Optimal Filtering for Incompletely Measured Polynomial States over Linear Observations

Michael Basin Dario Calderon-Alvarez
Department of Physical and Mathematical Sciences
Autonomous University of Nuevo Leon, San Nicolas de los Garza
Nuevo Leon, Mexico
mbasin@fcfm.uanl.mx dcalal@hotmail.com

Mikhail Skliar
Department of Chemical and Fuels Engineering
University of Utah, Salt Lake City, Utah, USA
Mikhail.Skliar@m.cc.utah.edu

Abstract

In this paper, the optimal filtering problem for incompletely measured polynomial system states over linear observations is treated proceeding from the general expression for the stochastic Ito differential of the optimal estimate and the error variance. In contrast to the previous works, the nonlinear polynomial states are allowed to be unmeasured in this problem. The procedure for obtaining a closed system of the filtering equations for any polynomial state over linear observations is then established, which yields the explicit closed form of the filtering equations in the particular case of a bilinear state equation. In the example, performance of the designed optimal filter is verified against a conventional extended Kalman-Bucy filter.[1]

1 Introduction

This paper presents the optimal finite-dimensional filter for incompletely measured polynomial system states over linear observations with an arbitrary, not necessarily invertible, observation matrix, thus generalizing the results of ([3, 2]). In contrast to [3], the nonlinear polynomial states are allowed to be unmeasured in this framework, whereas only linear unmeasurable components of polynomial states are allowed in [3]. This significantly complicates solution of the optimal filtering problem, since the direct derivation of the filtering equation from the Ito differentials, employed in ([3, 2]), is no longer possible, and a transformation of the original filtering system should first be conducted.

The optimal filtering problem is treated proceeding from the general expression for the stochastic Ito differential of the optimal estimate and the error variance [4]. As the first result, the Ito differentials for the optimal estimate and error variance corresponding to the stated filtering problem are derived. Next, a transformation of the observation equation is introduced to reduce the original problem to the previously solved one with an invertible observation matrix [2]. The procedure for obtaining a closed system of the filtering equations for any polynomial state over linear observations is then established, which yields the explicit closed form of the filtering equations in the particular case of a bilinear state equation. In the illustrative example, performance of the designed optimal filter is verified for a quadratic bidimensional state over linear observations against a conventional extended Kalman-Bucy filter, which is still the most common tool for obtaining suboptimal estimates for nonlinear system states. The simulation results show a definite advantage in favor of the designed optimal filter.

2 Filtering Problem for Incompletely Measured Polynomial States

Let (Ω, F, P) be a complete probability space with an increasing right-continuous family of σ-algebras $F_t, t \geq t_0$, and let $(W_1(t), F_t, t \geq t_0)$ and $(W_2(t), F_t, t \geq t_0)$ be independent Wiener processes. The F_t-measurable random process $(x(t), y(t))$ is described by a nonlinear differential equation with both polynomial drift and diffusion terms for the sys-

[1]The authors thank the US National Science Foundation (NSF) and the Mexican National Science and Technology Council (CONACyT) for financial support under Grants CTS-0117300 and 39388-A, 52953-A, respectively.

tem state

$$dx(t) = f(x,t)dt + b(t)dW_1(t), \quad x(t_0) = x_0, \qquad (1)$$

and a linear differential equation for the observation process

$$dy(t) = (A_0(t) + A(t)x(t))dt + B(t)dW_2(t). \qquad (2)$$

Here, $x(t) \in R^n$ is the state vector and $y(t) \in R^m$ is the linear observation vector, $m \le n$. The initial condition $x_0 \in R^n$ is a Gaussian vector such that x_0, $W_1(t) \in R^p$, and $W_2(t) \in R^q$ are independent. In contrast to the previously obtained results (see [2]), the observation matrix $A(t) \in R^{m\times n}$ is not supposed to be invertible or even square. It is assumed that $B(t)B^T(t)$ is a positive definite matrix, therefore, $m \le q$. All coefficients in (1)–(2) are deterministic functions of appropriate dimensions.

The nonlinear function $f(x,t)$ is considered polynomial of n variables, components of the state vector $x(t) \in R^n$, with time-dependent coefficients. Since $x(t) \in R^n$ is a vector, this requires a special definition of the polynomial for $n > 1$. In accordance with [2], a p-degree polynomial of a vector $x(t) \in R^n$ is regarded as a p-linear form of n components of $x(t)$:

$$f(x,t) = a_0(t) + a_1(t)x + \ldots + a_p(t)x \ldots_{p\ times} \ldots x, \qquad (3)$$

where $a_0(t)$ is a vector of dimension n, a_1 is a matrix of dimension $n\times n$, a_2 is a 3D tensor of dimension $n\times n\times n$, a_p is an $(p+1)$D tensor of dimension $n\times \ldots_{(p+1)\ times} \ldots \times n$, and $x\times \ldots_{p\ times} \ldots \times x$ is a pD tensor of dimension $n\times \ldots_{p\ times} \ldots \times n$ obtained by p times spatial multiplication of the vector $x(t)$ by itself.

The estimation problem is to find the optimal estimate $\hat{x}(t)$ of the system state $x(t)$, based on the observation process $Y(t) = \{y(s), t_0 \le s \le t\}$, that minimizes the Euclidean 2-norm $J = E[(x(t) - \hat{x}(t))^T (x(t) - \hat{x}(t)) \mid F_t^Y]$ at every time moment t. Here, $E[z(t) \mid F_t^Y]$ means the conditional expectation of a stochastic process $z(t) = (x(t) - \hat{x}(t))^T (x(t) - \hat{x}(t))$ with respect to the σ - algebra F_t^Y generated by the observation process $Y(t)$ in the interval $[t_0,t]$. As known [4], this optimal estimate is given by the conditional expectation $\hat{x}(t) = m(t) = E(x(t) \mid F_t^Y)$ of the system state $x(t)$ with respect to the σ - algebra F_t^Y generated by the observation process $Y(t)$ in the interval $[t_0,t]$. As usual, the matrix function $P(t) = E[(x(t) - m(t))(x(t) - m(t))^T \mid F_t^Y]$ is the estimation error variance.

3 Optimal Filter for Incompletely Measured Polynomial States

The following non-closed system of the optimal filtering equations is obtained using the formula for the Ito differential of the conditional expectation $m(t) = E(x(t) \mid F_t^Y)$ ([4])

$$dm(t) = E(f(x,t) \mid F_t^Y)dt + P(t)A^T(t)(B(t)B^T(t))^{-1} \times$$
$$(dy(t) - (A_0(t) + A(t)m(t))dt). \qquad (4)$$

$$dP(t) = (E((x(t) - m(t))(f(x,t))^T \mid F_t^Y) +$$
$$E(f(x,t)(x(t) - m(t))^T) \mid F_t^Y) +$$
$$+b(t)b^T(t) - P(t)A^T(t)(B(t)B^T(t))^{-1}A(t)P(t))dt +$$
$$E(((x(t) - m(t))(x(t) - m(t))(x(t) - m(t))^T \mid F_t^Y) \times$$
$$A^T(t)(B(t)B^T(t))^{-1}(dy(t) - A(t)m(t))dt). \qquad (5)$$

The equation (4) and (5) should be complemented with the initial conditions $m(t_0) = E(x(t_0) \mid F_{t_0}^Y)$ and $P(t_0) = E[(x(t_0) - m(t_0)(x(t_0) - m(t_0)^T \mid F_{t_0}^Y]$.

The equations (4) and (5) for the optimal estimate $m(t)$ and the error variance $P(t)$ form a non-closed system of the filtering equations. As shown in [2], a closed system of the filtering equations for a system state (1) with polynomial drift and state-independent diffusion over linear observations can be obtained, if the observation matrix $A(t)$ is invertible for any $t \ge t_0$. Since the observation matrix $A(t)$ in (2) is not necessarily invertible, the following transformations are introduced.

First, note that the matrix A can always be assumed a matrix of complete rank, m, which is equal to the dimension of the linearly independent observations $y(t) \in R^m$; if not so, excessive linearly dependent observations, corresponding to linearly dependent rows of the matrix A, must be removed from consideration. In doing so, the number of Wiener processes in the observation equations can also be reduced to m, the dimension of independent observations, by summarizing and re-numerating the Wiener processes in each observation equation (2). Therefore, the matrix B can always be assumed a square matrix of dimension $m\times m$, such that $B(t)B^T(t)$ is a positive definite matrix (see Section 2 for this condition). Next, the new matrices $\bar{A}(t)$ and $\bar{B}(t)$ are defined as follows. The matrix $\bar{A}(t) \in R^{n\times n}$ is obtained from $A(t) \in R^{m\times n}$ by adding $n - m$ linearly independent rows such that the resulting matrix $\bar{A}(t)$ is invertible. The matrix $\bar{B}(t) \in R^{n\times n}$ is made from the matrix $B(t) \in R^{m\times m}$ by placing $B(t)$ in the upper left corner of $\bar{B}(t)$, defining the other $n - m$ diagonal entries of $\bar{B}(t)$ equal to infinity, and setting to zero all other entries of $\bar{B}(t)$ outside the main diagonal or outside the submatrix $B(t)$. In other words, $\bar{B}(t) = diag[B(t), \beta I_{(n-m)\times(n-m)}]$, where $\beta = \infty$, and $I_{(n-m)\times(n-m)}$ is the identity matrix of dimension $(n-m)\times(n-m)$. Thus, the new observation equation is given by

$$\bar{y}(t) = (\bar{A}_0(t) + \bar{A}(t)x(t))dt + \bar{B}(t)dW_2(t), \qquad (6)$$

where $\bar{y}(t) \in R^n$, $\bar{A}_0(t) = [A_0^T(t), 0_{n-m}]^T \in R^n$, and 0_{n-m} is a vector of $n - m$ zeros.

The key point of the introduced transformation is that the new observation process $\bar{y}(t)$ is physically equivalent to the old one $y(t)$, since the fictitious last $n - m$ components of $\bar{y}(t)$ consist of pure noise in view of infinite intensities of white Gaussian noises in the corresponding $n-m$ equations, and the first m components of $\bar{y}(t)$ coincide with $y(t)$. In addition, the entire observation matrix $\bar{A}(t)$ is invertible, and the matrix $(\bar{B}(t)\bar{B}^T(t))^{-1} \in R^{n\times n}$ exists and equals to the $n\times n$ – dimensional square matrix, whose upper left corner is occupied by the submatrix $(B(t)B^T(t))^{-1} \in R^{m\times m}$ and all other entries are zeros.

In terms of the new observation equation (6), the filtering equations (4) and (5) take the form

$$dm(t) = E(f(x,t) \mid F_t^Y)dt + P(t)\bar{A}^T(t)(\bar{B}(t)\bar{B}^T(t))^{-1} (d\bar{y}(t) - (\bar{A}_0(t) + \bar{A}(t)m(t))dt), \quad (7)$$

$$dP(t) = (E((x(t)-m(t))(f(x,t))^T \mid F_t^Y) + E(f(x,t)(x(t)-m(t))^T) \mid F_t^Y) + +b(t)b^T(t) - P(t)\bar{A}^T(t)(\bar{B}(t)\bar{B}^T(t))^{-1}\bar{A}(t)P(t))dt + E(((x(t)-m(t))(x(t)-m(t))(x(t)-m(t))^T \mid F_t^Y)\times \bar{A}^T(t)(\bar{B}(t)\bar{B}^T(t))^{-1}(d\bar{y}(t) - \bar{A}(t)m(t))dt). \quad (8)$$

Since the new observation matrix $\bar{A}(t)$ is invertible for any $t \geq t_0$, the random variable $x(t)-m(t)$ is conditionally Gaussian with respect to the observation process $y(t)$ for any $t \geq t_0$ (see [2]). Hence, the technique for representing the superior conditional moments of $x(t)-m(t)$, based on the properties of Gaussian variables, is applicable to obtaining a closed system of the filtering equations for any polynomial function $f(x,t)$ in (1) (see [2] for more substantiation and details of this technique). In doing so, the following variance equation is obtained

$$dP(t) = (E((x(t)-m(t))(f(x,t))^T \mid F_t^Y) + E(f(x,t)(x(t)-m(t))^T) \mid F_t^Y) + b(t)b^T(t) - P(t)\bar{A}^T(t)(\bar{B}(t)\bar{B}^T(t))^{-1}\bar{A}(t)P(t))dt. \quad (9)$$

Finally, in view of definition of the matrices $\bar{A}(t)$ and $\bar{B}(t)$ and the new observation process $\bar{y}(t)$, the filtering equations (7),(9) can be written again in terms of the original observation equation (2) using $y(t)$, $A(t)$, and $B(t)$

$$dm(t) = E(f(x,t) \mid F_t^Y)dt + P(t)A^T(t)(B(t)B^T(t))^{-1}\times (dy(t) - (A_0(t) + A(t)m(t))dt), \quad (10)$$

$$dP(t) = (E((x(t)-m(t))(f(x,t))^T \mid F_t^Y) + E(f(x,t)(x(t)-m(t))^T) \mid F_t^Y) + b(t)b^T(t) - P(t)A^T(t)(B(t)B^T(t))^{-1}A(t)P(t))dt, \quad (11)$$

with the initial conditions $m(t_0) = E(x(t_0) \mid F_{t_0}^Y)$ and $P(t_0) = E[(x(t_0)-m(t_0)(x(t_0)-m(t_0)^T \mid F_{t_0}^Y]$.

In the next subsection, a closed form of the filtering equations will be obtained from (10) and (11) for a bilinear function $f(x,t)$ in the equation (1). It should be noted, however, that application of the same procedure would result in designing a closed system of the filtering equations for any polynomial function $f(x,t)$ in (1).

3.1 Optimal Filter for Bilinear States

Let the function

$$f(x,t) = a_0(t) + a_1(t)x + a_2(t)xx^T \quad (12)$$

be bilinear polynomial, where x is an n-dimensional vector, $a_0(t)$ is an n-dimensional vector, $a_1(t)$ is $n\times n$ - matrices, $a_2(t)$ is 3D tensor of dimension $n\times n\times n$. In this case, the representations for $E(f(x,t) \mid F_t^Y)$ and $E((x(t)-m(t))(f(x,t))^T \mid F_t^Y)$ as functions of $m(t)$ and $P(t)$ are derived as follows (see [2])

$$E(f(x,t) \mid F_t^Y) = a_0(t) + a_1(t)m(t) + a_2(t)m(t)m^T(t) + a_2(t)P(t), \quad (13)$$

$$E(f(x,t)(x(t)-m(t))^T) \mid F_t^Y) + E((x(t)-m(t))(f(x,t))^T \mid F_t^Y) = a_1(t)P(t) + P(t)a_1^T(t) + 2a_2(t)m(t)P(t) + 2(a_2(t)m(t)P(t))^T. \quad (14)$$

Substituting the expression (13) in (10) and the expression (14) in (11), the filtering equations for the optimal estimate $m(t)$ and the error variance $P(t)$ are obtained

$$dm(t) = (a_0(t) + a_1(t)m(t) + a_2(t)m(t)m^T(t) + a_2(t)P(t))dt + P(t)A^T(t)(B(t)B^T(t))^{-1}[dy(t) - A(t)m(t)dt], \quad (15)$$

$$m(t_0) = E(x(t_0) \mid F_t^Y)),$$

$$dP(t) = (a_1(t)P(t) + P(t)a_1^T(t) + 2a_2(t)m(t)P(t) + 2(a_2(t)m(t)P(t))^T + b(t)b^T(t))dt - P(t)A^T(t)(B(t)B^T(t))^{-1}A(t)P(t)dt. \quad (16)$$

$$P(t_0) = E((x(t_0)-m(t_0))(x(t_0)-m(t_0))^T \mid F_t^Y)).$$

4 Example

This section presents an example of designing the optimal filter for a quadratic bi-dimensional state over scalar linear observations and comparing it to a conventional extended Kalman-Bucy filter, which is still the most common tool for obtaining suboptimal estimates for nonlinear system states.

Let the bi-dimensional real state $x(t)$ satisfy the quadratic-linear system

$$\dot{x}_1(t) = x_2(t), \quad x_1(0) = x_{10}, \quad (17)$$

$$\dot{x}_2(t) = 0.1x_2^2(t) + \psi_1(t), \quad x_2(0) = x_{20},$$

and the scalar observation process be given by the linear equation

$$y(t) = x_1(t) + \psi_2(t), \quad (18)$$

where $\psi_1(t)$ and $\psi_2(t)$ are white Gaussian noises, which are the weak mean square derivatives of standard Wiener process (see [4]). The equations (17),(18) present the conventional form for the equations (1),(2), which is actually used in practice [1].

The filtering problem is to find the optimal estimate for the quadratic state(17), using incomplete linear observations (18) corrupted with independent and identically distributed disturbances modeled as white Gaussian noises. Since the solution of (17) goes to infinity at $T = 7.87$, the filtering horizon time is set to $T = 7.85$.

The filtering equations (15),(16) take the following particular form for the system (17),(18)

$$\dot{m}_1(t) = m_2(t) + P_{11}(t)[y(t) - m_1(t)], \tag{19}$$

$$\dot{m}_2(t) = 0.1m_2^2 + 0.1P_{22}(t) + P_{12}(t)[y(t) - m_1(t)],$$

with the initial condition $m(0) = E(x(0) \mid y(0)) = m_0$,

$$\dot{P}_{11}(t) = 2P_{12}(t) - P_{11}^2(t), \tag{20}$$

$$\dot{P}_{12}(t) = P_{22}(t) + 0.2m_2(t)P_{12}(t) - P_{11}(t)P_{12}(t),$$

$$\dot{P}_{22}(t) = 1 + 0.4m_2(t)P_{22}(t) - P_{12}^2(t),$$

with the initial condition $P(0) = E((x(0) - m(0))(x(0) - m(0))^T \mid y(0)) = P_0$.

The estimates obtained upon solving the equations (19)–(20) are also compared to the estimates satisfying the following extended Kalman-Bucy filtering equations for the quadratic-linear state (17) over the incomplete linear observations (18), which are obtained using the direct copy of the state dynamics (17) in the estimate equation and assigning the filter gain as the solution of the Riccati equation for the linearized system:

$$\dot{m}_{K1}(t) = m_{K2}(t) + P_{K11}(t)[y(t) - m_{K1}(t)], \tag{21}$$

$$\dot{m}_{K2}(t) = 0.1m_{K2}^2 + 0.1P_{K22}(t) + P_{K12}(t)[y(t) - m_{K1}(t)],$$

$$\dot{P}_{K11}(t) = 2P_{K12}(t) - P_{K11}^2(t), \tag{22}$$

$$\dot{P}_{K12}(t) = P_{K22}(t) + 0.2P_{K12}(t) - P_{K11}(t)P_{K12}(t),$$

$$\dot{P}_{K22}(t) = 1 + 0.4P_{K22}(t) - P_{K12}^2(t),$$

with the same initial conditions m_0 and P_0.

For each of the two filters (19)–(20) and (21)–(22), and the reference system (17)–(18), involved in simulation, the following initial values are assigned: $x_1(0) = 1.1$, $x_2(0) = 1.1$, $m_1(0) = 10.1$, $m_2(0) = 10.1$, $P_{11}(0) = 10$, $P_{12}(0) = 1$, $P_{22}(0) = 10$. Gaussian disturbances $\psi_1(t)$ and $\psi_2(t)$ in (17),(18) are realized using the built-in MatLab white noise function.

The following graphs are obtained: graphs of the errors between the the reference state components $x_1(t)$ and $x_2(t)$, satisfying the equations (17), and the optimal filter estimate components $m_1(t)$ and $m_2(t)$, satisfying the equations (19), are shown in Fig. 1; and graphs of the errors between the the reference state components $x_1(t)$ and $x_2(t)$, satisfying the equations (17), and the extended Kalman-Bucy filter estimate components $m_{K1}(t)$ and $m_{K2}(t)$, satisfying the equations (21), which are not shown. It can be observed that the error given by the optimal filter estimate (19) rapidly reaches and then maintains the zero mean value even in a close vicinity of the asymptotic time point $T = 7.87$, where the reference state (17) goes to infinity. On the contrary, the estimation errors given by the extended Kalman-Bucy filter behave unstably and diverge to infinity almost immediately at $T = 0.68$.

References

[1] K. J. Åström. *Introduction to Stochastic Control Theory*, volume 70 of *Mathematics in Science and Engineering*. Academic Press, New York, 1970.

[2] M. V. Basin, J. Perez, and M. Skliar. Optimal filtering for polynomial system states with polynomial multiplicative noise. *International Journal of Robust and Nonlinear Control*, 16:287–298, 2006.

[3] M. V. Basin and M. Skliar. Optimal filtering for partially measured polynomial system states. In *Proc. American Control Conf.*, pages 4022–4027, Portland, OR, 2005.

[4] V. S. Pugachev and I. N. Sinitsyn. *Stochastic Systems: Theory and Applications*. World Scientific, Singapore, 2001.

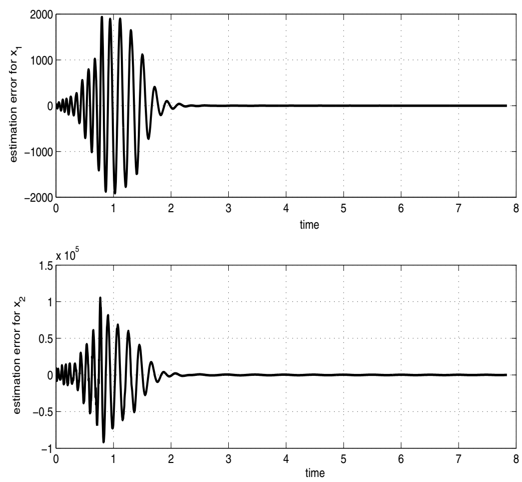

Figure 1. Graph of the error between the real states $x_1(t)$, $x_2(t)$ (17) and the optimal filter estimates $m_1(t)$, $m_2(t)$ (19), respectively, in the interval $[0, 7.84]$.