Preprints of the 21st IFAC World Congress (Virtual)
Berlin, Germany, July 12-17, 2020

Series Solution of Stochastic Dynamic Programming Equations

Arthur J Krener *

* *Department of Applied Mathematics, Naval Postgraduate School, Monterey, CA 93943 USA (e-mail:ajkrener@ nps.edu).*

Abstract: In this paper we consider discrete time stochastic optimal control problems over infinite and finite time horizons. We show that for a large class of such problems the Taylor polynomials of the solutions to the associated Dynamic Programming Equations can be computed degree by degree.

Keywords: Discrete Time Stochastic Optimal Control, Dynamic Programming, Linear Quadratic Regulator

1. INTRODUCTION

In this paper we consider discrete time stochastic optimal control problems over infinite and finite time horizons. We show that for a large class of such problems the Taylor polynomials of the solutions to the associated Dynamic Programming Equations can be computed degree by degree generalizing the method of Al'brekht (1). There is a vast literature dealing with such problems, we refer the reader to Bertsekas (2). Stochastic optimal control problems in continuous time are discussed in (3) and (6). Navasca generalized Al'brekht's method to deterministic, discrete time, infinite horizon optimal control problems (5).

We begin with a relatively simple stochastic, infinite horizon, optimal control problem and then move on to more complicated problems over infinite and finite horizons. Consider a discrete time, infinite horizon, stochastic Linear Quadratic Regulator with Bilinear Noise (DLQGB),

$$\min_{u(\cdot)} \frac{1}{2}\mathrm{E}\left\{\sum_0^\infty \ (x'Qx+2x'Su+u'Ru)\right\}$$

$$\text{subject to} \qquad x^+ = Fx+Gu+\sum_{k=1}^r w_k(C_kx+D_ku)$$

where $x(0)=x^0$ and $x^+(t)=x(t+1)$.

The state x is n dimensional, the control u is m dimensional and $w(t)=(w_1(t),\dots,w_r(t))'$ is r dimensional sequence of independent Gaussian random vectors of mean zero and covariance $I^{r\times r}$. The matrices are sized accordingly, in particular C_k is an $n\times n$ matrix and D_k is an $n\times m$ matrix for each $k=1,\dots,r$.

To the best of our knowledge discrete time infinite horizon problems with bilinear noise have not been considered before. In (**?**) we studied the continuous time version of this problem. The finite horizon version of this problem with noise entering linearly is well studied in both discrete (2) and continuous time (3), (6).

We restrict our attention to problems with bilinear noise so that we can use power series techniques to solve the dynamic programming equations of nonlinear optimal control problems. The class of infinite horizon nonlinear optimal control problems that are of interest are of the form

$$\min_{u(\cdot)} \mathrm{E}\left\{\sum_0^\infty l(x,u)\right\}$$

$$\text{subject to} \qquad x^+ = f(x,u)+\sum_{k=1}^r w_k\gamma_k(x,u)$$

where $x(0)=x^0$, $f(x,u)$ and $\gamma_k(x,u)$ are smooth functions of order $O(x,u)$ and $l(x,u)$ is a smooth function of order $O(x,u)^2$.

Associated to these problems are Bellman's Dynamic Programming equations for the optimal cost and optimal feedback. Assuming they exist, let $\pi(x)$ be the optimal cost starting at x and $u=\kappa(x)$ be the optimal feedback at x for this problem. Then they satisfy the Stochastic Infinite Horizon Dynamic Programming Equations (SIDPE),

$$\pi(x) = \min_u \mathrm{E}\left\{\pi(f(x,u))+\sum_{k=1}^r w_k\gamma_k(x,u))+l(x,u)\right\} \tag{1}$$

$$\kappa(x) = \mathrm{argmin}_u \mathrm{E}\left\{\pi(f(x,u))+\sum_{k=1}^r w_k\gamma_k(x,u))+l(x,u)\right\} \tag{2}$$

These equations differ from their deterministic counterparts because of the presence of the noise terms.

The class of finite horizon nonlinear optimal control problems that are of interest are of the form

$$\min_{u(\cdot)} \mathrm{E}\left\{\sum_0^T l(t,x,u)+\pi_T(x(T))\right\}$$

$$\text{subject to} \qquad x^+ = f(t,x,u)+\sum_{k=1}^r w_k\gamma_k(t,x,u)$$

* Research supported by AFOSR

$$x(t_0) = x^0$$

where $f(t,x,u)$ and $\gamma_k(t,x,u)$ are smooth n vector vaued functions with respect to x,u of order $O(x,u)$, $l(t,x,u)$ is a smooth scalar valued function with respect to x,u of order $O(x,u)^2$ and $\pi_T(x)$ is a smooth function with respect to x of order $O(x)^2$.

Assuming they exist, let $\pi(t_0,x^0)$ be the optimal cost given that $x(t_0) = x^0$ and $u(t) = \kappa(t,x)$ be the optimal feedback for this problem. Then they satisfy the Stochastic Finite Horizon Dynamic Programming Equations (SFDPE),

$$\pi(t_0,x^0) = \min_{u(t_0)} \mathrm{E}\left\{\pi\left(t_0+1,z^0\right) + l(t_0,x^0,u(t_0))\right\} \tag{3}$$

$$\kappa(t_0,x^0) = \mathrm{argmin}_{u(t_0)} \mathrm{E}\left\{\pi\left(t_0+1,z^0\right) + l(t_0,x^0,u(t_0))\right\} \tag{4}$$

where z^0 is the random vector

$$z^0 = f(t_0,x^0,u(t_0)) + \sum_{k=1}^r w_k(t_0)\gamma_k(t_0,x^0,u(t_0))$$

Again these equations differ from their deterministic counterparts because of the noise terms.

The rest of the this paper is organized as follows. In the next section we solve the infinite horizon discrete time linear quadratic regulator problems with bilinear noise (DLQGB). In this case the SIDPE reduces to stochastic discrete time algebraic Riccati equations (SDARE). To our knowledge the SDARE are new. We present an iterative method for solving SDARE using a solver for the corresponding deterministic algebraic Riccati equation (DARE) such as MATLAB's `dare.m`. This iteration may or may not converge depending on the size of the noise coefficients relative to the deterministic part of the system. Section 3 contains an example of a DLQRB. In Section 4, we turn our attention to nonlinear, nonquadratic problems over an infinite horizon. We show how the Taylor polynomials of the optimal cost $\pi(x)$ and the optimal feedback $u = \kappa(x)$ of the solution of (SIDPE) (1, 2) can be computed degree by degree up to the degree of smoothness of the problem by first solving an SDARE and then solving a sequence of linear algebraic equations. These equations are solvable if the noise coefficients are sufficiently small. Section 5 contains an example of a nonlinear, nonquadratic problem. In Section 6 we consider nonlinear, nonquadratic problems over a finite time interval. We show that the time varying Taylor polynomials of the optimal cost and optimal feedback can be found by first soving a Riccati difference equation and then solving a sequence of linear difference equations. Section 7 is the conclusion.

2. DISCRETE TIME LINEAR QUADRATIC REGULATOR WITH BILINEAR NOISE

If we can find a smooth scalar valued function $\pi(x)$ and a smooth m vector valued $\kappa(x)$ satisfying the Infinite Horizon Stochastic Dynamic Programming Equations (SIDPE) (1, 2) then by a standard verification argument (3) one can show that $\pi(x^0)$ is the optimal cost of starting at x^0 and $u(0) = \kappa(x^0)$ is the optimal control at x^0.

We make the standard assumptions of deterministic LQR,

(1) The matrix $[Q,S;S',R]$ is nonnegative definite.
(2) The matrix R is positive definite.
(3) The pair F, G is stabilizable.
(4) The pair $Q^{1/2}$, F is detectable where $Q = (Q^{1/2})'Q^{1/2}$.

Because of the linear dynamics and quadratic cost, we expect that $\pi(x)$ to be a quadratic function of x and $\kappa(x)$ to be a linear function of x,

$$\pi(x) = \frac{1}{2}x'Px, \qquad \kappa(x) = Kx$$

We plug these expressions into SIDPE and they simplify to

$$\begin{aligned} P &= Q + K'RK + (F+GK)'P(F+GK) \\ &\quad + \sum_{k=1^r}(C_k + D_kK)'P(C_k + D_kK) \end{aligned} \tag{5}$$

$$\begin{aligned} K &= -\left(R + G'PG + \sum_{k=1}^r D_k'PD_k\right)^{-1} \\ &\quad \times \left(G'PF + S' + \sum_{k=1}^r D_k'PC_k\right) \end{aligned} \tag{6}$$

We call these equations (5, 6) the Stochastic Discrete Time Algebraic Riccati Equations (SDARE). They reduce to the deterministic Discrete Time Algebraic Riccati Equations (DARE) if $C_k = 0$ and $D_k = 0$ for $k = 1,\ldots,r$.

Here is an iterative method for solving SDARE. Let $P_{(0)}$ be the solution of the first discrete time deterministic algebraic Riccati equation DARE

$$0 = P_{(0)}F + F'P_{(0)} + Q - (P_{(0)}G + S)R^{-1}(G'P_{(0)} + S')$$

and $K_{(0)}$ be solution of the second deterministic equation DARE

$$K_{(0)} = -R^{-1}(G'P_{(0)} + S')$$

Given $P_{(\tau-1)}$ define

$$Q_{(\tau)} = Q + \sum_{k=1}^r C_k'P_{(\tau-1)}C_k$$

$$R_{(\tau)} = R + \sum_{k=1}^r D_k'P_{(\tau-1)}D_k$$

$$S_{(\tau)} = S + \sum_{k=1}^r C_k'P_{(\tau-1)}D_k$$

Let $P_{(\tau)}$ be the solution of

$$\begin{aligned} 0 &= P_{(\tau)}F + F'P_{(\tau)} + Q_{(\tau)} \\ &\quad - (P_{(\tau)}G + S_{(\tau)})R_{(\tau)}^{-1}(G'P_{(\tau)} + S_{(\tau)}') \end{aligned}$$

and

$$K_{(\tau)} = -R_{(\tau)}^{-1}\left(G'P_{(\tau)} + S_{(\tau)}'\right)$$

If the iteration on $P_{(\tau)}$ converges, that is, for some τ, $P_{(\tau)} \approx P_{(\tau-1)}$ then $P_{(\tau)}$ and $K_{(\tau)}$ are approximate solutions to SDARE

The solutions $P_{(\tau)}$ of the DARE is the kernel of the optimal cost of deterministic LQRs and since

$$\begin{bmatrix} Q & S \\ S' & R \end{bmatrix} \leq \begin{bmatrix} Q_{(\tau-1)} & S_{(\tau-1)} \\ S'_{(\tau-1)} & R_{(\tau-1)} \end{bmatrix} \leq \begin{bmatrix} Q_{(\tau)} & S_{(\tau)} \\ S'_{(\tau)} & R_{(\tau)} \end{bmatrix}$$

it follows that $P_{(0)} \leq P_{(\tau-1)} \leq P_{(\tau)}$, the iteration is monotonically increasing. Computationally we have found that if matrices C_k and D_k are not too big relative to F, G, Q, R, S then the iteration conveges. But if the C_k and D_k are about the same size as F and G or larger the iteration can diverge. Further study of this issue is needed. The iteration does converge in the simple example in the next section.

It is well-known (6) that the first and second standard assumptions of LQR can be violated in a stochastic optimal control problem and still the optimal cost can be finite and positive. This is true for some SLQRB problems and the reason why can be seen in the above iteration. For some $\tau^* > 0$ it may happen that

$$\begin{bmatrix} Q_{(\tau^*)} & S_{(\tau^*)} \\ S'_{(\tau^*)} & R_{(\tau^*)} \end{bmatrix} \geq 0, \qquad R_{(\tau^*)} > 0$$

then this will happen for all $\tau > \tau^*$ even though this might not be true when $\tau = 0$. The MATLAB function `dare.m` only requires that the last two LQR assumptions hold so it can be used in the above iteration even if the first and/or second assumptions are violated.

3. DLQRB EXAMPLE

Here is a simple example with $n = 2, m = 1, r = 2$.

$$\min_u \frac{1}{2} \sum_0^\infty \|x\|^2 + u^2 \, dt$$

$$\text{subject to} \qquad \begin{aligned} x_1^+ &= x_1 + 0.1x_2 + 0.1w_1x_1 \\ x_2^+ &= x_2 + 0.1u + 0.1w_2(x_2 + u) \end{aligned}$$

In other words

$$Q = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad S = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad R = 1$$

$$F = \begin{bmatrix} 1 & 0.1 \\ 0 & 1 \end{bmatrix}, \quad G = \begin{bmatrix} 0 \\ 0.1 \end{bmatrix}$$

$$C_1 = \begin{bmatrix} 0.1 & 0 \\ 0 & 0 \end{bmatrix}, C_2 = \begin{bmatrix} 0 & 0 \\ 0 & 0.1 \end{bmatrix}$$

$$D_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad D_2 = \begin{bmatrix} 0 \\ 0.1 \end{bmatrix}$$

The solution of the noiseless DARE is

$$P = \begin{bmatrix} 18.3422 & 10.9046 \\ 10.9046 & 18.9110 \end{bmatrix}, \; K = -\begin{bmatrix} 0.9170 & 1.6821 \end{bmatrix}$$

The eigenvalues of the noiseless closed loop matrix $F + GK$ are $0.9054 \pm 0.0443i$ and are of norm 0.9065.

The above iteration essentially converges to the solution of the SDARE in about twenty iterations, the solution is

$$P = \begin{bmatrix} 22.3884 & 13.2764 \\ 13.2764 & 21.6311 \end{bmatrix}, \; K = \begin{bmatrix} -1.3276 & -2.1631 \end{bmatrix}$$

The eigenvalues of the noisy closed loop matrix $F + GK$ are $0.8918 \pm 0.0397i$ and are of norm 0.8927.

As expected the noisy system is more difficult to control than the noiseless system and the poles are smaller in norm. It should be noted that the above iteration diverged to infinity when the noise coefficients were increased from 0.1 to 1.

4. NONLINEAR STOCHASTIC INFINITE HORIZON DPE

Suppose the problem is not linear-quadratic, the dynamics is given by a nonlinear stochastic difference equation

$$x^+ = f(x, u) + \sum_{k=1}^r w_k \gamma_k(x, u)$$

and the criterion to be minimized is

$$\min_{u(\cdot)} \mathrm{E} \left\{ \sum_0^\infty l(x, u) \right\}$$

As before the noise $w(t) = (w_1, \ldots, w_r)'$ is a sequence of independent Gaussian vectors of zero mean and covariance $I^{r \times r}$.

We assume that $f(x, u), \gamma_k(x, u), l(x, u)$ are smooth functions that have Taylor polynomial expansions around $x = 0, u = 0$. We also assume that $f(x, u) = O(x, u)$, $\gamma_k(x, u) = O(x, u)$ and $l(x, u) = O(x, u)^2$ so their Taylor polynomial expansions are of the forms

$$\begin{aligned} f(x, u) &= Fx + Gu + f^{[2]}(x, u) \\ &\quad + \ldots + f^{[d]}(x, u) + O(x, u)^{d+1} \\ \gamma_k(x, u) &= C_k x + D_k u + \gamma_k^{[2]}(x, u) \\ &\quad + \ldots + \gamma_k^{[d]}(x, u) + O(x, u)^{d+1} \\ l(x, u) &= \frac{1}{2} \left(x'Qx + 2x'Su + u'Ru \right) + l^{[3]}(x, u) + \ldots \\ &\quad + l^{[d+1]}(x, u) + O(x, u)^{d+2} \end{aligned}$$

where $^{[d]}$ indicates the polynomial terms of homogeneous degree d.

Then if they exist the optimal cost $\pi(x)$ and optimal feedback $u = \kappa(x)$ satisfy SIDPE (1. 2). The quantity to be minimized is a smooth function of u hence (1. 2) imply

$$\begin{aligned} \pi(x) &= \mathrm{E} \left\{ \pi \left(f(x, \kappa(x)) + \sum_{k=1}^r w_k \gamma_k(x, \kappa(x)) \right) \right\} \\ &\quad + l(x, \kappa(x)) \end{aligned} \tag{7}$$

$$\begin{aligned} 0 &= \mathrm{E} \left\{ \frac{\partial \pi}{\partial x} \left(f(x, \kappa(x)) + \sum_{k=1}^r w_k \gamma_k(x, \kappa(x)) \right) \right. \\ &\quad \left. \times \left(\frac{\partial f}{\partial u}(x, \kappa(x)) + \sum_k w_k \frac{\partial \gamma_k}{\partial u}(x, \kappa(x)) \right) \right\} \\ &\quad + \frac{\partial l}{\partial u}(x, \kappa(x)) \end{aligned} \tag{8}$$

We call these the simplified Stochastic, Infinite Horizon Dynamic Programming Equations (sSIDPE). Of course the reverse implication is not necessarily true as the quantity to be minimized could have local minima or stationary points.

We assume that the optimal cost and optimal feedback have similar Taylor polynomial expansions

$$\pi(x) = \frac{1}{2}x'Px + \pi^{[3]}(x) + \ldots + \pi^{[d+1]}(x) + O(x)^{d+2}$$
$$\kappa(x) = Kx + \kappa^{[2]}(x) + \ldots + \kappa^{[d]}(x) + O(x)^{d+1}$$

We plug all these expansions into equations (7, 8). At lowest degrees, degree two in (7) and degree one in (8) we get the familiar SDARE (5, 6).

If (5, 6) are solvable then we may proceed to the next degrees, degree three in (7) and degree two in (8).

$$\begin{aligned}\pi^{[3]}(x) &= \mathrm{E}\left\{\pi^{[3]}\left((F+GK)x + \sum_k w_k(C_k + D_kK)x\right)\right\} \\ &\quad + \sum_k x'(C_k+D_kK)'P\gamma_k^{[2]}(x,Kx) + l^{[3]}(x,Kx) \\ &\quad + x'(F+GK)'Pf^{[2]}(x,Kx) \qquad (9)\end{aligned}$$

$$\begin{aligned}0 &= \mathrm{E}\left\{\frac{\partial\pi^{[3]}}{\partial x}\left((F+GK)x + \sum_k w_k(C_k+D_kK)x\right)\right. \\ &\quad \left.\times\left(G + \sum_k w_kD_k\right)\right\} \\ &\quad + x'(F+GK)'P\frac{\partial f^{[2]}}{\partial u}(x,Kx) + \frac{\partial l^{[3]}}{\partial u}(x,Kx) \\ &\quad + (\kappa^{[2]}(x))'\left(R + G'PG + \sum_k D_k'PD_k\right) \qquad (10)\end{aligned}$$

Notice the first equation (9) is a square linear equation for the unknown $\pi^{[3]}(x)$, the other unknown $\kappa^{[2]}(x)$ does not appear in it. If we can solve it for $\pi^{[3]}(x)$ then we can solve the second equation (9) for $\kappa^{[2]}(x)$ because of the second standard LQR assumption that R is invertible which implies $R + G'PG + \sum_k D_kPD_k$ is also be invertible. But again if the second LQR assumption does not hold, the matrix $R+G'PG+\sum_k D_kPD_k$ might still be invertible.

In the deterministic case the eigenvalues of the linear operator

$$\pi^{[3]}(x) \mapsto \pi^{[3]}\left((F+GK)x\right) \qquad (11)$$

are the products of three eigenvalues of $F+GK$. Under the standard LQR assumptions all the eigenvalues of $F+GK$ are in the open unit disc so any product of three eigenvalues of $F+GK$ has norm less than one. Hence the operator

$$\pi^{[3]}(x) \mapsto \pi^{[3]}(x) - \pi^{[3]}\left((F+GK)x\right) \qquad (12)$$

is invertible. If the noise coefficients $C_k,\ D_k$ are small relative to the eigenvalues of (11) then the operator

$$\begin{aligned}\pi^{[3]}(x) &\mapsto \pi^{[3]}(x) - \qquad (13) \\ &\mathrm{E}\left\{\pi^{[3]}\left((F+GK)x + \sum_k w_k(C_k+D_kK)x\right)\right\}\end{aligned}$$

will also be invertible and so we can solve (9) for $\pi^{[3]}(x)$ and then (10) for $\kappa^{[2]}(x)$.

The first SIDPE equation for $\pi^{[d+1]}(x)$ contains previously computed lower degree terms and the linear operator is

$$\begin{aligned}\pi^{[d+1]}(x) &\mapsto \pi^{[d+1]}(x) \qquad (14) \\ &-\mathrm{E}\left\{\pi^{[d+1]}\left((F+GK)x + \sum_k (C_k+D_kK)x\,w_k\right)\right\}\end{aligned}$$

The eigenvalues of deterministic part of this operator

$$\pi^{[d+1]}(x) \mapsto \pi^{[d+1]}(x) - \pi^{[d+1]}\left((F+GK)x\right) \qquad (15)$$

are of the form $1 - \lambda_{i_1}\cdots\lambda_{i_{d+1}}$ where λ_{i_j} are eigenvalues of $F+Gk$ which are strictly inside the unit disk. Hence (15) is always invertible and its stochastic perturbation (14) will be also if C_k and D_k are small enough.

5. NONLINEAR EXAMPLE

Here is a simple example with $n=2, m=1, r=2$. Consider a pendulum of length $1\ m$ and mass $1\ kg$ orbiting approximately 400 kilometers above Earth on the International Space Station (ISS). The "gravity constant" at this height is approximately $g = 8.7\ m/sec^2$. The pendulum can be controlled by a torque u that can be applied at the pivot and there is damping at the pivot with linear damping constant $c_1 = 0.1\ kg/sec$ and cubic damping constant $c_3 = 0.05\ kg\ sec/m^2$. Let x_1 denote the angle of pendulum measured counter clockwise from the outward pointing ray from the center of the Earth and let x_2 denote the angular velocity. The continuous time determistic equations of motion are

$$\begin{aligned}\dot{x}_1 &= x_2 \\ \dot{x}_2 &= lg\sin x_1 - c_1x_2 - c_3x_2^3 + u\end{aligned}$$

The goal is to find a feedback $u = \kappa(x)$ that stabilizes the pendulum to straight up in spite of the noises so we take the continuous time criterion to be

$$\min_u \frac{1}{2}\int_0^\infty \|x\|^2 + u^2\ dt$$

We time discretize this problem using Euler's method with a time step of 0.02 seconds to get the discrete time optimal control problem of minimizing

$$\min_u 0.01\sum_{t=0}^\infty \|x\|^2 + u^2$$

subject to

$$\begin{aligned}x_1^+ &= x_1 + 0.02x_2 \\ x_2^+ &= x_2 + 0.02\left(lg\sin x_1 - c_1x_2 - c_3x_2^3 + u\right)\end{aligned}$$

But the shape of the earth is not a perfect sphere and its density is not uniform so there are fluctuations in the "gravity constant". We model these relative fluctuations in the "gravity constant" by $0.1w_1$ although they are probably much smaller. There might also be relative fluctuations in the damping constants modeled by $0.1w_2$. We model these stochastically by two white noises,

$$\begin{aligned}x_1^+ &= x_1 + 0.02x_2 \\ x_2^+ &= x_2 + 0.02\left(lg\sin x_1 - c_1x_2 - c_3x_2^3 + u\right) \\ &\quad + 0.02\left(0.1w_1lg\sin x_1 - 0.1w_2(c_1x_2 + c_3x_2^3)\right)\end{aligned}$$

This is an example about how stochastic models with noise coefficients of order $O(x,u)$ can arise. If the noise is modeling an uncertain environment then its coefficients are likely to be $O(1)$. But if it is the model that is uncetain then noise coefficients are likely to be $O(x, u)$.

The linear coefficients of the dynamics are

$$F = \begin{bmatrix} 1 & 0.02 \\ 0.1740 & 0.9980 \end{bmatrix}, \quad G = \begin{bmatrix} 0 \\ 0.02 \end{bmatrix},$$
$$Q = \begin{bmatrix} 0.02 & 0 \\ 0 & 0.02 \end{bmatrix}, \quad R = 0.02, \quad S = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$
$$C_1 = \begin{bmatrix} 0 & 0 \\ 0.0174 & 0 \end{bmatrix}, C_2 = \begin{bmatrix} 0 & 0 \\ 0 & -0.0002 \end{bmatrix},$$
$$D_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad D_2 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

The above iteration converges in six steps to the solution of SDARE (5, 6),

$$P = \begin{bmatrix} 54.9340 & 17.9795 \\ 17.9795 & 6.0744 \end{bmatrix}, \; K = [\,-17.9795 \;\; -6.0744\,]$$

The eigenvalues of $F + GK$ are 0.9483 and 0.9282.

By way of comparison if we delete the noise terms from the problem then the solution to DARE is

$$P = \begin{bmatrix} 54.8930 & 17.9739 \\ 17.9739 & 6.0734 \end{bmatrix}, \; K = [\,-16.9694 \;\; -5.7253\,]$$

and the eigenvalues of $F + GK$ are 0.9510 and 0.9325.

The dynamics is an odd function of x, u so its quadratic and quartic terms are zero. The cubic terms are

$$f^{[3]}(x,u) = \begin{bmatrix} 0 \\ -0.029x_1^3 - 0.001x_2^3 \end{bmatrix}$$
$$\gamma_1^{[3]}(x,u) = \begin{bmatrix} 0 \\ -0.0029x_1^3 \end{bmatrix}$$
$$\gamma_2^{[3]}(x,u) = \begin{bmatrix} 0 \\ -0.0001x_1^3 \end{bmatrix}$$

and the quintic terms are

$$f^{[5]}(x,u) = \begin{bmatrix} 0 \\ 0.00145x_1^5 \end{bmatrix}$$
$$\gamma_1^{[5]}(x,u) = \begin{bmatrix} 0 \\ 0.000145x_1^5 \end{bmatrix}$$
$$\gamma_2^{[5]}(x,u) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Because the Lagrangian is an even function and the dynamics is an odd function of x, u we know that $\pi(x)$ is an even function of x and $\kappa(x)$ is an odd function of x.

We have computed the optimal cost $\pi(x)$ to degree 6 and the optimal feedback $\kappa(x)$ to degree 5,

$$\begin{aligned}
\pi(x) = {} & 27.4670x_1^2 + 17.9795x_1x_2 + 3.0372x_2^2 \\
& -4.4633x_1^4 - 2.7258x_1^3x_2 - 0.4995_1^2x_2^2 \\
& -0.0796x_1x_2^3 - 0.0169x_2^4 \\
& 0.3860x_1^6 + 01976x_1^5x_2 + 0.0266x_1^4x_2^2 + 0.0021_1^3x_2^3 \\
& -0.0003x_1^2x_2^4 - 0.0001x_1x_2^5 + 0.00004x_2^6 \\
\kappa(x) = {} & -17.9795x_1 - 6.0744x_2 \\
& 2.7244x_1^3 + 0.9604x_1^2x_2 + 0.1913x_1x_2^2 + 0.0557x_2^3 \\
& -0.17347x_1^5 + -0.0359x_1^4x_2 + 0.0056x_1^3x_2^2 \\
& +0.0048x_1^2x_2^3 + 0.0010x_1x_2^4 - 0.0001x_2^5
\end{aligned}$$

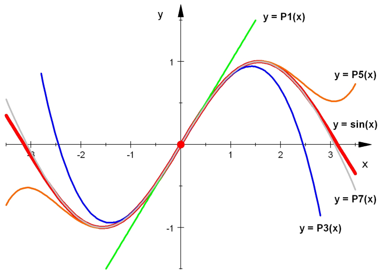

Fig. 1. Taylor approximations of sin(x)

In making this computation we are approximating $\sin x_1$ by its Taylor polynomials

$$\sin x_1 = x_1 - \frac{x_1^3}{6} + \frac{x_1^5}{120} + \ldots$$

The alternating signs of the odd terms in these polynomials are reflected in the nearly alternating signs in the Taylor polynomials of the optimal cost $\pi(x)$ and optimal feedback $\kappa(x)$. If we take a first degree approximation to $\sin x_1$ we are overestimating the gravitational force pulling the pendulum from its upright position so $\pi^{[2]}(x)$ overestimates the optimal cost and the feedback $u = \kappa^{[1]}(x)$ is stronger than it needs to be. This could be a problem if there is a bound on the magnitude of u that we ignored in the analysis. If we take a third degree approximation to $\sin x_1$ then $\pi^{[2]}(x) + \pi^{[4]}(x)$ under estimates the optimal cost and the feedback $u = \kappa^{[1]}(x) + \kappa^{[3]}(x)$ is weaker than it needs to be. If we take a fifth degree approximation to $\sin x_1$ then $\pi^{[2]}(x) + \pi^{[4]}(x) + \pi^{[6]}(x)$ over estimates the optimal cost but by a smaller margin than $\pi^{[2}(x)$. The feedback $u = \kappa^{[1]}(x) + \kappa^{[3]}(x) + \kappa^{[5]}(x)$ is stronger than it needs to be but by a smaller margin than $u = \kappa^{[1]}(x)$.

6. NONLINEAR STOCHASTIC FINITE HORIZON DPE

Consider the finite horizon stochastic nonlinear optimal control problem,

$$\min_{u(\cdot)} \mathrm{E} \left\{ \sum_{t=0}^{T-1} l(t,x,u) + \pi_T(x(T)) \right\}$$

subject to

$$x^+ = f(t,x,u) + \sum_{k=1}^{r} w_k \gamma_k(t,x,u)$$

Again we assume that f, l, γ_k, π_T are sufficiently smooth.

If they exist and are smooth the optimal cost $\pi(t, x)$ of starting at x at time t and the optimal feedback $u(t) = \kappa(t, x(t))$ satisfy the Stochastic Finite Horizon Dynamic Programming Equations (SFDPE) (3, 4)

The quantity to be minimized is a smooth function of u hence (3. 4) imply

$$\pi(t,x) = l(t,x,\kappa(t,x)) \tag{16}$$
$$+\mathrm{E}\left\{\pi\left(t+1, f(t,x,\kappa(t,x)) + \sum_{k=1}^{r} w_k\gamma_k(t,x,\kappa(t,x))\right)\right\}$$
$$0 = \mathrm{E}\left\{\frac{\partial\pi}{\partial x}\left(t+1, f(t,x,\kappa(t,x)) + \sum_{k=1}^{r} w_k\gamma_k(t,x,\kappa(t,x))\right)\right.$$
$$\left.\times\left(\frac{\partial f}{\partial u}(t,x,\kappa(t,x)) + \sum_k w_k\frac{\partial\gamma_k}{\partial u}(t,x,\kappa(t,x))\right)\right\}$$
$$+\frac{\partial l}{\partial u}(t,x,\kappa(t,x)) \tag{17}$$

We call these the simplified Stochastic, Finite Horizon Dynamic Programming Equations (sSFDPE). Of course the reverse implication is not necessarily true as the quantity to be minimized could have local minima or stationary points.

These equations are solved backward in time from the final condition

$$\pi(T,x) = \pi_T(x)$$

Again we assume that we have the following Taylor expansions

$$f(t,x,u) = F(t)x + G(t)u + f^{[2]}(t,x,u) + f^{[3]}(t,x,u) + \dots$$
$$l(t,x,u) = \frac{1}{2}\left(x'Q(t)x + 2x'S(t)u + u'R(t)u\right) + l^{[3]}(t,x,u)$$
$$+l^{[4]}(t,x,u) + \dots$$
$$\gamma_k(t,x,u) = C_k(t)x + D_k(t)u + \gamma_k^{[2]}(t,x,u) + \dots$$
$$\pi_T(x) = \frac{1}{2}x'P_Tx + \pi_T^{[3]}(x) + \pi_T^{[4]}(x) + \dots$$
$$\pi(t,x) = \frac{1}{2}x'P(t)x + \pi^{[3]}(t,x) + \pi^{[4]}(t,x) + \dots$$
$$\kappa(t,x) = K(t)x + \kappa^{[2]}(t,x) + \kappa^{[3]}(t,x) + \dots$$

where $^{[r]}$ indicates terms of homogeneous degree r in x,u with coefficients that are continuous functions of t. The key assumption is that $\gamma_k(t,0,0) = 0$ for then (16, 17, 18) are amenable to power series methods.

We plug these expansions into the simplified Finite Horizon Stochastic Dynamic Programming Equations (16, 17) and collect terms of lowest degree, that is, degree two in (16), degree one in (17) and degree two in (18). We plug these into SIDPE which simplifies to

$$P(t) + Q(t) + K'(t)S(t) + S(t)K'(t) + K'(t)R(t)K(t)$$
$$+(F(t)+G(t)K(t))'P(t+1)(F(t)+G(t)K(t))$$
$$+\sum_{k=1^r}(C_k(t)+D_k(t)K(t))'P(t+1)(C_k(t)+D_k(t)K(t))$$

$$K(t) = -\left(\bar{R}(t)\right)^{-1}$$
$$\times\left(G'(t)P(t+1)F(t) + S'(t) + \sum_{k=1}^{r} D_k'(t)P(t+1)C_k(t)\right)$$

where

$$\bar{R}(t) = R(t) + G'(t)P(t+1)G(t)$$
$$+\sum_{k=1}^{r} D_k'(t)P(t+1)D_k(t)$$

We call these equations the stochastic discrete time Riccati difference equations (SDRDE). These difference equations are solved backward in time from the terminal condition $P(T) = P_T$.

Then we may proceed to the next degrees,

$$\pi^{[3]}(t,x) = \mathrm{E}\left\{\pi^{[3]}\left(t+1, z(t,x,w)\right)\right\}$$
$$+x'(F(t)+G(t)K(t))'P(t+1)f^{[2]}(t,x,Kx)$$
$$+\sum_k x'(C_k(t)+D_k(t)K(t))'P(t+1)\gamma_k^{[2]}(t,x,Kx)$$
$$+l^{[3]}(t,x,Kx)$$
$$0 = \mathrm{E}\left\{\frac{\partial\pi^{[3]}}{\partial x}\left(t, z(t,x,w)\right)\left(G(t) + \sum_k w_kD_k(t)\right)\right\}$$
$$+x'P(t+1)\frac{\partial f^{[2]}}{\partial u}(t,x,K(t)x) + \frac{\partial l^{[3]}}{\partial u}(t,x,K(t)x)$$
$$+(\kappa^{[2]}(t,x))'R(t)$$

where

$$z(t,x,w) = (F(t)+G(t)K(t))x$$
$$+\sum_k w_k(C_k(t)+D_k(t)K(t))x$$

Notice again the unknown $\kappa^{[2]}(t,x)$ does not appear in the first equation which is linear difference equation for $\pi^{[3]}(t,x)$ running backward in time from the terminal condition, $\pi^{[3]}(t,x) = \pi_T^{[3]}(x)$. We can solve it and if $R(t)$ is invertible then we can solve the second equation for $\kappa^{[2]}(t,x)$. The higher degree terms can be found in a similar fashion.

7. CONCLUSION

We have shown how the Taylor polynomials of the optimal cost and optimal feedback for some stochastic, discrete time optimal control problems can be computed degree by degree.

REFERENCES

[1] E. G. Al'brekht, *On the optimal stabilization of nonlinear systems*, PMM-J. Appl. Math. Mech., 25:1254-1266, 1961.

[2] D. Bertsekas, *Dynamic Programming and Optimal Control*, Vol. 1, 4th Ed. Athena Scientific, 2017.

[3] W. Fleming and R. Rishel, *Deterministic and Stochastic Optimal Control*, Springer. New York, 1975.

[4] A. J. Krener, *Stochastic HJB Equations and Regular Singular Points*, , in Modeling, Stochastic Control, Optimization, and Applications, G. Yin and Q. Zhang, eds., IMA Volumes in Mathematics and its Applications, Springer Nature, Switzerland, pages 351-368.

[5] C. Navasca, *Local Solutions of the Dynamic Programming Equations and the Hamilton-Jacobi-Bellman PDEs* PhD Thesis, University of California, Davis, 2002.

[6] J. Yong and X. J. Zhou, *Stochastic Controls, Hamiltonian Systems and HJB Equations*, Springer. New York, 1999.