ON STOCHASTIC BELLMAN EQUATION: USING CONDITIONAL CERTAINTY PROPERTY EXPECTATIONS

J. S. Apanapudor* and J. Tsetemi
Department of Mathematics, Delta State University, Abraka, Nigeria
*Corresponding author: japanapudor@yahoo.com

Received: October 25, 2019 **Accepted:** January 29, 2020

Abstract: We review and extend the idea of Linear Quadratic Regulator (LQR) by suppressing the time dependence, assuming that our solution is smooth and the gradient is zero. We develop necessary and sufficient conditions for optimal state feedback solution using the Stochastic Optimality Principle. Furthermore, we obtain the Stochastic Hamilton-Jacobi-Bellman (SHJB) equation, using conditional expectations. Finally, we observed that if no perturbation is present, then the SHJB equation reduces to the deterministic Hamilton-Jacobi-Bellman (HJB) equation.

Keywords: Linear Quadratic Regulator (LQR), optimal feedback, stochastic optimality principle

Introduction

It is assumed within the Linear Quadratic Regulator environment, that the state of the system is available for feedback and that it is uncorrupted by any perturbation signal. Now consider the optimality of the control when the state is perturbed by a random process; with this, the system state becomes a Markov process and the principle of Optimality is employed to derive a stochastic version of the Hamilton – Jacobi – Bellman equation (HJB) equation (Golsten, 1950; Anderson and Moore, 1989; Fleming and Rishel, 1975; Bellman, 1957). The deterministic case of (SHJB) is a special case where dw, the Wienerincrement is replaced by its average value, zero. When a Stochastic problem is solved as a deterministic problem with stochastic variables replaced by their averages (Sontag, 1998), we noticed that the certainty – equivalence principle holds. Certainty equivalence is normally employed in engineering design, the above establishes validity of certainty equivalence (CE) for additive perturbation Stochastic LQR (APSLQR) problems for the deterministic and Stochastic LQR problem, (Perota *et al.*, 1975 and for Stochastic HJB equation (Davis, 1977; Fleming and Rishel, 1975).

The rest of this paper will be organized as follows: the next section will dwell on the Linear Quadratic Regulator; followed by discussion on the Stochastic Bellman equation. Then, we discuss an optimization equation for SHJB control problems. Finally, the last section deals with the additive perturbation, leading to our conclusion.

Linear quadratic regulator (LQR)

The basic LQR problem seeks to state the feedback law of the form

$$u = -kx \qquad (1)$$

which minimizes the performance criteria

$$J(x,u) = x^T(t)Rx(t) + \int_0^\tau \left(x^T Px + u^T Qu\right) dt \qquad (2)$$

subject to the system dynamics given by

$$\dot{x} = Ax(t) + Du(t) \qquad (3)$$

where the matrices P and R are positive – semi- definite and Q is positive definite matrix. X is an n - dimensional input vector. The final time τ is fixed and the final state $x(\tau)$ is free. Considering the LQR problem, the Lagrangian form of it, is

$$L(x,u,\tau) = x^T Px + u^T Qu \qquad (4)$$

The system vector field F is,

$$F(x,u,\tau) = Ax + Du \qquad (5)$$

and the terminal condition

$$R(x(\tau)) = x^T(\tau)Rx(\tau) \qquad (6)$$

The matrices A, D, P, Q and R consist of the input data to the LQR problem and in general the time, τ varies.

Now suppose we suppress the time dependence as follows since this is time invariant problem, we know that

$$\frac{\delta v}{\delta \tau} = 0 \qquad (7)$$

sothat the HJB equation takes the form

$$\frac{\delta v}{\delta \tau} = 0 = \min_u \left\{ L(x,u,\tau) + \frac{\delta v}{\delta x} F^T(x,u,\tau) \right\}$$

$$= \min_u \left\{ x^T Px + u^T Qu + \left[\frac{\delta v}{\delta x}\right]^T (Ax + Du) \right\} \qquad (8)$$

Since our solution is assumed smooth, the minimization will be attained by setting the gradient to zero. If Q is positive definite, the necessary condition for a minimal point is also a sufficient condition, as long as the gradient is zero and solving for u produces

$$u^* = -\frac{1}{2} Q^{-1} D^T \frac{\delta w^*}{\delta x} \qquad (9)$$

Bear in mind that the gradient of a quadratic form

$$u^T Hu + d^T u \qquad (10)$$

with respect to u is equal to $2Hu + d$ (11)

Better still, we state that $u^T Hu + d^T u = 2Hu + d$.

This implies that equations (10) and (11) are equal. It is a fact that an integral quadratic form evaluated for a linear system is a quadratic form in the initial state of the system, hence it is reasonable to assume that

$$W^*(t,x) = x^T V(t) x \qquad (12)$$

where $V(t)$ is a symmetric matrix valued function of time τ.

The gradient of W^* is then $2V(t)x$. If W takes on this quadratic form and its gradient is substituted back into (8), the HJB equation, becomes (13)

$$-x^T \dot{V} x = x^T \left[A^T V + VA + P - VDQ^{-1} D^T V \right] x \qquad (13)$$

with boundary conditions

$$W^*(\tau,x) = x^T V(\tau) x = x^T Rx \qquad (14)$$

Since we have used the matrix identity,

$$2x^T VAx = x^T \left[A^T V + VA\right] x \qquad (15)$$

in deriving the first equation.

The true state of the two equations for all values of x, led to the following matrix Riccati equation and the final boundary value for $V(t)$.

$$-V = A^T V + VA + P - VWDQ^{-1}D^T Q \quad (16)$$

$$V(\tau) = R \quad (17)$$

and the optimal state feedback control is given by

$$u^* = -k(t)x, \quad (18)$$

where $k(t) = Q^{-1}D^T V(t)$ (19)

We note that*V(t)* is unknown. The product 2*V(t)x*(t) stands for $\frac{\partial J^*}{\partial x}$ and subsequently, we are able to express $\frac{\partial J^*}{\partial x}$ explicitly in terms of *t* and *x*, since *V(t)* becomes explicitly derivable from (16). This leads to the feedback law.

The Stochastic Bellman equation

The performance measure for the Stochastic Control problem is defined as

$$J = E\left\{ R(x(\tau)) / x(t) = x + \int_t^\tau L(x(r), u(r), r)dr \right\} \quad (20)$$

Assume that τ is fixed and that L and R have the same properties, which also implies in the deterministic LQR problem. With the condition on $x(t)$, equation (20) becomes a function of the initial state x, in addition to the initial time.Now consider the following basic definitions.

Definition 3.1Areal valued stochastic process *W(t)* is called a Wiener process or Brownian motion if

(i) $W(0) = 0$,

(ii) Each sample path is continuous,

(iii) $W(t)$ is Gaussian with $\mu = 0, \sigma^2 = t$ (that is, $W(t)$ is $N(0,t)$),

(iv) For all choices of times $0 < t_1 < t_2 < \cdots < t_m$ the random variables $W(t_1), W(t_2), \cdots, W(t_m) - W(t_{m-1})$ are independent random variables.

Condition (iv) implies that *W* has "independent increments" and we heuristically interpret the one-dimensional "white noise" $\xi(.)$ as equalling $\frac{dW(t)}{dt}$.

To understand stochastic differential equation (SDE) driven by "white noise" that will enable have full meaning of stochastic control problem and its optimality, consider first of all

$$\begin{cases} \dot{X}(t) = f(X(t)) + \sigma\xi(t), (t > 0) \\ X(0) = x^0 \end{cases} \quad (21a)$$

where we informally think of $\xi = \dot{W}$.

Definition 3.2 A stochastic process $X(.)$ solves (21a) if for all times $t \geq 0$, we have

$$X(t) = x_0 + \int_0^t f(X(s))ds + \sigma W(t). \quad (21b)$$

Remarks: (i) It is possible to solve (21b) by the method of successive approximation. With this, set $X_0(.) \equiv x$, and inductively define

$$X_{k+1}(t) = x_0 + \int_0^t f(X_k(s))ds + \sigma W(t).$$

It turns out that $X_k(t)$ converges to a limit $X(t)$ for all $t \geq 0$ and $X(.)$ solves the integral identity (21b).

(ii) Consider a more general SDE

$$\dot{X}(t) = f(X(s)) + G(X(s))\xi(s), (t > 0), \quad (21c)$$

which is read formally as

$$\frac{dX(t)}{dt} = f(X(t)) + G(X(t))\frac{dW(t)}{dt}$$

and then $dX = F(X,u,t)dt + G(X,u,t)dW(t)$.

The Optimal Stochastic Control Problem considered is to minimize equation (20) with respect to the control input, subject to the constraint that the state $x(r)$ satisfies the stochastic differential equation

$$dX = F(X,u,t)dt + G(X,u,t)dW \quad (21)$$

where $dW(t)$ is a Wiener increment as defined earlier, with co-variance matrix $W(t)dt$.

We state below a version of the stochastic principle of Optimality that will be used to develop the condition and optimal state feedback solution.

Proposition: Stochastic optimality principle

If $u^*(r)$ is optimal over the interval $[t, \tau]$, conditioned on the initial state $x(t)$, then $u^*(t)$ is necessarily optimal over the subinterval $[t + \Delta t, \tau]$,

for any $\Delta t \ni \tau - t \geq \Delta t > 0$ (22)

Proof:

With the Markov property of $x(t)$ and the conditioning in the performance measure, it follows that the performance measure value over the subinterval $[t + \Delta t, \tau]$ is conditioned on $x(t + \Delta t)$ completely independent of the value of the $u(t)$ over the interval $[t, t + \Delta t]$. This is really all that is needed to make the deterministic proof of the optimality principle become applicable to stochastic problems.

Derivation of an optimization equation for the stochastic control problem

Assume equation (12) represents the optimal value of the performance measure (20). Since W^* depends on the initial state

$$x(t) = x \quad (23)$$

because of the conditioning on the expected value. As in the deterministic case, assume $u[t, \tau]$, the control input is defined over the interval $[t, \tau]$. Next suppress the arguments of the functions;

$L(x,u,t), R(x), F(x,u,t)$ and $G(x,u,t)$ in the expressions that follows and we have from the definition of the value function $W^*(t,x)$ that;

$$W^*(t,x) = \min_{u\in F(t,\tau)} E\left\{\int_t^{t+\Delta t} Ldr + (\int_{t+\Delta t}^{\tau} Ldr + R) / x(t)\right\} \quad (24)$$

Applying the nested expectations property, we have

$$W^*(t,x) = \min_u E\left\{\int_t^{t+\Delta t} Ldr\right\} + E\left\{(\int_{t+\Delta t}^{\tau} Ldr + R) / \{x(t+\Delta t)\} / x(t)\right\} \quad (25)$$

A careful look at equations (24) and (25) indicates that we have written the conditioning directly in terms of these equations. Then by bringing into focus the stochastic optimality principle, equation (25) becomes

$$W^*(t,x) = \min_u E\left\{\int_t^{t+\Delta t} Ldr + W^*(x(t+\Delta t), t+\Delta t) / x(t)\right\} \quad (26)$$

From equation (26), $x(t+\Delta t)$ which seems as an argument of W^* is a random vector, given by $x(t+\Delta t) = x + \Delta x,$ from the stochastic differential equation. The stochastic increment Δx may be approximated as

$$\Delta x = f\Delta x + G\Delta y \quad (27)$$

and by applying a multivariable Taylor – Series expansion on W* around the point (t, x) and an approximation of the integral can be obtained for our final optimization equation, since the covariance of a Wiener process is linear in Δt .Hence we must utilize the Taylor series expansion up to the third term in Δx . The emerging expansion takes the form

$$W^*(x+\Delta x, t+\Delta t) = W^*(t,x) + \frac{\delta W^*}{\delta t}\Delta t + \left[\frac{\delta W^*}{\delta x}\right]^T \Delta x + \frac{1}{2}(\Delta x)^T H(\Delta x) \quad (28)$$

where the matrix H denotes the Hessian matrix whose i, j – th entries are

$$H = \left[\frac{\partial^2 W^*}{\partial x_i \partial x_j}\right] \quad (29)$$

As such, our approximation of the HJB equation becomes

$$W^*(t,x) = \min_{u(t)} E\left\{\begin{array}{l} L(x,u,t)\Delta t + W^*(t,x) + \\ \frac{\delta W^*}{\delta x}\Delta x + \frac{1}{2}(\Delta x)^T H \frac{1}{2}(\Delta x) / x \end{array}\right\} \quad (30)$$

On applying the conditional certainty property of expectation to equation (30), we have,

$$E\{L(x,u,t)\Delta t / x\} = L(x,u,t)\Delta t E\left\{\left[\frac{\delta W^*}{\delta x}\right]^T (F\Delta t + G\Delta y) / x\right\} t$$

$$= \left[\frac{\delta W^*}{\delta x}\right]^T F(x,u,t)\Delta t \quad (31)$$

Similarly, using the fact that for any vector Z and any symmetric matrix A, we have

$$z^T Az = tr\left[A \geq z^T\right] \quad (32)$$

Then;

$$E\left[(\Delta x)^T H(\Delta x) / x\right] = tr\left[HE\left((\Delta x)(\Delta x)^T / x\right)\right] \quad (33)$$

Next by using ΔY, we have a zero mean and a covariance matrix Y Δt, we have,

$$E[(\Delta x)(\Delta x)^T / x] = E[(F\Delta t + G\Delta y)(F\Delta t + G\Delta y) / x] \quad (34)$$

$$= FF^T(\Delta t)^2 + GYG^T\Delta t \quad (35)$$

It is obvious that the first term in equation (35) is of order two as such we can neglect it in the limit. Taking the limit as $\Delta t \to 0$ and using all of the above computations for conditional expectation, we obtain the Stochastic Hamilton – Jacobi – Bellman equation;

$$-\frac{\delta W^*}{\delta t} = \min_u \left\{\begin{array}{l} L(x,u,t) + \left[\frac{\delta W^*}{\delta x}\right]^T F(x,u,t) + \\ \frac{1}{2} tr\left(HG(x,u,t)YG^T(x,u,t)\right) \end{array}\right\} \quad (36)$$

with boundary condition

$$W^*(\tau, x) = R(x), \forall x.$$

We obtain the condition which results directly from our definition for Jand the property of the conditional expectation if no disturbance is present $(y = 0)$; then the stochastic HJB equation reduces to the deterministic HJB equation. The deterministic HJB equation of the stochastic HJB equation is solved by first computing the feedback function

$$u^* = k(\frac{\delta W^*}{\delta x}, x, t) \quad (37)$$

which minimizes the term in parentheses in equation (36) and when this feedback controller is substituted back into the HJB equation, we eventually solve the emerging PDE for $W^*(t,x)$.

Additive Perturbations

Consider the optimal control problem where the performance criteria is quadratic, the system is linear and the perturbations seem additive. Then

$$J = E\left\{x^T(\tau)Rx(\tau) + \int_t^{\tau}\left(x^T Px + u^T Qu\right)dt / x\right\} \quad (38)$$

Where $dx = (Ax + Du)dt + Gdw \quad (39)$

Equation (38) can be written in terms of the functions, L, R, F and G, where each denotes

$$L(x,u,t) = \left(x^T Px + u^T Qu\right), \quad (40)$$

$$R(x) = x^T Rx, \quad (41)$$

$$F(x,u,t) = Ax + Du, \quad (42)$$

$$G(x,t) = G(t), \quad (43)$$

W is a Wiener process with a different covariance matrix Wdt. (44)

Thus, the input for the optimal stochastic linear quadratic problem includes P, Q, R, A, D, G and W. The minimization step required in the Stochastic HJB involves minimizing

$$u^T Qu + \left[\frac{\delta W^*}{\delta x}\right]^T Du \quad (45)$$

as long as these are the only terms containing u. These terms are the same terms used in the deterministic LQR problem, where the optimal value of u was found as that in equation (9). In the same vein, one may assume a solution for W^* of the form in equation (12)

$$W^* = x^T V(t)x \quad (46)$$

as in the deterministic case. This approach, however introduces a term independent of x on the right hand side of the stochastic HJB equation that would not be balanced on LHS, the best way for this value function is

$$W^* = x^T V(t)x + C(t) \quad (47)$$

Let U^* and W^* be as defined above, on substituting back into the stochastic HJB equation and using the fact that

$$\frac{\delta x^T Qx}{\delta x} = 2Qx \quad (48)$$

and

$$H = 2Q \quad (49)$$

After some matrix manipulations, we have;

$$-x^T Qx - C = x^T(A^T V + VA + P - VDQ^{-1}D^T V)x + tr(VGYG^T) \quad (50)$$

Equating the coefficients of like powers of x, we have the optimization equations as

$$-\dot{V} = A^T V + VA + P - VDQ^{-1}D^T V \quad (51)$$

$$-C = tr(PGYG^T) \quad (52)$$

With boundary conditions as

$$V(\tau) = R \quad (53)$$

$$C(\tau) = 0 \quad (54)$$

and the optimal state feedback control given by

$$U^*(t) = -Q^{-1}D^T V(t)x(t) \quad (55)$$

where $V(t)$ is given as the solution of the differential Riccati equation.

Conclusion

By suppressing the time dependence in LQR, our solution becomes smooth, since the gradient is zero. Also we observed that if no perturbation, the stochastic Hamilton – Jacobi – Bellman equation reduces to the deterministic Hamilton – Jacobi – Bellman equation. In addition, by applying the conditional certainty property expectation, we were able to derive the stochastic Jacobi Bellman equation. Furthermore, the system is uncorrupted by any signal perturbation on the condition that if we assumed that the LQR system is known. Finally, the optimality of the control is perturbed when the state is perturbed by a random process and with this assumption the system state becomes a Markov process.

Conflict of Interest

Authors declare that there is no conflict of interest related to this study.

References

Bellman R 1957. Dynamic Programming, Princeton University Press.

Chung WH & Speyer JL 1998. Stochastic Processes, Estimation and Control.

Fleming WH & Rishel RW 1975. Deterministic and Stochastic Optimal Control, New York – Heidelberg – Berlin. Springer – Verlag.

Golsten H 1950. Classical Mechanics, Addison – Wesley.

Perota P, Ahdallah C & Cerone V 1975. Linear Quadratic Regulator Control: An Introduction Prentice – Hall.

Sontag E 1998. Mathematical Control Theory: Deterministic Finite Dimensional Systems, 2nd Edition, Springer.