Filomat 31:19 (2017), 5979–5992
https://doi.org/10.2298/FIL1719979D

Published by Faculty of Sciences and Mathematics,
University of Niš, Serbia
Available at: http://www.pmf.ni.ac.rs/filomat

Filtering Method for Linear and Non-Linear Stochastic Optimal Control Of Partially Observable Systems

A. Delavarkhalafi[a], A. Poursherafatan[a]

[a]*Department of Applied Mathematics, Yazd University, Yazd, Iran.*

Abstract. This paper studies two linear methods for linear and non-linear stochastic optimal control of partially observable problem (SOCPP). At first, it introduces the general form of a SOCPP and states it as a functional matrix. A SOCPP has a payoff function which should be minimized. It also has two dynamic processes: state and observation. In this study, it is presented a deterministic method to find the control factor which has named feedback control and stated a modified complete proof of control optimality in a general SOCPP. After finding the optimal control factor, it should be substituted in the state process to make the partially observable system. Next, it introduces a linear filtering method to solve the related partially observable system with complete details. Finally, it is presented a heuristic method in discrete form for estimating non-linear SOCPPs and it is stated some examples to evaluate the performance of introducing methods.

1. Introduction

The optimal control theory had a great interest in the 1951s. This significant science has found many applications in various fields [1]. As an example of this issue we can see its applications in finance, physics, engineering and other sciences [2]. Today, this science plays an efficient role in many sciences. Due to the importance of this science, it has developed into the stochastic branches.

Everybody knows that everything in the universe has a random scheme. For example, we are unaware of the next minute of our life. It is the reason for randomness of all the events. After identifying the optimal control and understanding the importance of random comment in the various sciences, the stochastic optimal control has been interested in [3] and [4].

After a while by identifying capital markets and other fields, stochastic optimal control of jump diffusion has been raised. In this issue we can see jumps in the system [5] and [6]. Therefore, the optimal control is developed by understanding the needs of the world.

Nowadays by recognizing the practical courses in science and having partial information in most cases, stochastic optimal control of partially observable systems has become an important topic. A lot of researches have been done on this subject [7], [8] and [9]. In [10], Wang has used maximum principle to find the optimal control in a partially observable control problem. Tang has also found a way to obtain the control factor in a SOCPP.

He has found the relations between adjoint processes and the maximum principle rule to establish this

2010 *Mathematics Subject Classification*. Primary 93Exx
Keywords. Stochastic optimal control; Partially observable systems; Feedback control; Linear filtering; Kalman filter.
Received: 09 September 2016; Accepted: 05 January 2017
Communicated by Miljana Jovanović
Email addresses: Delavarkh@yazd.ac.ir (A. Delavarkhalafi), Ali70.Sherafat@gmail.com (A. Poursherafatan)

method [11]. In [12], Stockbridge has presented a technique to find the optimal control by characterizing linear programming formula. Cimen surveys a method by Kalman-Busy filter to obtain the optimal control for non-linear SOCPP [13]. In [14], a SOCPP is solved explicitly. In this method it does not use Hamilton-Jacobi-Bellman equation or a stochastic maximum principle. There are several researches as the above ones are in the field of partially observable stochastic optimal control (See [15], [16] and [17]).
In this paper, we are extended the study of Bensoussan [18] to a discrete method, where the SOCPP is presented there. In fact, we are modified studies [18] to obtain a discrete method. A SOCPP has a payoff function that should be minimized. It also has two dynamic differential equations. The first one is state process and the second one is observation process. State process is an equation with own variable, but the observation process is an equation of state process. There is this kind of Problems to minimize the cost function by using the information about state and observation processes. They have two parts, the first part finds control factor and the second one obtains a solution for state process which is according to the observation process. After finding the optimal control factor, it should be substituted in the state process to have a partially observable system. Enough information about the proof of control optimality and the uniqueness of linear filtering for evaluating the related partially observable systems are stated in the paper. Finally, it is introduced two methods for estimation. All methods have a discrete form. But the second method is used for non-linear cases in witch obtianing control factors is difficult to estimate or the analytical solutions of control parameters do not exist.

2. Preliminaries

In this section all needed assumptions are introduced. So assume that (Ω, A, P) is a probability space and all processes and variables are chosen in it. A SOCPP in general form is described as follows.
The state process x , with the control factor v has the bellow system,

$$dx = (Fx + f + Gv)dt + \sigma_0 dw, \quad x(0) = \xi, \tag{1}$$

and the observation dynamic z is as follows,

$$dz = (Hx + h)dt + \sigma_1 db, \quad z(0) = 0, \tag{2}$$

where ξ , is a vector of random variable with Gaussian probability law, with mean value x_0 and covariance matrix P_0 . The processes $w(\cdot)$ and $b(\cdot)$ are Wiener processes with the corresponding noise covariance matrices $Q(\cdot)$ and $R(\cdot)$ respectively. The matrix $R(\cdot)$ is uniformly positive definite. It should be mentioned that the variable ξ and processes $w(\cdot), b(\cdot)$ are mutually independent [18].
The matrix functions $F, f, G, v, \sigma_0, z, H, h, \sigma_1, w(\cdot), b(\cdot), \xi$ are $n \times n, n \times 1, n \times n, n \times 1, n \times d, k \times 1, k \times n, k \times 1, k \times k, d \times 1, k \times 1, n \times 1$ dimensional respectively.
At time t the controller knows the observations $z(r)$ for $0 \leq r \leq t$, where 0 is the start time. In fact, all the researches are in the space of observable information. The corresponding noise covariance matrices of processes $w(\cdot)$ and $b(\cdot)$ in equations (1) and (2) are defined by $Q = (\sigma_0 \times \sigma_0{}^*)_{n\times n}$ and $R = (\sigma_1 \times \sigma_1{}^*)_{k\times k}$ respectively [2].
Let introduce the stochastic optimal control problem with partial observation systems. The system is described by the dynamics (1) and (2) where the control $v(\cdot) \in L^2(0, T; R^n)$, $x(\cdot) \in L^2(0, T; R^n)$ and $z(\cdot) \in L^2(0, T; R^k)$. There is a minimum payoff function in all control problems. In this paper the suitable payoff function in general form is as follows,

$$J(v(\cdot)) = E\Big\{ \int_0^T (x^* Mx + v^* Nv + 2mx + 2nv)dt + x^*(T)M_T x(T) + 2m_T x(T)\Big\}. \tag{3}$$

where $M(\cdot)$ is symmetric non negative, $N(\cdot)$ is symmetric uniformly positive definite and all functions of time F, G, f, M, N, m, n are bounded [2] and [18].

2.1. Framework

In this subsection the dynamic systems (1) and (2) are seperated in to two systems. The first one is the 'without control' system and the other one is 'difference' system. These processes help us to evaluate and explore the optimal control factor.
Consider the state process (1) and observable dynamic (2). Define the without control processes $\alpha(\cdot)$ and $\beta(\cdot)$ from processes (1) and (2) as follows [18].

$$d\alpha = [F(t)\alpha + f(t)]dt + \sigma_0 dw, \quad \alpha(0) = \xi. \tag{4}$$

$$d\beta = [H(t)\alpha + h(t)]dt + \sigma_1 db, \quad \beta(0) = 0. \tag{5}$$

The processes $\alpha(\cdot)$ and $\beta(\cdot)$ are corresponding to the dynamics (1) and (2) respectively.
Assume that $\nu(t) \in L^2(0, T\times\Omega, dt\otimes dP; R^k)$ is a square integrable process. Then define the Difference processes x_1, z_1 as follows,

$$dx_1 = d(x - \alpha) = (F(x - \alpha) + G\nu)dt = (Fx_1 + G\nu)dt, \quad x(0) - \alpha(0) = x_1(0) = 0, \tag{6}$$

$$dz_1 = d(z - \beta) = H(x - \alpha)dt = Hx_1 dt, \quad z(0) - \beta(0) = z_1(0) = 0. \tag{7}$$

It is clear that the above relations are solved for each sample path.
The processes $x(\cdot)$ and $z(\cdot)$ in separation mode are defined as follows (for any control $\nu(\cdot)$).

$$x(t) = \alpha(t) + x_1(t). \tag{8}$$

$$z(t) = \beta(t) + z_1(t). \tag{9}$$

The above processes (8) and (9) are the state and observation corresponding to the control $\nu(\cdot)$.
First, consider the family of σ - algebras Z^t_ν , consists of all the information (until time t) from observation process $z(\cdot)$.

$$Z^t_\nu = \sigma(z(s), s \le t). \tag{10}$$

The optimal control is selected in the space of admissible controls. The definitions of admissibility are in the following:

Definition 2.1. *(natural definition): Control $\nu(\cdot)$ is admissible if and only if $\nu(t)$ is adapted to Z^t_ν .*

After that, introduce the family of σ - algebras F^t , consists of all the information (until time t) from without control observation process $\beta(\cdot)$.

$$F^t = \sigma(\beta(s), s \le t). \tag{11}$$

Definition 2.2. *: If control $\nu(\cdot)$ is admissible then $F^t \subset Z^t_\nu$.*

Definition 2.3. *(more restrictive definition): Control $\nu(\cdot)$ is admissible if and only if $\nu(t)$ is adapted to F^t and to Z^t_ν .*

All of the above definitions are the result of (8) and (9) expressions. If $\nu(\cdot)$ is admissible, then $F^t = Z^t_\nu$, it is one of the important conclusion of expression (9). This conclusion makes it easy to obtain the Kalman filter $\hat{x}$ as follows [19].

$$\hat{x}(t) = E[x(t)|Z^t_\nu] = E[x(t)|F^t]. \tag{12}$$

This paper studies the control on the filter to obtain the Kalman filter process. So, consider the 'innovation process' $I(\cdot)$ as follows,

$$I(t) = \beta(t) - \int_0^T (H\hat{\alpha} + h)ds = z(t) - \int_0^T (H\hat{x} + h)ds. \tag{13}$$

In equation (13) the first equality is obtained from the expression (5) and the second one is obtained from the expression (2) as follows,

$$\beta(t) - \underbrace{\beta(0)}_{0} = \int_0^t [H(t)\alpha + h(t)]dt + \int_0^t \sigma_1 db \xrightarrow{(12)} I(t) = \int_0^t \sigma_1 db = \beta(t) - \int_0^t [H(t)\hat{\alpha} + h(t)]dt.$$

$$Z(t) - \underbrace{Z(0)}_{0} = \int_0^t [H(t)x + h(t)]dt + \int_0^t \sigma_1 db \xrightarrow{(12)} I(t) = \int_0^t \sigma_1 db = Z(t) - \int_0^t [H(t)\hat{x} + h(t)]dt.$$

It is clear that $I(t)$ is a F^t wiener process with covariance matrix $R(\cdot)$. The control factor is defined with two parameters Λ and λ in the linear feedback form as follows,

$$v(t) = \Lambda(t)x + \lambda(t). \tag{14}$$

Impose this feedback control on process (1).

$$dx = [(F + G\Lambda)x + f + G\lambda]dt + \sigma_0 dw. \tag{15}$$

Expression (15) is not the real solution. In fact the real solution is a solution of state process which is according to the observation z. But before obtaining the real solution we should find optimal control. The following theorem is about the admissibility of control.

Theorem 2.4. *If $v(\cdot)$ is admissible in the sense of definition* (2.3)*, then the Kalman filter $\hat{x}(t)$ is the solution of,*

$$d\hat{x} = (F\hat{x} + f + Gv)dt + PH^*R^{-1}(dz - (H\hat{x} + h)dt), \quad \hat{x}(0) = 0, \tag{16}$$

where P is the solution of bellow differential equation with initial point $P(0)$.

$$\begin{aligned} &-\frac{dP}{dt} + FP + PF^* - PH^*R^{-1}HP + Q = 0 \\ &P(0) = E[(x_0 - E[x_0])(x_0 - E[x_0])^*]. \end{aligned} \tag{17}$$

Proof. please see the theorem and proof in [7], [9] and [18]. □

By this theorem the real equation of solution with the 'innovation process' is as follows,

$$d\hat{x} = [(F + G\Lambda)\hat{x} + f + G\lambda]dt + PH^*R^{-1}dI, \quad \hat{x}(0) = x_0. \tag{18}$$

The solution of expression (18) is the Kalman filter of x and is corresponding to the observation z; But what functions should be used instead of $\Lambda(\cdot)$ and $\lambda(\cdot)$ parameters in control factors. This issue is expressed in the next subsection. The following theorem is a complete proof of optimality and uniqueness of feedback control.

Theorem 2.5. *(optimality of control): For a SOCPP with a payoff function* (3)*, state process* (1) *and observable dynamic* (2) *there is an optimal and unique feedback control in the form of* (14)*, with the following control parameters, among the set of admissible control in the sense of definition* (2.3)*.*

$$\begin{aligned} &\Lambda(t) = -N^{-1}(t)G^*(t)\Pi(t). \\ &\lambda(t) = -N^{-1}(t)n^*(t) - N^{-1}(t)G^*(t)r(t). \end{aligned}$$

Where Π and r are the solutions of the following differential equations.

$$\begin{aligned} &\frac{d\Pi}{dt} + \Pi F + F^*\Pi - \Pi G N^{-1} G^* \Pi + M = 0, \quad \Pi(T) = M_T. \\ &\frac{dr}{dt} + (F^* - \Pi G N^{-1} G^*)r + \Pi(f - GN^{-1}n^*) + m^* = 0, \quad r(T) = m_T. \end{aligned}$$

Proof. Beforehand, state the control factor form in expression (14). Assume that the control parameters Λ and λ are unspecified in this form. Denote the control factor (14) by $u(\cdot)$ instead of $\nu(\cdot)$. The corresponding state and observation processes are denoted by $y(\cdot)$ and $\zeta(\cdot)$ respectively.

$$dy = (Fy + f + Gu)dt + \sigma_0 dw, \quad y(0) = \xi, \tag{19}$$

$$d\zeta = (Hy + h)dt + \sigma_1 db, \quad \zeta(0) = 0. \tag{20}$$

The Kalman filter of state process corresponding to the observation process $y(\cdot)$ and control factor $u(t) = \Lambda(t)\hat{y}(t) + \lambda(t)$ is as follows.

$$d\hat{y} = [(F + G\Lambda)\hat{y} + f + G\lambda]dt + PH^* R^{-1}(d\zeta - (H\hat{y} + h)dt), \quad \hat{y}(0) = x_0, \tag{21}$$

Make a 'change of control function' to prove the optimality of this case. Suppose that process $\mu(\cdot)$ is adapted to F^t . Now if the process $\eta(\cdot)$ is adapted to F^t and it has the bellow differential equation, then make sure that the 'change of control function' $\nu(t)$ is also adapted to F^t.

$$d\eta = [(F + G\Lambda)\eta + f + G\lambda + G\mu]dt + PH^* R^{-1} dI, \quad \eta(0) = x_0, \tag{22}$$

where the 'innovation process' is as follows,

$$dI = dz - (H\eta + h)dt, \tag{23}$$

and the change of control function is,

$$\nu(t) = \Lambda(t)\eta(t) + \lambda(t) + \mu(t). \tag{24}$$

In fact the state and observation processes which are corresponding to the control factor $\nu(\cdot)$ and $\eta(t)$ is the Kalman filter of process x as expressed in the expression (12).
Furthermore $\eta(t)$ is adapted to F^t. Impose the additional condition that process $\mu(\cdot)$ is adapted to Z^t_ν, then we can understand from the innovation process (23) and equation (22) that $\eta(t)$ is also adapted to Z^t_ν.
So the process $\eta(t)$ is adapted to the both families of σ - algebras F^t and Z^t_ν. Therefore, $\nu(\cdot)$ is admissible in the sense of definition (2.3). The corresponding state (x) and observation (z) processes for any $\mu(\cdot)$ are as follows,

$$dx = (Fx + f + G\nu)dt + \sigma_0 dw, \quad x(0) = \xi, \tag{25}$$

$$dz = [H(t)x + h(t)]dt + \sigma_1 db, \quad z(0) = 0. \tag{26}$$

The process $\hat{x}$ is the Kalman filter corresponding to the state process $x(\cdot)$ and control $\nu(\cdot)$.

$$d\hat{x} = [(F + G\Lambda)\hat{x} + f + G\lambda + G\mu]dt + PH^* R^{-1}(dz - (H\hat{x} + h)dt), \quad \hat{x}(0) = x_0, \tag{27}$$

Find the following useful comment by comparing Kalman filters (21) and (27).

$$dz - (H\hat{x} + h)dt = d\zeta - (H\hat{y} + h)dt = d\beta - (H\hat{\alpha} + h)dt. \tag{28}$$

By subtracting these two Kalman filters make a new process ($\tilde{x}$) as follows [18].

$$\frac{d\tilde{x}}{dt} = (F + G\Lambda)\tilde{x} + G\mu, \quad \tilde{x}(0) = 0, \tag{29}$$

where,

$$\tilde{x}(t) = \hat{x}(t) - \hat{y}(t). \tag{30}$$

Moreover, by using Kalman filter (12) for the 'separated processes' (8) and (9) find the following two new relations for $\hat{x}$ and $\hat{y}$.

$$\hat{x}(t) = \hat{\alpha}(t) + x_1(t),$$
$$\hat{y}(t) = \hat{\alpha}(t) + y_1(t).$$

Now, due to the expressions (8) and (9) and the above relations, find the following equalities,

$$d(x_1 - y_1) = F(x_1 - y_1) + G(\nu - u) = F(x_1 - y_1) + G\Lambda(x_1 - y_1) + G\mu, \ \ x_1(0) - y_1(0) = 0.$$

From the above equation conclude that,

$$\hat{x}(t) - \hat{y}(t) = x_1(t) - y_1(t) = x(t) - y(t) = \tilde{x}(t). \tag{31}$$

To compute the value of cost functions, use expression (30) and substitute $\hat{x}(t) = \tilde{x}(t) + \hat{y}(t)$, in $J(\nu(\cdot))$ relation and simplified it as follows,

$$\begin{aligned} J(\nu(\cdot)) = K(\mu(\cdot)) = E\Big\{ \int_0^T & [(y(t) + \tilde{x}(t))^* M(y(t) + \tilde{x}(t)) \\ & + (\Lambda\hat{y} + \lambda + \Lambda\hat{x} + \mu)^* N(\Lambda\hat{y} + \lambda + \Lambda\hat{x} + \mu) \\ & +2m(y + \tilde{x}) + 2n(\Lambda\hat{y} + \lambda + \Lambda\hat{x} + \mu)]dt \\ & + (y(T) + \tilde{x}(T))^* M_T(y(T) + \tilde{x}(T)) \\ & + 2m_T(y(T) + \tilde{x}(T))\Big\}. \end{aligned}$$

By simplifying the above relation the corresponding payoff function is,

$$J(\nu(\cdot)) = J(u(\cdot)) + E\Big\{ \int_0^T [\tilde{x}^* M\tilde{x} + (\Lambda\tilde{x} + \mu)^* N(\Lambda\tilde{x} + \mu)]dt + \tilde{x}(T)^* M_T \tilde{x}(T)\Big\} + 2X,$$

where,

$$X = E\Big\{ \int [\tilde{x}^*(t) M y(t) + (\Lambda\tilde{x} + \mu)^* N(\Lambda\hat{y} + \lambda) + m\tilde{x} + n(\Lambda\tilde{x} + \mu)]dt + \tilde{x}^*(T) M_T y(T) + m_T \tilde{x}(T)\Big\}.$$

From equality (31) and the above relations find the following relation,

$$E\tilde{x}^*(t) M y(t) = E\Big\{\tilde{x}^*(t) M E[y(t)|F^T]\Big\} = E\tilde{x}^*(t) M \hat{y}(t). \tag{32}$$

Define,

$$p(t) = \Pi(t) y(t) + r(t), \tag{33}$$

and use it for expression X,

$$X = \int_0^T (\tilde{x}^* M y + (\Lambda\tilde{x} + \mu)^* N(\Lambda\hat{y} + \lambda) + m\tilde{x} + n(\Lambda\tilde{x} + \mu))dt + \tilde{x}^*(T) p(T). \tag{34}$$

then the differential form of $\tilde{x}(t)p(t)$ is produced as follows,

$$\frac{d}{dx}\tilde{x}^*(t)p(t) = \tilde{x}^*(F^* + \Lambda^* G^*)p + \mu^* G^* p + \tilde{x}^*\Big[\frac{d\Pi}{dt} y + \Pi(F + G\Lambda)y + \Pi(F + G\lambda) + \frac{dr}{dt}\Big].$$

So by substituting the above relation in (34) find the below expression,

$$\begin{aligned} X = \int_0^T & \Big\{\tilde{x}^* M y + (\Lambda\tilde{x} + \mu)^* N(\Lambda\hat{y} + \lambda) + m\tilde{x} + n(\Lambda\tilde{x} + \mu) \\ & + \tilde{x}^*(F^* + \Lambda^* G^*)p + \mu^* G^* p \\ & + \tilde{x}^*\Big[\frac{d\Pi}{dt} y + \Pi(F + G\Lambda)y + \Pi(F + G\lambda) + \frac{dr}{dt}\Big]\Big\}dt. \end{aligned}$$

In the above relation, set the coefficients of $\tilde{x}^*$ and μ^* to zero. This yields the bellow relations,

$$My + \Lambda^* N(\Lambda y + \lambda) + m^* + \Lambda^* n^* + (F^* + \Lambda^* G^*)(\Pi y + r) + \frac{d\Pi}{dt} y + \Pi(f + G\Lambda)y + \Pi(f + G\lambda) + \frac{dr}{dt} = 0, \quad (35)$$

$$N(\Lambda y + \lambda) + n^* + G^*(\Pi y + r) = 0. \quad (36)$$

Match the coefficients of y and constant terms of equalities (35) and (36), then get the following relations,

$$M + \Lambda^* N\Lambda + (F^* + \Lambda^* G^*)\Pi + \frac{d\Pi}{dt} + \Pi(F + G\Lambda) = 0, \quad (37)$$

$$\Lambda^* N\lambda + m^* + \Lambda^* n^* + (F^* + \Lambda^* G^*)r + \Pi(f + G\lambda) + \frac{dr}{dt} = 0, \quad (38)$$

$$N\Lambda + G^*\Pi = 0, \quad (39)$$

$$N\lambda + n^* + G^* r = 0. \quad (40)$$

By solving equations (39) and (40) respect to the parameters Λ and λ, we can achieve the following results,

$$\begin{aligned} &\Lambda(t) = -N^{-1}(t)G^*(t)\Pi(t), \\ &\lambda(t) = -N^{-1}(t)n^*(t) - N^{-1}(t)G^*(t)r(t). \end{aligned} \quad (41)$$

Now by substituting equalities (41) in expressions (37) and (38) the following differential equations are obtained,

$$\frac{d\Pi}{dt} + \Pi F + F^*\Pi - \Pi G N^{-1} G^* \Pi + M = 0, \quad \Pi(T) = M_T. \quad (42)$$

$$\frac{dr}{dt} + (F^* - \Pi G N^{-1} G^*)r + \Pi(f - GN^{-1}n^*) + m^* = 0. \quad r(T) = m_T. \quad (43)$$

In fact, by solving equations (42) and (43), find the parameters of equality (41) for optimal feedback control (14). Therefore, find that $X = 0$ and,

$$J(v(\cdot)) = K(\mu(\cdot)) = J(u(\cdot)) + E\Big\{ \int_0^T [\tilde{x}^* M \tilde{x} + (\Lambda\tilde{x} + \mu)^* N(\Lambda\tilde{x} + \mu)]dt + \tilde{x}(T)^* M_T \tilde{x}(T) \Big\}. \quad (44)$$

It is clear that the optimal process $\mu(\cdot)$ is zero for minimizing the cost function $J(v(\cdot))$. Therefore, it is possible to know that the change of control function is equal to the feedback control. It means that the control feedback is optimal and unique. □

From the above results, deduce that minimizing the cost function (3) which is according to the state and observation processes (1) and (2) has a unique solution. Its control factor is equality (14) with control parameters (41).
After finding optimal feedback control it should be substituted in the state process. Now there are two dynamic systems so that the second one is according to the first. In the next section a method will be introduced to simulate this partially observation process.

2.2. Partially observation systems

The continuous partially observation systems are described as follows,

$$dx = [F(t)x + f(t)]dt + \delta dw, \quad x(0) = \xi, \quad (45)$$

$$dy = [H(t)x(t) + h(t)]dt + \gamma db(t), \quad y(0) = 0, \quad (46)$$

where ξ, is a vector random variable with Gaussian probability law, with mean value x_0 and covariance matrix P_0. The processes $w(\cdot)$ and $b(\cdot)$ are wiener processes with the corresponding noise covariance matrices

$Q(\cdot)$ and $R(\cdot)$ respectively. The matrix $R(\cdot)$ is uniformly positive definite. It should be mentioned that the variable ξ and processes $w(\cdot)$ and $b(\cdot)$ are mutually independent. The matrix functions $\delta, z, H, h, \gamma, w(\cdot), b(\cdot), \xi$ are $n \times d, k \times 1, k \times n, k \times 1, k \times k, d \times 1, k \times 1, n \times 1$ dimensional respectively. At time t the controller from where observations $y(r)$ for $0 \leq r \leq t$ where 0 is the start time. The corresponding noise covariance matrices of processes $w(\cdot)$ and $b(\cdot)$ in equations (45) and (46) are defined by $Q = (\delta \times \delta^*)_{n\times n}$ and $R = (\gamma \times \gamma^*)_{k\times k}$ respectively [2].
For simulating this kind of system, convert it to the discrete time mode. Consider these discrete systems as follows,

$$x_{k+1} = F_k x_k + f_k + \delta_k w_k, \quad k = 0, 1, \cdots, N-1, \quad x_0 = \xi, \tag{47}$$

$$y_k = H_k x_k + h_k + \gamma_k b_k, \quad k = 0, 1, \cdots, N-1. \tag{48}$$

where $x_k \in \mathbb{R}^n$ and $y_k \in \mathbb{R}^k$ and the matrix functions $F_k, f_k, \delta_k, w_k, \xi, H_k, h_k, \gamma_k$ and b_k are $n \times n, n \times 1, n \times d, d \times 1, n \times 1, k \times n, k \times 1, k \times k$ and $k \times 1$ dimensional respectively. The variables w_k and b_k are Gaussian with mean value 0 and covariance matrices Q_k and R_k respectively and the matrix R_k is positive definite. The variables ξ, w_k and b_k are mutually independent [23].
Let us consider the sequence of σ - algebras,

$$y^k = \sigma(y_0, \ldots, y_{k-1}), \quad k = 1, \ldots, N. \tag{49}$$

The solution of partially observable systems is to compute following relation [24] and [25].

$$\hat{x}_N = E[\ x_N \mid y^N\ ].$$

In fact the solution of discrete systems (47) and (48) is according to the observation process until time N. In this way, use the linear filters [21]. A best linear filter is defined as follows,

$$F_s = \bar{x}_N + \sum_{k=0}^{N-1} S_k(y_k - \bar{y}_k), \tag{50}$$

where $\left\{S_0, \ldots, S_{N-1}\right\}$ is a set of matrices which characterized by S in the text and $\bar{x}_N$ and $\bar{y}_k$ represent the means of x_N and y_k respectively. The sequences $\bar{x}_k$ and $\bar{y}_k$ are defined as follows [20].

$$\bar{x}_{k+1} = F_k \bar{x}_k + f_k, \quad k = 0, \ldots, N-1, \tag{51}$$

$$\bar{y}_k = H_k \bar{x}_k + h_k. \tag{52}$$

Choose S in order to minimize the bellow function. It is a rule to obtain the best linear filter [18].

$$L(S) = E(x_N - F_s)^*(x_N - F_s). \tag{53}$$

Let Λ_{kl} be the correlation matrix of the process x_k, which is in $L(\mathbb{R}^n; \mathbb{R}^n)$.

$$\Lambda_{kl} = E(x_k - \bar{x}_k)(x_l - \bar{x}_l)^*. \tag{54}$$

It should be expressed that, $\Lambda^*_{kl} = \Lambda_{lk}$ and if $M \in L(\mathbb{R}^n; \mathbb{R}^n)$, the following relation is held,

$$E(x_k - \bar{x}_k)^* M(x_l - \bar{x}_l) = tr\Lambda_{lk} M. \tag{55}$$

Then by using the expression (53), (54) and (55) deduce that,

$$L(S) = E\left((x_N - \bar{x}_N)(x_N - \bar{x}_N)^* + \left(\sum_{k=0}^{N-1} S_k(y_k - \bar{y}_k) \right)^2 - 2\left(\sum_{k=0}^{N-1} S_k(y_k - \bar{y}_k) \right)(x_N - \bar{x}_N)^* \right),$$

where,

$$E\left(\sum_{k=0}^{N-1} S_k(y_k - \bar{y}_k)\right)^2 = E\left(\sum_{k,l=0}^{N-1} S_l(y_l - \bar{y}_l)S_k^*(y_k - \bar{y}_k)^*\right) = \sum_{k,l=0}^{N-1} tr\Lambda_{lk}H_k^*S_k^*S_lH_l + \sum_{k=0}^{N-1} trR_kS_k^*S_k,$$

and,

$$E\left(\left(\sum_{k=0}^{N-1} S_k(y_k - \bar{y}_k)\right)(x_N - \bar{x}_N)^*\right) = \sum_{k=0}^{N-1} E\Big(S_k(y_k - \bar{y}_k)(x_N - \bar{x}_N)^*\Big) = \sum_{k=0}^{N-1} tr\Lambda_{kN}S_kH_k.$$

Therefore,

$$L(S) = tr\Lambda_{NN} + \sum_{k=0}^{N-1} trR_kS_k^*S_k + \sum_{k,l=0}^{N-1} tr\Lambda_{lk}H_k^*S_k^*S_lH_l - 2\sum_{k=0}^{N-1} tr\Lambda_{kN}S_kH_k. \tag{56}$$

It is important to find that, there is a unique S minimizing the functional $L(S)$.

Theorem 2.6. *There exist a unique S minimizing the functional $L(S)$.*

Proof. Suppose that the set of S is a finite dimensional vector space with the scalar product $(S, \tilde{S})$ that states as follows,

$$(S, \tilde{S}) = \sum_k trS_k^*\tilde{S}_k. \tag{57}$$

Since the functional $L(S) \geq 0$ and the equation (56) are quadratic then it can be concluded that $tr\Lambda_{NN} \geq 0$ and $\sum_{k,l} tr\Lambda_{lk}H_k^*S_k^*S_lH_l \geq 0$. Furthermore, according to the following expressions, find $\sum_{k=0}^{N-1} trR_kS_k^*S_k \geq 0$. (See [18].)

$$\sum_{k=0}^{N-1} trR_kS_k^*S_k = \sum_{K=0}^{N-1}\sum_{h=1}^{n}\sum_{i,j=1}^{m} R_{k,ij}S_{k,hi}S_{k,hj} \geq \alpha\sum_{K=0}^{N-1}\sum_{h=1}^{n}\sum_{i=1}^{m} (S_{k,hi})^2 = \alpha||S||^2 \geq 0. \tag{58}$$

As mentioned previously, know that $L(S) \geq 0$ and it should be minimized.
Furthermore, it can be found that $tr\Lambda_{NN} \geq 0$, $\sum_{k,l} tr\Lambda_{lk}H_k^*S_k^*S_lH_l \geq 0$ and $\sum_{k=0}^{N-1} trR_kS_k^*S_k \geq 0$. By these concepts find that sum of the expressions with an unknown variable in (56) is zero.
Therefore, there is a unique minimum S which is applied in the following expression,

$$\sum_{k=0}^{N-1} trR_kS_k^*S_k + \sum_{k,l=0}^{N-1} tr\Lambda_{lk}H_k^*S_k^*S_lH_l - 2\sum_{K=0}^{N-1} tr\Lambda_{kN}S_kH_k = 0, \quad \forall S. \tag{59}$$

□

Now the unique linear filter (50) is existed, but the set S should be found as follows.

$$\frac{d}{dS_i}L(S_0, S_1, \ldots, S_{N-1}) = 0, \quad \forall i = 0, 1, \ldots, N-1. \tag{60}$$

The above expression is a convenient rule in multiple cases to obtain the minimum or maximum amount of the function [22]. As stated obviously, $L(S)$ is quadratic, so make sure that by using the relation (60) the minimum of $L(S)$ is obtained.

3. Estimation methods

In this section we introduce two methods for a general stochastic optimal problem with partial observation in filtering case. Suppose the following stochastic optimization problem,

$$\min_{v\in U} J(v(\cdot)) = \int_0^T (x^*Mx + v^*Nv + 2mx + 2nv)dt + x^*(T)M_T x(T) + 2m_T x(T).$$
$$dx = (Fx + f + Gv)dt + \sigma_0 dw, \quad x(0) = \xi.$$
$$dz = (Hx + h)dt + \sigma_1 db, \quad z(0) = 0.$$

Where, U is the space of admissible controls. First, find the feedback control $u(x,t) = \Lambda(t)x(t) + \lambda(t)$ with the following parameters,

$$\Lambda(t) = -N^{-1}(t)G^*(t)\Pi(t),$$
$$\lambda(t) = -N^{-1}(t)n^*(t) - N^{-1}(t)G^*(t)r(t).$$

To do this, $\Pi(t)$ and $r(t)$ must be obtained. They are the solutions of the following differential equations,

$$\frac{d\Pi}{dt} + \Pi F + F^*\Pi - \Pi G N^{-1} G^* \Pi + M = 0, \quad \Pi(T) = M_T.$$
$$\frac{dr}{dt} + (F^* - \Pi G N^{-1} G^*)r + \Pi(f - GN^{-1}n^*) + m^* = 0, \quad r(T) = m_T.$$

Now the feedback control can be constructed. Substitute the feedback control in the state and observation processes and obtain the following dynamic system,

$$dx = (Fx + f + G(-N^{-1}G^*\Pi x - N^{-1}n^* - N^{-1}G^*r))dt + \delta dw, \quad x(0) = \xi. \tag{61}$$

$$dz = (Hx + h)dt + \gamma db, \quad z(0) = 0. \tag{62}$$

3.1. linear filter method

At first, we should discrete the above system for $k = 0, 1, \ldots, N-1$ then substitute $\Pi(t_k)$ and $r(t_k)$ in the discrete form of (61) and (62) as follows,

$$x_{k+1} = F_k x_k + f_k + G_k(-N_k^{-1}G_k^*\Pi(t)x_k - N_k^{-1}n_k^* - N_k^{-1}G_k^* r(t)) + \delta_k w_k, \quad x_0 = \xi. \tag{63}$$

$$y_k = H_k x_k + h_k + \gamma_k b_k. \tag{64}$$

As it was stated in the previous sections, the needed solution is $\hat{x}_N = \bar{x}_N + \sum_{k=0}^{N-1} S_k(y_k - \bar{y}_k)$, where $\bar{x}_k$ and $\bar{y}_k$ are defined in equations (51) and (52). All arguments in the solution $\hat{x}_N$ are known expect S_k for $k = 0, 1, \ldots, N-1$. Chose dynamic system as (63) and (64) and obtain $S_k(t)$ for $k = 0, 1, \ldots, N-1$. In fact, in this method achieve $\hat{x}_N$ as a function of t in each sample path. Of course, the solution is a curve in each sample path.

3.2. Adams-Moulton method for non-linear problem

The second method is due to the difficulty of solutions $\Pi(t)$ and $r(t)$ in 1-dimensional cases. Sometimes the exact solutions of these differential equations are too long and some computational problems are occurred in estimations. Therefore, using one step Adams-Moulton method helps us in estimations. This 1-step method for solving $y' = f(x,y)$ is as follows.

$$y_{n+1} - y_n = \frac{h}{2}(f_{n+1} + f_n). \tag{65}$$

There is a problem in the dynamics of $\Pi(t)$ and $r(t)$. In fact, the known value of these differential equations are $\Pi(T) = M_T$ and $r(T) = m_T^*$. It means that they are not initial value. Therefore construct a backward method according to the relation (65) for the differential equations $\Pi(t)$ and $r(t)$ as follows. We use the following form of these dynamics to obtain the backward methods,

$$\Pi' = -\Pi F - F^*\Pi + \Pi G N^{-1} G^* \Pi - M,$$
$$r' = -(F^* - \Pi G N^{-1} G^*) r - \Pi(f - G N^{-1} n^*) - m^*.$$

Now, apply the Adams- Moulton method on the above equations and obtain Π_n and r_n as follow.

$$\Pi_n = \sqrt{\frac{\Pi_{n+1} + N^{-1}G^2\frac{h}{2}{\Pi_{n+1}}^2 + hF\Pi_{n+1} + \frac{2M}{N^{-1}G^2} + (\frac{1+hF}{N^{-1}G^2h})^2}{-N^{-1}G^2\frac{h}{2}}} + \frac{1+hF}{N^{-1}G^2h}, \tag{66}$$

$$\Pi_N = M_T. \tag{67}$$

$$r_n = \frac{r_{n+1} - \frac{h}{2}(N^{-1}G^2\Pi_{n+1} - F)r_{n+1} - \frac{h}{2}\Pi_{n+1}(GN^{-1}h - f) - \frac{h}{2}\Pi_n(GN^{-1}h - f) - hm}{1 + \frac{h}{2}(N^{-1}G^2\Pi_n - F)}, \tag{68}$$

$$r_N = m_T^*. \tag{69}$$

By using the above relations, find the numerical solutions of $\Pi(t)$ and $r(t)$ in the backward mode. Then substitute the numerical results in (63) and (64) relations for $k = 0, 1, \ldots, N-1$.

$$x_{k+1} = F_k x_k + f_k + G_k(-N_k^{-1}G_k^*\Pi_k x_k - N_k^{-1}n_k^* - N_k^{-1}G_k^* r_k) + \delta_k w_k, \quad x_0 = \xi. \tag{70}$$

$$y_k = H_k x_k + h_k + \gamma_k b_k. \tag{71}$$

After obtaining the processes $x_k, y_k, \bar{x}_k$ and $\bar{y}_k$ from the relations (70), (71), (51) and (52) simulate the filtering process $\hat{x}_N$ for $n = 0, 1, \ldots, N-1$.

$$\hat{x}_n = \bar{x}_n + \sum_{k=0}^{n-1} S_k(y_k - \bar{y}_k). \tag{72}$$

This method simulates the filtering process $\hat{x}_n$ point wise. In fact, the graph of this process turned around the exact solution of the partial observation system. Deduce that the expected exact solution of the system is the curve fitting of some paths witch is obtained by the second method.

In fact, these two methods are the same. In these methods we use the linear filters to estimate the Kalman filter of the corresponding process. The only difference between the two methods is substituting the solutions of differential equations (42) and (43) in the relation (63). In the first method we use the regular solution of (42) and (43) and in the second one we use the numerical method of them. We use the numerical method, because, in some non-linear cases the regular solution of differential equations (42) and (43) dose not exist.

4. Numerical Example

In this section we take two examples to evaluate methods. Example (4.1) solves with both methods. In this example, we will conclude that the methods are equivalent. But in the example (4.2), the control parameters are not solvable, because, it can be solved only by the second method.

Example 4.1. *Consider the following example and obtain $\hat{x}_n$.*

$$min_{v \in U} J(v(\cdot)) = min_{v \in U} E\left(\int_0^1 (\frac{v^2}{2})dt + 2x(1) \right),$$
$$dx = (x + v + 1)dt + dw, \quad x(0) = 0.$$
$$dz = xdt + db, \quad z(0) = 0.$$

The differential equations of Π and r are solvable. So, this example can be solved by the both methods. The approximations of $\hat{x}$ with the linear filter method and Adamz-Moulton method are as follows,

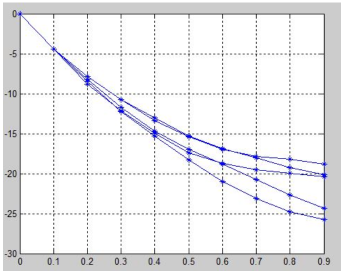

Figure1. Five paths of linear filtering method in ten points.

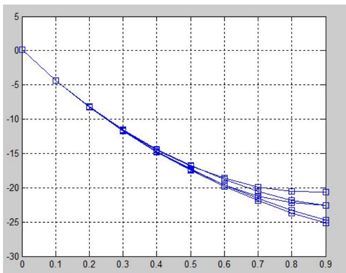

Figure2. Five paths of Adamz- Moulton method in ten points.

By comparing these two methods in the bellow figure, conclude that they are equivalent.

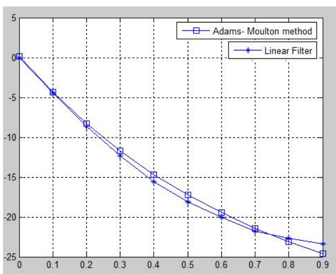

Figure3. Comparison of Two Methods.

Example 4.2. *Consider the bellow example. In this example the differential equations of control parameters don't have analytical solutions. So, this example can be solved only by the second method.*

$$min_{\nu \in U} J(\nu(\cdot)) = min_{\nu \in U} E\left(\int_0^1 (2x^2 + \nu^2)dt + 2x(1) \right),$$
$$dx = \Big(-sin(\pi t + 1)x + \sqrt{t+1}\nu + 1 \Big)dt + dw, \quad x(0) = 0.1.$$
$$dz = (x+1)dt + db, \quad z(0) = 0.$$

According to the state process, we can understand that the exact solution has a sinusoidal form. So, we should see this sinusoidal form in the estimation.

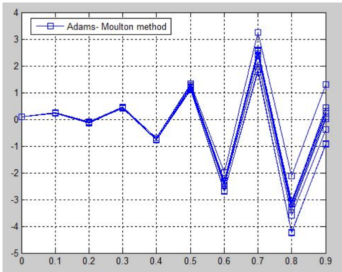

Figure4. Five paths of second method in ten points.

5. Conclusion

In this paper we studied the general form of SOCPPs. First the linear filtering method is introduced to estimate this kind of problem. But this method can be used when the dynamics of control parameters have analytical solutions. The second method is used in non-linear cases when the analytical solutions do not exist. According to the first example we can conclude that, these methods are equivalent.

References

[1] Li, W., Todorov, E. (2007). Iterative linearization methods for approximately optimal control and estimation of non-linear stochastic system. International Journal of Control, 80 (9), 1439-1453.

[2] Fleming, W. H., Rishel, R. W. (2012). Deterministic and stochastic optimal control (Vol. 1). Springer Science & Business Media.

[3] Bertsekas, D. P., Shreve, S. E. (1978). Stochastic optimal control: The discrete time case (Vol. 23, pp. 15-24). New York: Academic Press.

[4] Wang, T., Zhang, H., Luo, Y. (2016). Infinite-time stochastic linear quadratic optimal control for unknown discrete- time systems using adaptive dynamic programming approach. Neurocomputing, 171, 379-386.

[5] ksendal, B. K., Sulem, A. (2005). Applied stochastic control of jump diffusions (Vol. 498). Berlin: Springer.

[6] Confortola, F., Fuhrman, M. (2014). Backward stochastic differential equations associated to jump Markov processes and applications. Stochastic Processes and their Applications, 124(1), 289-316.

[7] Fleming, W. H., Soner, H. M. (2006). Controlled Markov processes and viscosity solutions (Vol. 25). Springer Science & Business Media.

[8] Bouchard, B., Touzi, N. (2011). Weak dynamic programming principle for viscosity solutions. SIAM Journal on Control and Optimization, 49(3), 948-962.

[9] ksendal, B. (2003). Stochastic differential equations. In Stochastic differential equations (pp. 65-84). Springer Berlin Heidelberg.

[10] Wang, G., Xiong, J., Zhang, S. (2016). Partially observable stochastic optimal control. International Journal of Numerical analysis and modeling, 13 (4), 493-512.

[11] Tang, S. (1998). The maximum principle for partially observed optimal control of stochastic differential equations. SIAM Journal on Control and optimization, 36 (5), 1596-1617.

[12] Stockbridge, R. H. (2005). A separation principle for partially observed control of singular stochastic processes. Nonlinear Analysis: Theory, Methods & Applications, 63 (5), e2057-e2065.

[13] imen, T., Banks, S. P. (2005). Stochastic optimal control of partially observable nonlinear systems. IFAC Proceedings Volumes, 38 (1), 687-692.
[14] Duncan, T. E., Pasik-Duncan, B. (2014). Some results on optimal control for a partially observed linear stochastic system with an exponential quadratic cost. IFAC Proceedings Volumes, 47 (3), 8695-8698.
[15] Lesser, K., Oishi, M. (2014). Reachability for partially observable discrete time stochastic hybrid systems. Automatica, 50(8), 1989-1998.
[16] Duncan, T. E., Pasik-Duncan, B. (2014). Some results on optimal control for a partially observed linear stochastic system with an exponential quadratic cost. IFAC Proceedings Volumes, 47(3), 8695-8698.
[17] Ben-Zvi, T., Chernonog, T., Avinadav, T. (2016). A two-state partially observable Markov decision process with three actions. European Journal of Operational Research, 254(3), 957-967.
[18] Bensoussan, A. (2004). Stochastic control of partially observable systems. Cambridge University Press.
[19] Xiong, J. (2008). An introduction to stochastic filtering theory (Vol. 18). Oxford University Press on Demand.
[20] Evans, L. C. (2006). An introduction to stochastic differential equations version 1.2. Lecture Notes, UC Berkeley.
[21] Billingsley, P. (2008). Probability and measure. John Wiley & Sons.
[22] Swokowski, E. W. (1979). Calculus with analytic geometry. Taylor & Francis.
[23] Bierman, G. J. (1973). A comparison of discrete linear filtering algorithms. IEEE Transactions on Aerospace and Electronic Systems, (1), 28-37.
[24] Wang, G., Wu, Z. (2008). KalmanBucy filtering equations of forward and backward stochastic systems and applications to recursive optimal control problems. Journal of Mathematical Analysis and Applications, 342(2), 1280-1296.
[25] Kulikov, G. Y., Kulikova, M. V. (2017). Accurate cubature and extended Kalman filtering methods for estimating continuous-time nonlinear stochastic systems with discrete measurements. Applied Numerical Mathematics, 111, 260-275.