 \cdot 3 \cdot 5 \cdot \cdots \cdot (s-1)$, if s is even. It is easy to see that no matter how small m_2 is, the even moments grows without bound as $s \to \infty$. Therefore, it is inappropriate to let all the higher moments to be zero, even in the Gaussian case.

In this paper, we propose a novel suboptimal method (NSM) for the PF problems by observing the evolution of the conditional *central* moments of the states. Instead of augmenting the states by their higher moments as in the Carleman approach, we derive the evolution of the higher *central* moments, and omit the high enough ones to form a finite-dimensional system. Another novelty of this paper is that we provide an explanation why the optimal method can be derived only for the linear/bilinear filtering problems from the viewpoint of the evolution of the conditional higher central moments. According to Theorem 2, the stochastic differentials of the conditional central moments form a closed system automatically without truncation, if the linear/bilinear filtering problem is considered.

2. Filtering model and notations

The model we consider here is:

$$\begin{cases} dx_t = f(x_t, t)dt + g(t)dv_t \\ dy_t = h(x_t, t)dt + dw_t, \end{cases} \tag{2.1}$$

where x_t, v_t, y_t, and w_t are $\mathbb{R}^n-$, $\mathbb{R}^p-$, $\mathbb{R}^m-$, and $\mathbb{R}^m-$valued processes, respectively, and $f : \mathbb{R}^n \times \mathbb{R}_+ \to \mathbb{R}^n$, $g : \mathbb{R}_+ \to \mathbb{R}^{n\times p}$, $h : \mathbb{R}^n \times \mathbb{R}_+ \to \mathbb{R}^m$ are polynomials with respect to x. Assume that $\{v_t, t \geqslant 0\}$ and $\{w_t, t \geqslant 0\}$ are Brownian motion processes with $\mathrm{Var}[dv_t] = Q(t)dt$ and $\mathrm{Var}[dw_t] = R(t)dt$, respectively. Moreover, $\{v_t, t \geqslant 0\}$, $\{w_t, t \geqslant 0\}$ and x_0 are independent, and $y_0 = 0$.

The conditional expectation for certain process x_t is denoted as $\widehat{x_t} := E[x_t \mid Y_t]$ for short, where $Y_t := \{y_s : 0 \leq s \leq t\}$ is the observation history. Also the a priori conditional expectation is denoted as $\widehat{(\circ)}_- := E[\circ \mid Y_{t_-}]$, where $Y_- := \{y_s : 0 \leq s \leq t_-\}$.

In this paper, we shall use the Kronecker algebra for conciseness. For the quick survey on the Kronecker product and its properties can be found in Carravetta, Germani, and Raimondi (1996). For the readers' convenience, we include some simple facts of Kronecker algebra here. The Kronecker product $\otimes$ is defined for any two matrices $M_{r\times s}$ and $N_{p\times q}$:

$$M \otimes N := \begin{bmatrix} m_{11}N & \cdots & m_{1s}N \\ \cdots & \cdots & \cdots \\ m_{r1}N & \cdots & m_{rs}N \end{bmatrix}.$$

Let $M^{[i]}$ denote the ith Kronecker power of the matrix M, which is defined as

$$M^{[0]} = 1; \qquad M^{[i]} = M \otimes M^{[i-1]} = M^{[i-1]} \otimes M.$$

The stack of the matrix $M_{r\times s} := [m_1, m_2, \ldots, m_s]$ is defined as

$$\mathrm{st}(M) = \begin{bmatrix} m_1^T & m_2^T & \cdots & m_s^T \end{bmatrix}^T,$$

where m_i is the i^{th} column of M. The inverse operation of the stack can reduce a vector into a matrix with proper size. That is, $M = \mathrm{st}_{r\times s}^{-1}\mathrm{st}(M)$. The following are properties of Kronecker product and the stack operation:

$$(A + B) \otimes (C + D) = A \otimes C + A \otimes D + B \otimes C + B \otimes D \tag{2.2}$$

$$(A \otimes B) \otimes C = A \otimes (B \otimes C) \tag{2.3}$$

$$(A \cdot C) \otimes (B \cdot D) = (A \otimes B) \cdot (C \otimes D), \tag{2.4}$$

$$u \otimes v = \mathrm{st}(v \cdot u^T), \tag{2.5}$$

where A, B, C, and D are matrices with suitable size, u and v are vectors. As the usual matrix multiplication, Kronecker product is not commutative. Given any two matrices $A \in \mathbb{R}^{r_a\times s_a}$ and $B \in \mathbb{R}^{r_b\times s_b}$, then

$$B \otimes A = C_{r_a,r_b}^T (A \otimes B) C_{s_a,s_b},$$

where $C_{a,b}$ is an orthonormal commutative matrix in $\{0, 1\}^{ab\times ab}$ with its entry (h, l) given by

$$\{C_{a,b}\}_{h,l} = \begin{cases} 1, & \text{if } l = (|h-1|_b)a + \left(\left[\dfrac{h-1}{b}\right] + 1\right) \\ 0, & \text{otherwise,} \end{cases}$$

where $[\cdot]$ and $|\cdot|_s$ denote the integer part and s-modulo, respectively. The Kronecker power of a binomial, $(a + b)^{[i]}$ allows the following expansion:

$$(a + b)^{[i]} = \sum_{j=0}^{i} M_j^i \left(a^{[j]} \otimes b^{[i-j]}\right), \tag{2.6}$$

for any $a, b \in \mathbb{R}^n$, where $M_j^i \in \mathbb{R}^{n^i\times n^i}$ can be recursively computed

$$M_h^h = M_0^h = I_{n^h},$$
$$M_j^h = (M_j^{h-1} \otimes I_n) + (M_{j-1}^{h-1} \otimes I_n)(I_{n^{j-1}} \otimes G_{h-j}),$$

and $\{G_l\}$ is a sequence that satisfies the following equations:

$$G_1 = C_{n,n}^T, \qquad G_l = (I_n \otimes G_{l-1}) \cdot (G_1 \otimes I_{n^{l-1}}).$$

See detailed derivation in Carravetta et al. (1996).

In the derivation of our NSM for PF problems, the Itô formula for the computation of stochastic differentials is needed (see Liptser & Shiryayev, 1977). For any vector function $\psi(x_t) : \mathbb{R}^n \to \mathbb{R}^r$, we have

$$d\psi = (\nabla_x \otimes \psi)|_{x_t}\, dx_t + \frac{1}{2}\left(\nabla_x^{[2]} \otimes \psi\right)\Big|_{x_t} (dx_t \otimes dx_t). \tag{2.7}$$

The differential operator $\nabla_x^{[i]}\otimes$ applied to ψ is defined as

$$\nabla_x^{[0]} \otimes \psi = \psi, \qquad \nabla_x^{[i+1]} \otimes \psi = \nabla_x \otimes \left(\nabla_x^{[i]} \otimes \psi\right), \quad i \geq 1,$$

with $\nabla_x = \left[\frac{\partial}{\partial x_1}\ \frac{\partial}{\partial x_2}\ \cdots\ \frac{\partial}{\partial x_n}\right]$.

3. NSM for PF problems

3.1. Derivation of stochastic differentials for higher central moments

It is well-known from Jazwinski (1970) that the conditional mean of the process x_t satisfies

$$\begin{aligned} d\hat{x}_t &= \hat{f}dt + \widehat{P^{[1]}h^T}R^{-1}\left(dy - \hat{h}dt\right) \\ &= \left[\hat{f} + \widehat{P^{[1]}h^T}R^{-1}(h - \hat{h})\right]dt + \widehat{P^{[1]}h^T}R^{-1}dw_t, \end{aligned} \tag{3.8}$$