

Stochastic Differential Dynamic Programming

Evangelos Theodorou, Yuval Tassa & Emo Todorov

Abstract— We present a generalization of the classic Differential Dynamic Programming algorithm. We assume the existence of state- and control-dependent process noise, and proceed to derive the second-order expansion of the cost-to-go. We find the correction terms that arise from the stochastic assumption. Despite having quartic and cubic terms in the initial expression, we show that these vanish, leaving us with the same quadratic structure as standard DDP.

I. INTRODUCTION

Optimal Control describes the choice of actions which minimizes future costs. In the continuous nonlinear case, *local methods* are the only class of algorithms which successfully solve general, high-dimensional Optimal Control problems. These methods are based on the observation that optimal solutions form extremal trajectories, i.e. are solutions to a calculus-of-variations problem.

Differential Dynamic Programming, or DDP, is a powerful local dynamic programming algorithm, which generates both open and closed loop control policies along a trajectory. The DDP algorithm, introduced in [1], computes a quadratic approximation of the cost-to-go around a trajectory and correspondingly, a local linear-feedback controller. Serving as a basis for many popular control frameworks [2], and recently used for advanced aerodynamic control [3] As in the famous Linear-Quadratic-Gaussian case, fixed additive noise has no effect on the controllers generated by DDP, and when described in the literature, the dynamics are usually assumed deterministic.

While the impartiality to noise can be considered a feature if the noise is indeed fixed, in many cases varying noise covariance is an important feature of the problem, as with control-multiplicative noise which is common in biological systems [4]. This latter case was addressed within the iterative-LQG framework [4], but the more general case of state-dependent noise appears to have never been appropriately tackled.

In this paper, we derive the DDP algorithm for general state- and control-dependent noise terms. We find that despite the potential of cubic and quartic terms, these cancel out, allowing us to maintain the quadratic form of the approximation.

E. Theodorou is with the Computational Learning and Motor Control Lab, Departments of Computer Science and Neuroscience, University of Southern California etheodor@usc.edu

Y. Tassa is with the Interdisciplinary Center for Neural Computation, Hebrew University, Jerusalem, Israel tassa@alice.nc.huji.ac.il

E. Todorov is with the Department of Computer Science and Engineering and the Department of Applied Mathematics, University of Washington Seattle, WA, todorov@cs.washington.edu

II. STOCHASTIC DIFFERENTIAL DYNAMIC PROGRAMMING

For our analysis of stochastic Differential Dynamics Programming we will consider systems described by the stochastic continuous-time dynamics:

$$d\mathbf{x} = f(\mathbf{x}, \mathbf{u})dt + F(\mathbf{x}, \mathbf{u})d\omega \quad (1)$$

where $\mathbf{x} \in \mathfrak{R}^{n \times 1}$ is the state, $\mathbf{u} \in \mathfrak{R}^{p \times 1}$ is the control and $d\omega \in \mathfrak{R}^{m \times 1}$ is brownian noise. To enhance the readability of our derivations we will write the dynamics as a function $\Phi \in \mathfrak{R}^{n \times 1}$ of the state, control and instantiation of the noise:

$$\Phi(\mathbf{x}, \mathbf{u}, d\omega) \equiv f(\mathbf{x}, \mathbf{u})dt + F(\mathbf{x}, \mathbf{u})d\omega \quad (2)$$

It will sometimes be convenient to write the matrix $F(\mathbf{x}, \mathbf{u}) \in \mathfrak{R}^{n \times p}$ in terms of its rows or columns:

$$F(\mathbf{x}, \mathbf{u}) = \begin{bmatrix} F_r^1(\mathbf{x}, \mathbf{u}) \\ \vdots \\ F_r^n(\mathbf{x}, \mathbf{u}) \end{bmatrix} = \left[F_c^1(\mathbf{x}, \mathbf{u}), \dots, F_c^p(\mathbf{x}, \mathbf{u}) \right]$$

Every element of the vector $\Phi(\mathbf{x}, \mathbf{u}, d\omega) \in \mathfrak{R}^{n \times 1}$ can now be expressed as:

$$\Phi^j(\mathbf{x}, \mathbf{u}, d\omega) = f^j(\mathbf{x}, \mathbf{u})\delta t + F_r^j(\mathbf{x}, \mathbf{u})d\omega$$

Given a nominal trajectory of states and controls $(\bar{\mathbf{x}}, \bar{\mathbf{u}})$ we expand the dynamics around this trajectory to second order:

$$\begin{aligned} \Phi(\bar{\mathbf{x}} + d\mathbf{x}, \bar{\mathbf{u}} + d\mathbf{u}, d\omega) = \\ \Phi(\bar{\mathbf{x}}, \bar{\mathbf{u}}, d\omega) + \nabla_x \Phi \cdot \delta \mathbf{x} + \nabla_u \Phi \cdot \delta \mathbf{u} \\ + \mathbf{O}(\delta \mathbf{x}, \delta \mathbf{u}, d\omega) \end{aligned}$$

where $\mathbf{O}(\delta \mathbf{x}, \delta \mathbf{u}, d\omega) \in \mathfrak{R}^{n \times 1}$ contains all the second order terms in the deviations in states, controls and noise¹. Writing this term element-wise:

$$\mathbf{O}(\delta \mathbf{x}, \delta \mathbf{u}, d\omega) = \begin{pmatrix} O^{(1)}(\delta \mathbf{x}, \delta \mathbf{u}, d\omega) \\ \vdots \\ O^{(n)}(\delta \mathbf{x}, \delta \mathbf{u}, d\omega) \end{pmatrix},$$

we can express the elements $O^{(j)}(\delta \mathbf{x}, \delta \mathbf{u}, d\omega) \in \mathfrak{R}$ as:

$$\begin{aligned} O^{(j)}(\delta \mathbf{x}, \delta \mathbf{u}, d\omega) = \\ \frac{1}{2} \begin{pmatrix} \delta \mathbf{x} \\ \delta \mathbf{u} \end{pmatrix}^T \begin{pmatrix} \nabla_{\mathbf{xx}} \Phi^j & \nabla_{\mathbf{xu}} \Phi^j \\ \nabla_{\mathbf{ux}} \Phi^j & \nabla_{\mathbf{uu}} \Phi^j \end{pmatrix} \begin{pmatrix} \delta \mathbf{x} \\ \delta \mathbf{u} \end{pmatrix}. \end{aligned}$$

¹Not to be confused with “big-O”.

We would now like to express the derivatives of Φ in terms of the given quantities. Beginning with the first-order terms, we find that:

$$\begin{aligned}\nabla_{\mathbf{x}}\Phi &= \nabla_{\mathbf{x}}f(\mathbf{x}, \mathbf{u})dt + \nabla_{\mathbf{x}}\left(\sum_{i=1}^m F_c^i d\omega^i\right)\sqrt{dt} \\ \nabla_{\mathbf{u}}\Phi &= \nabla_{\mathbf{u}}f(\mathbf{x}, \mathbf{u})dt + \nabla_{\mathbf{u}}\left(\sum_{i=1}^m F_c^i d\omega^i\right)\sqrt{dt}\end{aligned}$$

Next we find the second order derivatives and we have that:

$$\begin{aligned}\nabla_{\mathbf{xx}}\Phi^{(j)} &= \nabla_{\mathbf{xx}}f^{(j)}(\mathbf{x}, \mathbf{u})dt + \nabla_{\mathbf{xx}}\left(F_r^{(j)}(\mathbf{x}, \mathbf{u})d\omega\right)\sqrt{dt} \\ \nabla_{\mathbf{uu}}\Phi^{(j)} &= \nabla_{\mathbf{uu}}f^{(j)}(\mathbf{x}, \mathbf{u})dt + \nabla_{\mathbf{uu}}\left(F_r^{(j)}(\mathbf{x}, \mathbf{u})d\omega\right)\sqrt{dt} \\ \nabla_{\mathbf{ux}}\Phi^{(j)} &= \nabla_{\mathbf{ux}}f^{(j)}(\mathbf{x}, \mathbf{u})dt + \nabla_{\mathbf{ux}}\left(F_r^{(j)}(\mathbf{x}, \mathbf{u})d\omega\right)\sqrt{dt} \\ \nabla_{\mathbf{xu}}\Phi^{(j)} &= \left(\nabla_{\mathbf{ux}}\Phi^{(j)}\right)^T\end{aligned}$$

The discrete-time dynamics are now formulated as:

$$\begin{aligned}\delta\mathbf{x}_{t+\delta t} &= \\ &\left(I_{n\times n} + \nabla_{\mathbf{x}}f(\mathbf{x}, \mathbf{u})\delta t + \nabla_{\mathbf{x}}\left(\sum_{i=1}^m F_c^{(i)} d\omega^{(i)}\sqrt{\delta t}\right)\right)\delta\mathbf{x}_t \\ &+ \left(\nabla_{\mathbf{u}}f(\mathbf{x}, \mathbf{u})\delta t + \nabla_{\mathbf{u}}\left(\sum_{i=1}^m F_c^{(i)} d\omega^{(i)}\sqrt{\delta t}\right)\right)\delta\mathbf{u}_t + \\ &+ F(\mathbf{x}, \mathbf{u})\sqrt{\delta t}d\omega + \mathbf{O}(\delta\mathbf{x}, \delta\mathbf{u}, d\omega, \delta t)\end{aligned}$$

with $\delta t = t_{k+1} - t_k$ corresponding to a small discretization interval. Note that the quadratic term \mathbf{O} is now a function of δt . The discretized dynamics can be written in a more compact form by grouping the state, control and noise dependent terms, and leaving the second order term separate:

$$\begin{aligned}\delta\mathbf{x}_{t+\delta t} &= \\ A_t\delta\mathbf{x}_t + B_t\delta\mathbf{u}_t + \Gamma_t d\omega + \mathbf{O}(\delta\mathbf{x}, \delta\mathbf{u}, d\omega, \delta t)\end{aligned}\quad (3)$$

where the matrices $A_t \in \mathfrak{R}^{n\times n}$, $B_t \in \mathfrak{R}^{n\times p}$ and $\Gamma_t \in \mathfrak{R}^{n\times m}$ are defined as

$$\begin{aligned}A_t &= I_{n\times n} + \nabla_{\mathbf{x}}f(\mathbf{x}, \mathbf{u})\delta t \\ B_t &= \nabla_{\mathbf{u}}f(\mathbf{x}, \mathbf{u})\delta t \\ \Gamma_t &= \left[\Gamma^{(1)} \quad \Gamma^{(2)} \quad \dots \quad \Gamma^{(m)} \right]\end{aligned}$$

with $\Gamma^{(i)} \in \mathfrak{R}^{n\times 1}$ defined $\Gamma^{(i)} = \nabla_{\mathbf{u}}F_c^{(i)}\delta\mathbf{u}_t + \nabla_{\mathbf{x}}F_c^{(i)}\delta\mathbf{x}_t + F_c^{(i)}$. For the derivation of the optimal control it is useful to express Γ_t as the summation of terms that depend on variations in state and controls and terms that are independent of such variations. More precisely we will have that:

$$\Gamma_t = \Delta_t(\delta\mathbf{x}, \delta\mathbf{u}) + F(\mathbf{x}, \mathbf{u})\quad (4)$$

where each column vector of Δ_t is defined as $\Delta_t^{(i)}(\delta\mathbf{x}, \delta\mathbf{u}) = \nabla_{\mathbf{u}}F_c^{(i)}\delta\mathbf{u}_t + \nabla_{\mathbf{x}}F_c^{(i)}\delta\mathbf{x}_t$.

III. VALUE FUNCTION SECOND ORDER APPROXIMATION

As in classical DDP, the derivation of stochastic DDP requires the second order expansion of the cost-to-go function around a nominal trajectory $\bar{\mathbf{x}}$:

$$\begin{aligned}V(\bar{\mathbf{x}} + \delta\mathbf{x}) &= \\ V(\bar{\mathbf{x}}) + \nabla_{\mathbf{x}}V^T\delta\mathbf{x} + \frac{1}{2}\delta\mathbf{x}^T V_{\mathbf{xx}}\delta\mathbf{x}\end{aligned}\quad (5)$$

Substitution of the discretized dynamics (3) in the second order Value function expansion (5) results in:

$$\begin{aligned}V(\bar{\mathbf{x}}_{t+\delta t} + \delta\mathbf{x}_{t+\delta t}) &= V(\bar{\mathbf{x}}_{t+\delta t}) \\ &+ \nabla_{\mathbf{x}}V^T(A_t\delta\mathbf{x}_t + B_t\delta\mathbf{u}_t + \Gamma_t d\omega + \mathbf{O}) \\ &+ (A_t\delta\mathbf{x}_t + B_t\delta\mathbf{u}_t + \Gamma_t d\omega + \mathbf{O})^T \\ &\times \nabla_{\mathbf{xx}}V(A_t\delta\mathbf{x}_t + B_t\delta\mathbf{u}_t + \Gamma_t d\omega + \mathbf{O})\end{aligned}\quad (6)$$

Next we will compute $E(V(\bar{\mathbf{x}}_{t+\delta t} + \delta\mathbf{x}_{t+\delta t}))$ which requires the calculation of the expectation of the all the terms that appear in the equation above. This is what the rest of the analysis is dedicated to. More precisely in the next two sections we will calculate the expectation of the terms:

$$E(\nabla_{\mathbf{x}}V^T\delta\mathbf{x}_{t+\delta t})\quad (7)$$

and

$$E(\delta\mathbf{x}_{t+\delta t}^T \nabla_{\mathbf{xx}}V\delta\mathbf{x}_{t+\delta t})\quad (8)$$

where the state deviation $\delta\mathbf{x}_{t+\delta t}$ at time instant $t + \delta t$ is given by the linearized dynamics:

$$\delta\mathbf{x}_{t+\delta t} = A_t\delta\mathbf{x}_t + B_t\delta\mathbf{u}_t + \Gamma_t d\omega + \mathbf{O}\quad (9)$$

The analysis that follows in section III-A consist of the computation of the expectation of the 4 terms that result from the substitution of the linearized dynamics (9) into (7). In section III-B we compute the expectation of the 16 terms that result from the substitution of (9) into (8).

A. Expectation of the first order term of the value function expansion $\nabla_{\mathbf{x}}V^T\delta\mathbf{x}_{t+\delta t}$.

The expectation of the first order term results in:

$$\begin{aligned}E(\nabla_{\mathbf{x}}V^T(A_t\delta\mathbf{x}_t + B_t\delta\mathbf{u}_t + \Gamma_t d\omega + \mathbf{O})) &= \\ \nabla_{\mathbf{x}}V^T(A_t\delta\mathbf{x}_t + B_t\delta\mathbf{u}_t + E(\mathbf{O}))\end{aligned}\quad (10)$$

In order to find the expectation of $\mathbf{O} \in \mathfrak{R}^{n\times 1}$ we need to find the expectation of each one of the elements of this column vector. Thus we will have that:

$$\begin{aligned}E(O^{(j)}(\delta\mathbf{x}, \delta\mathbf{u}, d\omega, \delta t)) &= \\ E\left(\frac{1}{2}\begin{pmatrix} \delta\mathbf{x} \\ \delta\mathbf{u} \end{pmatrix}^T \begin{pmatrix} \nabla_{\mathbf{xx}}\Phi^{(j)} & \nabla_{\mathbf{xu}}\Phi^{(j)} \\ \nabla_{\mathbf{ux}}\Phi^{(j)} & \nabla_{\mathbf{uu}}\Phi^{(j)} \end{pmatrix} \begin{pmatrix} \delta\mathbf{x} \\ \delta\mathbf{u} \end{pmatrix}\right) &= \\ = \frac{\delta t}{2}\begin{pmatrix} \delta\mathbf{x} \\ \delta\mathbf{u} \end{pmatrix}^T \begin{pmatrix} \nabla_{\mathbf{xx}}f^{(j)} & \nabla_{\mathbf{xu}}f^{(j)} \\ \nabla_{\mathbf{ux}}f^{(j)} & \nabla_{\mathbf{uu}}f^{(j)} \end{pmatrix} \begin{pmatrix} \delta\mathbf{x} \\ \delta\mathbf{u} \end{pmatrix} &= \tilde{O}^j\end{aligned}\quad (11)$$

Therefore we will have that:

$$E(\nabla_{\mathbf{x}} V^T \delta \mathbf{x}_{t+\delta t}) = \nabla_{\mathbf{x}} V^T (A_t \delta \mathbf{x}_t + B_t \delta \mathbf{u}_t + \tilde{\mathbf{O}}) \quad (12)$$

Where the term $\tilde{\mathbf{O}}$ is defined as:

$$\tilde{\mathbf{O}}(\delta \mathbf{x}, \delta \mathbf{u}, \delta t) = \begin{pmatrix} \tilde{\mathbf{O}}^{(1)}(\delta \mathbf{x}, \delta \mathbf{u}, \delta t) \\ \dots \\ \dots \\ \tilde{\mathbf{O}}^{(n)}(\delta \mathbf{x}, \delta \mathbf{u}, \delta t) \end{pmatrix} \quad (13)$$

The term $\nabla_{\mathbf{x}} V^T \tilde{\mathbf{O}}$ is quadratic in variations in the states and controls $\delta \mathbf{x}, \delta \mathbf{u}$ and thus there are the symmetric matrices $\mathcal{F} \in \mathbb{R}^{n \times n}$, $\mathcal{Z} \in \mathbb{R}^{m \times m}$ and $\mathcal{L} \in \mathbb{R}^{m \times n}$ such that:

$$\nabla_{\mathbf{x}} V^T \tilde{\mathbf{O}} = \frac{1}{2} \delta \mathbf{x}^T \mathcal{F} \delta \mathbf{x} + \frac{1}{2} \delta \mathbf{u}^T \mathcal{Z} \delta \mathbf{u} + \delta \mathbf{u}^T \mathcal{L} \delta \mathbf{x} \quad (14)$$

with

$$\mathcal{F} = \left(\sum_{j=1}^n \nabla_{\mathbf{xx}} f^{(j)} V_{x_j} \right) \quad (15)$$

$$\mathcal{Z} = \left(\sum_{j=1}^m \nabla_{\mathbf{uu}} f^{(j)} V_{x_j} \right) \quad (16)$$

$$\mathcal{L} = \left(\sum_{j=1}^m \nabla_{\mathbf{ux}} f^{(j)} V_{x_j} \right) \quad (17)$$

From the analysis above we can see that the expectation $\nabla_{\mathbf{x}} V^T \delta \mathbf{x}_{t+\delta t}$ is a quadratic function with respect to variations in states and controls $\delta \mathbf{x}, \delta \mathbf{u}$. As we will prove in the next section the expectation of $\delta \mathbf{x}_{t+\delta t}^T \nabla_{\mathbf{xx}} V^T \delta \mathbf{x}_{t+\delta t}$ is also a quadratic function of variations in states and controls $\delta \mathbf{x}, \delta \mathbf{u}$.

B. Expectation of the second order term of the value function expansion $\delta \mathbf{x}_{t+\delta t}^T \nabla_{\mathbf{xx}} V^T \delta \mathbf{x}_{t+\delta t}$.

In this section we compute all the terms that appear due to the second approximation of the value function. We will have that

$$E(\delta \mathbf{x}_{t+\delta t}^T \nabla_{\mathbf{xx}} V \delta \mathbf{x}_{t+\delta t}) \quad (18)$$

where $\delta \mathbf{x}_{t+\delta t}$ is given by 9. Substitution of 9 to the equation above results in 16 terms. More precisely we will have that

$$E(\delta \mathbf{x}_{t+\delta t}^T \nabla_{\mathbf{xx}} V^T \delta \mathbf{x}_{t+\delta t}) = \mathcal{E}_1 + \mathcal{E}_2 + \mathcal{E}_3 + \mathcal{E}_4 + \mathcal{E}_5 \quad (19)$$

where the terms $\mathcal{E}_1, \mathcal{E}_2, \mathcal{E}_3, \mathcal{E}_4$ and \mathcal{E}_5 are given by the equations:

$$\begin{aligned} \mathcal{E}_1 &= E(\delta \mathbf{x}_t^T A_t^T V_{xx} A_t \delta \mathbf{x}_t) + E(\delta \mathbf{u}_t^T B_t^T V_{xx} B_t \delta \mathbf{u}_t) + \\ &E(\delta \mathbf{x}_t^T A_t^T V_{xx} B_t \delta \mathbf{u}_t) + E(\delta \mathbf{u}_t^T B_t^T V_{xx} A_t \delta \mathbf{x}_t) \\ \mathcal{E}_2 &= E(d\omega^T \Gamma_t^T V_{xx} A_t \delta \mathbf{x}) + E(d\omega^T \Gamma_t^T V_{xx} B_t \delta \mathbf{u}) + \\ &E(\delta \mathbf{x}^T A_t^T V_{xx} \Gamma_t d\omega) + E(\delta \mathbf{u}^T B_t^T V_{xx} \Gamma_t d\omega) + \\ &E(d\omega^T \Gamma_t^T V_{xx} \Gamma_t d\omega) \\ \mathcal{E}_3 &= E(\mathbf{O}^T V_{xx} \Gamma_t d\omega) + E(d\omega^T \Gamma_t^T V_{xx} \mathbf{O}) \\ \mathcal{E}_4 &= E(\delta \mathbf{x}_t^T A_t^T V_{xx} \mathbf{O}) + E(\delta \mathbf{u}_t^T B_t^T V_{xx} \mathbf{O}) + \\ &E(\mathbf{O}^T V_{xx} B_t \delta \mathbf{u}_t) + E(\mathbf{O}^T V_{xx} A_t \delta \mathbf{x}_t) \\ \mathcal{E}_5 &= E(\mathbf{O}^T V_{xx} \mathbf{O}) \end{aligned} \quad (20)$$

As we can see we have classified these terms into 5 classes so that to make our analysis easier and more clear. In the 1th class we have all these terms that depend neither on $d\omega$ and nor on $\mathbf{O}(\delta \mathbf{x}, \delta \mathbf{u}, d\omega, \delta t)$. These are the terms that define \mathcal{E}_1 . In the 2th category \mathcal{E}_2 there are the terms that depend on $d\omega$ but not on $\mathbf{O}(\delta \mathbf{x}, \delta \mathbf{u}, d\omega, \delta t)$. In the third class \mathcal{E}_3 , there are terms that depends both on $\mathbf{O}(\delta \mathbf{x}, \delta \mathbf{u}, d\omega, \delta t)$ and $d\omega$. In the 4th class \mathcal{E}_4 , we have terms that depend on $\mathbf{O}(\delta \mathbf{x}, \delta \mathbf{u}, d\omega, \delta t)$. Finally in the 5th class \mathcal{E}_5 , we have all these terms that depend on $\mathbf{O}(\delta \mathbf{x}, \delta \mathbf{u}, d\omega, \delta t)$ quadratically. The expectation operator will cancel all the terms that include noise up the first order. Moreover, the mean operator for terms that depend on the noise quadratically will result in covariance.

We start with all these terms that belong to the \mathcal{E}_1 class. More precisely we will have that:

$$\begin{aligned} E(\delta \mathbf{x}_t^T A_t^T V_{xx} A_t \delta \mathbf{x}_t) &= \delta \mathbf{x}_t^T A_t^T V_{xx} A_t \delta \mathbf{x}_t \\ E(\delta \mathbf{u}_t^T B_t^T V_{xx} B_t \delta \mathbf{u}_t) &= \delta \mathbf{u}_t^T B_t^T V_{xx} B_t \delta \mathbf{u}_t \\ E(\delta \mathbf{x}_t^T A_t^T V_{xx} B_t \delta \mathbf{u}_t) &= \delta \mathbf{x}_t^T A_t^T V_{xx} B_t \delta \mathbf{u}_t \\ E(\delta \mathbf{u}_t^T B_t^T V_{xx} A_t \delta \mathbf{x}_t) &= \delta \mathbf{u}_t^T B_t^T V_{xx} A_t \delta \mathbf{x}_t \end{aligned} \quad (21)$$

We continue our analysis by calculating all the terms of the class \mathcal{E}_2 . More precisely we will have:

$$\begin{aligned} E(d\omega^T \Gamma_t^T V_{xx} A_t \delta \mathbf{x}) &= 0 \\ E(d\omega^T \Gamma_t^T V_{xx} B_t \delta \mathbf{u}) &= 0 \\ E(d\omega^T \Gamma_t^T V_{xx} A_t \delta \mathbf{x})^T &= 0 \\ E(d\omega^T \Gamma_t^T V_{xx} B_t \delta \mathbf{u})^T &= 0 \end{aligned} \quad (22)$$

The terms above are equal to zero since the brownian noise is zero mean. The expectation of the term that does not depend on $\mathbf{O}(\delta \mathbf{x}, \delta \mathbf{u}, d\omega, \delta t)$ and it is quadratic with respect to the noise is given as follows:

$$E \left(d\omega^T \Gamma_t^T V_{xx} \Gamma_t d\omega \right) = \mathbf{trace} \left(\Gamma_t^T V_{xx} \Gamma_t \Sigma_\omega \right) \quad (23)$$

For those terms that depend both on $\mathbf{O}(\delta\mathbf{x}, \delta\mathbf{u}, d\omega, \delta t)$ and on the noise class \mathcal{E}_3 we will have:

$$\begin{aligned} E \left(\mathbf{O}^T V_{xx} \Gamma_t d\omega \right) &= E \left(\mathbf{trace} \left(V_{xx} \Gamma_t d\omega \mathbf{O}^T \right) \right) \\ &= \mathbf{trace} \left(V_{xx} \Gamma_t E \left(d\omega \mathbf{O}^T \right) \right) \end{aligned} \quad (24)$$

By writing the term $\mathbf{O}(\delta\mathbf{x}, \delta\mathbf{u}, d\omega, \delta t)$ in a matrix form and putting the noise vector insight the this matrix we have:

$$\begin{aligned} E \left(\mathbf{O}^T V_{xx} \Gamma_t d\omega \right) &= \\ \mathbf{trace} \left(V_{xx} \Gamma_t E \left[\begin{array}{ccc} d\omega O^{(1)} & \dots & d\omega O^{(n)} \end{array} \right] \right) \end{aligned} \quad (25)$$

Calculation of the expectation above requires to find the terms $E \left(\sqrt{\delta t} d\omega O^{(j)} \right)$ more precisely we will have:

$$\begin{aligned} E \left(\sqrt{\delta t} d\omega O^{(j)} \right) &= \\ \frac{1}{2} E \left(\sqrt{\delta t} d\omega \delta \mathbf{x}^T \Phi_{\mathbf{xx}}^{(i)} \delta \mathbf{x} \right) &+ \frac{1}{2} E \left(\sqrt{\delta t} d\omega \delta \mathbf{u}^T \Phi_{\mathbf{uu}}^{(i)} \delta \mathbf{u} \right) + \\ E \left(\sqrt{\delta t} d\omega \delta \mathbf{u}^T \Phi_{\mathbf{ux}}^{(i)} \delta \mathbf{x} \right) \end{aligned} \quad (26)$$

We first calculate the term:

$$\begin{aligned} E \left(\sqrt{\delta t} d\omega \delta \mathbf{x}^T \nabla_{\mathbf{xx}} \Phi^{(i)} \delta \mathbf{x} \right) & \\ = E \left(\sqrt{\delta t} d\omega \delta \mathbf{x}^T \left(\nabla_{\mathbf{xx}} f^{(i)} \delta t + \nabla_{\mathbf{xx}} F_r^{(i)} d\omega \sqrt{\delta t} \right) \delta \mathbf{x} \right) & \\ = E \left(\sqrt{\delta t} d\omega \delta \mathbf{x}^T \left(\nabla_{\mathbf{xx}} f^{(i)} \delta t \right) \delta \mathbf{x} \right) & \\ + E \left(\sqrt{\delta t} d\omega \delta \mathbf{x}^T \left(\nabla_{\mathbf{xx}} F_r^{(i)} d\omega \sqrt{\delta t} \right) \delta \mathbf{x} \right) \end{aligned} \quad (27)$$

The term $E \left(\sqrt{\delta t} d\omega \delta \mathbf{x}^T \left(\nabla_{\mathbf{xx}} f^{(i)} \delta t \right) \delta \mathbf{x} \right) = 0$ since it depends linearly on the noise and $E(d\omega) = 0$. The 2th term $E \left(\sqrt{\delta t} d\omega \delta \mathbf{x}^T \left(\nabla_{\mathbf{xx}} F_r^{(i)} d\omega \sqrt{\delta t} \right) \delta \mathbf{x} \right)$ depends quadratically in the noise and thus the expectation operator will result in the variance on the noise. We follow the analysis:

$$\begin{aligned} E \left(\sqrt{\delta t} d\omega \delta \mathbf{x}^T \nabla_{\mathbf{xx}} \Phi^{(i)} \delta \mathbf{x} \right) &= \\ E \left(\sqrt{\delta t} d\omega \delta \mathbf{x}^T \nabla_{\mathbf{xx}} \left(F_r^{(i)} d\omega \sqrt{\delta t} \right) \delta \mathbf{x} \right) \end{aligned} \quad (28)$$

Since the $d\omega = (d\omega_1, \dots, d\omega_m)^T$ and $F_r^{(i)} = (F^{(i1)}, \dots, F^{(im)})$ we will have that:

$$\begin{aligned} E \left(\sqrt{\delta t} d\omega \delta \mathbf{x}^T \nabla_{\mathbf{xx}} \Phi^{(i)} \delta \mathbf{x} \right) &= \\ E \left(\delta t d\omega \delta \mathbf{x}^T \nabla_{\mathbf{xx}} \left(\sum_{j=1}^m F^{(ij)} d\omega^j \right) \delta \mathbf{x} \right) & \\ E \left(\delta t d\omega \delta \mathbf{x}^T \left(\sum_{j=1}^m \nabla_{\mathbf{xx}} \left(F^{(ij)} d\omega^j \right) \right) \delta \mathbf{x} \right) & \\ E \left(\delta t d\omega \delta \mathbf{x}^T \left(\sum_{j=1}^m d\omega^j \nabla_{\mathbf{xx}} \left(F^{(ij)} \right) \right) \delta \mathbf{x} \right) \end{aligned} \quad (29)$$

By writing $d\omega$ in vector form we have that:

$$\begin{aligned} E \left(\sqrt{\delta t} d\omega \delta \mathbf{x}^T \nabla_{\mathbf{xx}} \Phi^{(i)} \delta \mathbf{x} \right) &= \\ E \left(\delta t \begin{bmatrix} d\omega_1 \\ \dots \\ d\omega_2 \end{bmatrix} \delta \mathbf{x}^T \left(\sum_{j=1}^m d\omega^j \nabla_{\mathbf{xx}} \left(F^{(ij)} \right) \right) \delta \mathbf{x} \right) \end{aligned} \quad (30)$$

The term $\delta \mathbf{x}^T \left(\sum_{j=1}^m d\omega^j \nabla_{\mathbf{xx}} \left(F^{(ij)} \right) \right) \delta \mathbf{x}$ is scalar and it can multiply each one of the elements of the noise vector.

$$\begin{bmatrix} \delta t E \left(d\omega_1 \delta \mathbf{x}^T \left(\sum_{j=1}^m d\omega^j \nabla_{\mathbf{xx}} \left(F^{(ij)} \right) \right) \delta \mathbf{x} \right) \\ \dots \\ \delta t E \left(d\omega_2 \delta \mathbf{x}^T \left(\sum_{j=1}^m d\omega^j \nabla_{\mathbf{xx}} \left(F^{(ij)} \right) \right) \delta \mathbf{x} \right) \end{bmatrix} \quad (31)$$

Since $E(d\omega_i d\omega_i) = \sigma_{\omega_i}^2$ and $E(d\omega_i d\omega_j) = 0$ we can show that:

$$\begin{aligned} E \left(\sqrt{\delta t} d\omega \delta \mathbf{x}^T \nabla_{\mathbf{xx}} \Phi^{(i)} \delta \mathbf{x} \right) &= \\ \begin{bmatrix} \delta t \sigma_{d\omega_1}^2 \delta \mathbf{x}^T \nabla_{\mathbf{xx}} F_r^{(i1)} \delta \mathbf{x} \\ \dots \\ \delta t \sigma_{d\omega_m}^2 \delta \mathbf{x}^T \nabla_{\mathbf{xx}} F_r^{(im)} \delta \mathbf{x} \end{bmatrix} \end{aligned} \quad (32)$$

In similar way we can show that:

$$\begin{aligned} E \left(\sqrt{\delta t} d\omega \delta \mathbf{x}^T \nabla_{\mathbf{uu}} \Phi^{(i)} \delta \mathbf{x} \right) &= \\ \begin{bmatrix} \delta t \sigma_{d\omega_1}^2 \delta \mathbf{u}^T \nabla_{\mathbf{uu}} F_r^{(i1)} \delta \mathbf{u} \\ \dots \\ \delta t \sigma_{d\omega_m}^2 \delta \mathbf{u}^T \nabla_{\mathbf{uu}} F_r^{(im)} \delta \mathbf{u} \end{bmatrix} \end{aligned} \quad (33)$$

and

$$E \left(\sqrt{\delta t} d\omega \delta \mathbf{u}^T \nabla_{\mathbf{xu}} \Phi^{(i)} \delta \mathbf{x} \right) = \quad (34)$$

$$\begin{bmatrix} \delta t \sigma_{d\omega_1}^2 \delta \mathbf{u}^T \nabla_{\mathbf{ux}} F_r^{(i1)} \delta \mathbf{x} \\ \dots \\ \delta t \sigma_{d\omega_m}^2 \delta \mathbf{u}^T \nabla_{\mathbf{ux}} F_r^{(im)} \delta \mathbf{x} \end{bmatrix}$$

Since we have calculated all the terms of expression (27) we can proceed with the computation of (24). According to the analysis above the term $E \left(\mathbf{O}^T V_{xx} \Gamma_t d\omega \right)$ can be written as follows:

$$E \left(\mathbf{O}^T V_{xx} \Gamma_t d\omega \right) = \quad (35)$$

$$\text{trace} \left(V_{xx} \Gamma_t \left(\mathcal{M} + \mathcal{N} + \mathcal{G} \right) \right)$$

Where the matrices $\mathcal{M} \in \mathbb{R}^{m \times n}$, $\mathcal{N} \in \mathbb{R}^{m \times n}$ and $\mathcal{G} \in \mathbb{R}^{m \times n}$ are defined as follows:

$$\mathcal{M} = \quad (36)$$

$$\delta t \begin{bmatrix} \sigma_{d\omega_1}^2 \delta \mathbf{x}^T \nabla_{\mathbf{xx}} F_r^{(11)} \delta \mathbf{x} & \dots & \sigma_{d\omega_1}^2 \delta \mathbf{x}^T \nabla_{\mathbf{xx}} F_r^{(1n)} \delta \mathbf{x} \\ \dots & \dots & \dots \\ \sigma_{d\omega_m}^2 \delta \mathbf{x}^T \nabla_{\mathbf{xx}} F_r^{(m1)} \delta \mathbf{x} & \dots & \sigma_{d\omega_m}^2 \delta \mathbf{x}^T \nabla_{\mathbf{xx}} F_r^{(mn)} \delta \mathbf{x} \end{bmatrix}$$

Similarly

$$\mathcal{N} = \quad (37)$$

$$\delta t \begin{bmatrix} \sigma_{d\omega_1}^2 \delta \mathbf{x}^T \nabla_{\mathbf{xu}} F_r^{(1,1)} \delta \mathbf{u} & \dots & \sigma_{d\omega_1}^2 \delta \mathbf{x}^T \nabla_{\mathbf{xu}} F_r^{(1,n)} \delta \mathbf{u} \\ \dots & \dots & \dots \\ \sigma_{d\omega_m}^2 \delta \mathbf{x}^T \nabla_{\mathbf{xu}} F_r^{(m,1)} \delta \mathbf{u} & \dots & \sigma_{d\omega_m}^2 \delta \mathbf{x}^T \nabla_{\mathbf{xu}} F_r^{(m,n)} \delta \mathbf{u} \end{bmatrix}$$

and

$$\mathcal{G} = \quad (38)$$

$$\delta t \begin{bmatrix} \sigma_{d\omega_1}^2 \delta \mathbf{u}^T \nabla_{\mathbf{uu}} F_r^{(1,1)} \delta \mathbf{u} & \dots & \sigma_{d\omega_1}^2 \delta \mathbf{u}^T \nabla_{\mathbf{uu}} F_r^{(1,n)} \delta \mathbf{u} \\ \dots & \dots & \dots \\ \sigma_{d\omega_m}^2 \delta \mathbf{u}^T \nabla_{\mathbf{uu}} F_r^{(m,1)} \delta \mathbf{u} & \dots & \sigma_{d\omega_m}^2 \delta \mathbf{u}^T \nabla_{\mathbf{uu}} F_r^{(m,n)} \delta \mathbf{u} \end{bmatrix}$$

Based on (4) the term Γ_t depends on Δ which is a function of the variations in states and control up to the 1th order. In addition the matrices \mathcal{M} , \mathcal{N} and \mathcal{G} are also functions of the deviations in state and controls up to the 2th order. The product of Δ with each one of the matrices \mathcal{M} , \mathcal{N} and \mathcal{G} will result into 3th order terms that can be neglected. By neglecting these terms we can show that:

$$E \left(\mathbf{O}^T V_{xx} \Gamma_t d\omega \right) = \quad (39)$$

$$= \text{trace} \left(V_{xx} (\Delta + F) (\mathcal{M} + \mathcal{N} + \mathcal{G}) \right)$$

$$= \text{trace} \left(V_{xx} F (\mathcal{M} + \mathcal{N} + \mathcal{G}) \right)$$

Each element (i, j) of the product $\mathcal{C} = V_{xx} F$ can be expressed as $\mathcal{C}^{(i,j)} = \sum_{r=1}^n V_{xx}^{(i,r)} F^{(r,j)}$ where $\mathcal{C} \in \mathbb{R}^{n \times p}$. Furthermore the element (μ, ν) of the product $\mathcal{H} = \mathcal{C} \mathcal{M}$ is

formulated $\mathcal{H}^{(\mu, \nu)} = \sum_{k=1}^n \mathcal{C}^{(\mu, k)} \mathcal{M}^{(k, \nu)}$ with $\mathcal{H} \in \mathbb{R}^{n \times n}$. Thus, the term $\text{trace} (V_{xx} F \mathcal{M})$ can be now expressed as:

$$\text{trace} (V_{xx} F \mathcal{M}) = \sum_{\ell=1}^n \mathcal{H}^{(\ell, \ell)} \quad (40)$$

$$= \sum_{\ell=1}^n \sum_{k=1}^m \mathcal{C}^{(\ell, k)} \mathcal{M}^{(k, \ell)}$$

$$= \sum_{\ell=1}^n \sum_{k=1}^m \left(\sum_{r=1}^n V_{xx}^{(k, r)} F^{(r, \ell)} \right) \mathcal{M}^{(k, \ell)}$$

Since $\mathcal{M}^{(k, \ell)} = \delta t \sigma_{d\omega_1}^2 \delta \mathbf{x}^T \nabla_{\mathbf{xx}} F^{(k, \ell)} \delta \mathbf{x}$ the vectors $\delta t \sigma_{d\omega_1}^2 \delta \mathbf{x}^T$ and $\delta \mathbf{x}$ do not depend on k, ℓ, r and they can be taken outside the sum. Thus we can show that:

$$\text{trace} (V_{xx} F \mathcal{M}) \quad (41)$$

$$= \sum_{\ell=1}^n \sum_{k=1}^m \left(\left(\sum_{r=1}^n V_{xx}^{(k, r)} F^{(r, \ell)} \right) \sigma_{d\omega_1}^2 \delta t \delta \mathbf{x}^T \nabla_{\mathbf{xx}} F^{(k, \ell)} \delta \mathbf{x} \right)$$

$$= \delta \mathbf{x}^T \sigma_{d\omega_1}^2 \delta t \sum_{\ell=1}^n \sum_{k=1}^m \left(\left(\sum_{r=1}^n V_{xx}^{(k, r)} F^{(r, \ell)} \right) \nabla_{\mathbf{xx}} F^{(k, \ell)} \right) \delta \mathbf{x}$$

$$= \delta \mathbf{x}^T \tilde{\mathbf{M}} \delta \mathbf{x}$$

where $\tilde{\mathbf{M}}$ is a matrix of dimensionality $\tilde{\mathbf{M}} \in \mathbb{R}^{n \times n}$ and it is defined as:

$$\tilde{\mathbf{M}} = \sigma_{d\omega_1}^2 \delta t \sum_{\ell=1}^n \sum_{k=1}^m \left(\left(\sum_{r=1}^n V_{xx}^{(k, r)} F^{(r, \ell)} \right) \nabla_{\mathbf{xx}} F^{(k, \ell)} \right) \quad (42)$$

By following the same algebraic steps it can be shown that:

$$\text{trace} (V_{xx} F \mathcal{N}) = \delta \mathbf{x}^T \tilde{\mathbf{N}} \delta \mathbf{u} \quad (43)$$

with $\tilde{\mathbf{N}}$ matrix of dimensionality $\tilde{\mathbf{N}} \in \mathbb{R}^{n \times p}$ defined as:

$$\tilde{\mathbf{N}} = \sigma_{d\omega_1}^2 \delta t \sum_{\ell=1}^n \sum_{k=1}^m \left(\left(\sum_{r=1}^n V_{xx}^{(k, r)} F^{(r, \ell)} \right) \nabla_{\mathbf{xu}} F^{(k, \ell)} \right) \quad (44)$$

and

$$\text{trace} (V_{xx} F \mathcal{G}) = \delta \mathbf{u}^T \tilde{\mathbf{G}} \delta \mathbf{u} \quad (45)$$

with $\tilde{\mathbf{G}}$ matrix of dimensionality $\tilde{\mathbf{N}} \in \mathbb{R}^{p \times p}$ defined as:

$$\tilde{\mathbf{G}} = \sigma_{d\omega_1}^2 \delta t \sum_{\ell=1}^n \sum_{k=1}^m \left(\left(\sum_{r=1}^n V_{xx}^{(k, r)} F^{(r, \ell)} \right) \nabla_{\mathbf{uu}} F^{(k, \ell)} \right) \quad (46)$$

Thus the term $E \left(\mathbf{O}^T V_{xx} \Gamma_t d\omega \right)$ is formulated as:

$$E \left(\mathbf{O}^T V_{xx} \Gamma_t d\omega \right) = \frac{1}{2} \delta \mathbf{x}^T \tilde{\mathbf{M}} \delta \mathbf{x} + \frac{1}{2} \delta \mathbf{u}^T \tilde{\mathbf{G}} \delta \mathbf{u} + \delta \mathbf{x}^T \tilde{\mathbf{N}} \delta \mathbf{u} \quad (47)$$

Similarly we can show that:

$$E \left(d\omega^T \Gamma_t^T V_{xx} \mathbf{O} \right) = \frac{1}{2} \delta \mathbf{x}^T \tilde{\mathbf{M}} \delta \mathbf{x} + \frac{1}{2} \delta \mathbf{u}^T \tilde{\mathbf{G}} \delta \mathbf{u} + \delta \mathbf{x}^T \tilde{\mathbf{N}} \delta \mathbf{u} \quad (48)$$

Next we will find the expectation for all terms that depend on $\mathbf{O}(\delta \mathbf{x}, \delta \mathbf{u}, d\omega, \delta t)$ and not on the noise. Consequently, we will have that:

$$\begin{aligned} E \left(\delta \mathbf{x}_t^T A_t^T V_{xx} \mathbf{O} \right) &= \delta \mathbf{x}_t^T A_t^T V_{xx} \tilde{\mathbf{O}} = 0 \\ E \left(\delta \mathbf{u}_t^T B_t^T V_{xx} \mathbf{O} \right) &= \delta \mathbf{u}_t^T B_t^T V_{xx} \tilde{\mathbf{O}} = 0 \\ E \left(\mathbf{O}^T V_{xx} A_t \delta \mathbf{x}_t \right) &= \tilde{\mathbf{O}}^T V_{xx} A_t \delta \mathbf{x}_t = 0 \\ E \left(\mathbf{O}^T V_{xx} B_t \delta \mathbf{u}_t \right) &= \tilde{\mathbf{O}}^T V_{xx} B_t \delta \mathbf{u}_t = 0 \end{aligned} \quad (49)$$

where the quantity $\tilde{\mathbf{O}}$ has been defined in (13). All the 4 terms above are equal to zero since they have variations in state and control of the order higher than 2 and therefore they can be neglected.

Finally we compute the terms of the 5th class and therefore we have the expression

$$\begin{aligned} \mathcal{E}_5 &= E \left(\mathbf{O}^T V_{xx} \mathbf{O} \right) \\ &= E \left(\text{trace} \left(V_{xx} \mathbf{O} \mathbf{O}^T \right) \right) \\ &= \text{trace} \left(V_{xx} E \left(\mathbf{O} \mathbf{O}^T \right) \right) \\ &= \text{trace} \left(V_{xx} E \left(\begin{bmatrix} O^{(1)} \\ \dots \\ O^{(n)} \end{bmatrix} \begin{bmatrix} O^{(1)} \\ \dots \\ O^{(n)} \end{bmatrix}^T \right) \right) \end{aligned} \quad (50)$$

The product $O^{(i)} O^{(j)}$ is a function of variation in state and control of order 4 since each term $O^{(i)}$ is a function of variation in states and control of order 2. Consequently, the term $\mathcal{E}_5 = E \left(\mathbf{O}^T V_{xx} \mathbf{O} \right)$ is equal to zero.

With the computation of the expectation of term that is quadratic WRT \mathbf{O} we have calculated all the terms of the second order expansion of the cost to go function. In the next section we provide the optimal controls.

IV. OPTIMAL CONTROLS

In this section we provide the form of the optimal controls and we show how previous results are special cases of our generalized stochastic DDP formulation. Furthermore after we have calculated the all the terms of expansion of the cost to go function $V(\mathbf{x}_t)$ at state \mathbf{x}_t we can show that its form remains is quadratic WRT variations in the state $\delta \mathbf{x}_t$ under the constrain of the nonlinear stochastic dynamics in (1). More precisely we can show that:

$$V(\bar{\mathbf{x}}_{t+\delta t} + \delta \mathbf{x}_{t+\delta t}) = V(\bar{\mathbf{x}}_{t+\delta t}) \quad (51)$$

$$+ \nabla_x V^T A_t \delta \mathbf{x}_t + \nabla_x V^T B_t \delta \mathbf{u}_t \quad (52)$$

$$+ \frac{1}{2} \delta \mathbf{x}^T \mathcal{F} \delta \mathbf{x} + \frac{1}{2} \delta \mathbf{u}^T \mathcal{Z} \delta \mathbf{u} + \delta \mathbf{x}^T \mathcal{L} \delta \mathbf{x}$$

$$+ \frac{1}{2} \delta \mathbf{x}_t^T A_t^T V_{xx} A_t \delta \mathbf{x}_t + \frac{1}{2} \delta \mathbf{u}_t^T B_t^T V_{xx} B_t \delta \mathbf{u}_t$$

$$+ \frac{1}{2} \delta \mathbf{x}_t^T A_t^T V_{xx} B_t \delta \mathbf{u}_t + \frac{1}{2} \delta \mathbf{u}_t^T B_t^T V_{xx} A_t \delta \mathbf{x}_t$$

$$+ \frac{1}{2} \text{trace} \left(\Gamma_t^T V_{xx} \Gamma_t \Sigma_\omega \right)$$

$$+ \frac{1}{2} \delta \mathbf{x}^T \tilde{\mathbf{M}} \delta \mathbf{x} + \frac{1}{2} \delta \mathbf{u}^T \tilde{\mathbf{G}} \delta \mathbf{u} + \delta \mathbf{x}^T \tilde{\mathbf{N}} \delta \mathbf{u} \quad (53)$$

The unmaximized state, action value function is defined as follows:

$$Q(\mathbf{x}_k, \mathbf{u}_k) = \ell(\mathbf{x}_k, \mathbf{u}_k) + V(\mathbf{x}_{k+1}) \quad (54)$$

Given a trajectory in states and controls $\bar{\mathbf{x}}, \bar{\mathbf{u}}$ we can approximate the state action value function as follows:

$$\begin{aligned} Q(\bar{\mathbf{x}} + \delta \mathbf{x}, \bar{\mathbf{u}} + \delta \mathbf{u}) &= Q_0 + \delta \mathbf{u}^T Q_{\mathbf{u}} + \delta \mathbf{x}^T Q_{\mathbf{x}} \\ &\frac{1}{2} \begin{bmatrix} \delta \mathbf{x}^T & \delta \mathbf{u}^T \end{bmatrix} \begin{bmatrix} Q_{xx} & Q_{xu} \\ Q_{ux} & Q_{uu} \end{bmatrix} \begin{bmatrix} \delta \mathbf{x} \\ \delta \mathbf{u} \end{bmatrix} \end{aligned} \quad (55)$$

By equating the coefficients with similar powers between the state action value function $Q(\mathbf{x}_k, \mathbf{u}_k)$ and the immediate reward and cost to go $\ell(\mathbf{x}_k, \mathbf{u}_k)$ and $V(\mathbf{x}_{k+1})$ respectively we can show that:

$$Q_{\mathbf{x}} = \ell_x + A_t \nabla_x V \quad (56)$$

$$Q_{\mathbf{u}} = \ell_u + A_t \nabla_x V$$

$$Q_{\mathbf{xx}} = \ell_{xx} + A_t^T V_{xx} A_t + \mathcal{F} + \tilde{\mathbf{M}}$$

$$Q_{\mathbf{xu}} = \ell_{xu} + A_t^T V_{xu} B_t + \mathcal{L} + \tilde{\mathbf{N}}$$

$$Q_{\mathbf{uu}} = \ell_{uu} + B_t^T V_{uu} B_t + \mathcal{Z} + \tilde{\mathbf{G}}$$

where we have assume a local quadratic approximation of the immediate reward $\ell(\mathbf{x}_k, \mathbf{u}_k)$ according to the equation:

$$\begin{aligned} \ell(\bar{\mathbf{x}} + \delta \mathbf{x}, \bar{\mathbf{u}} + \delta \mathbf{u}) &= \ell_0 + \delta \mathbf{u}^T \ell_{\mathbf{u}} + \delta \mathbf{x}^T \ell_{\mathbf{x}} \\ &\frac{1}{2} \begin{bmatrix} \delta \mathbf{x}^T & \delta \mathbf{u}^T \end{bmatrix} \begin{bmatrix} \ell_{xx} & \ell_{xu} \\ \ell_{ux} & \ell_{uu} \end{bmatrix} \begin{bmatrix} \delta \mathbf{x} \\ \delta \mathbf{u} \end{bmatrix} \end{aligned} \quad (57)$$

The local variations in control $\delta \mathbf{u}^*$ that maximize the state, action value function are expressed by the equation that follows:

$$\begin{aligned} \delta \mathbf{u}^* &= \underset{\mathbf{u}}{\text{argmax}} Q(\bar{\mathbf{x}} + \delta \mathbf{x}, \bar{\mathbf{u}} + \delta \mathbf{u}) \\ &= -Q_{\mathbf{uu}}^{-1} (Q_{\mathbf{u}} + Q_{\mathbf{ux}} \delta \mathbf{x}) \end{aligned} \quad (58)$$

The optimal control variations have the form $\delta \mathbf{u}^* = \mathbf{l} + \mathbf{L} \delta \mathbf{x}$ where $\mathbf{l} = -Q_{\mathbf{uu}}^{-1} Q_{\mathbf{u}}$ is the open loop gain and $\mathbf{L} = -Q_{\mathbf{uu}}^{-1} Q_{\mathbf{ux}}$ is the closed loop - feedback gain.

For the special cases where the stochastic dynamics have only additive noise $F(\mathbf{u}, \mathbf{x}) = F$ then the terms $\tilde{\mathbf{M}}, \tilde{\mathbf{N}}, \tilde{\mathbf{G}}$ will be zero since they are functions of $\nabla_{\mathbf{xx}} F$ and $\nabla_{\mathbf{xu}} F$

and $\nabla_{\mathbf{u}\mathbf{u}}F$ and it holds that $\nabla_{\mathbf{x}\mathbf{x}}F = 0$, $\nabla_{\mathbf{x}\mathbf{u}}F = 0$ and $\nabla_{\mathbf{u}\mathbf{u}}F = 0$. In such a type of systems the control does not depend on the statistical characteristics of the noise. In cases of deterministic systems again $\tilde{\mathbf{M}}, \tilde{\mathbf{N}}, \tilde{\mathbf{G}}$ will be zero because these terms depend on the variance of the noise $\sigma_{d\omega_i} = 0$, $\forall i = 1, \dots, m$. Finally if the noise is only control depended then $\tilde{\mathbf{M}}, \tilde{\mathbf{N}}$ will be zero since $\nabla_{\mathbf{x}\mathbf{x}}F(\mathbf{u}) = 0$ and $\nabla_{\mathbf{x}\mathbf{u}}F(\mathbf{u}) = 0$ while if it is state dependent then $\tilde{\mathbf{N}}, \tilde{\mathbf{G}}$ will be zero since $\nabla_{\mathbf{x}\mathbf{u}}F(\mathbf{x}) = 0$ and $\nabla_{\mathbf{u}\mathbf{u}}F(\mathbf{x}) = 0$.

V. DISCUSSION

In this paper we explicitly derived the equations describing the second order expansion of the cost-to-go, given state and control dependent noise. Our main result is that the expressions remain quadratic WRT $\delta\mathbf{x}$ and $\delta\mathbf{u}$, so the basic structure of the algorithm, a quadratic cost-to-go approximation with a linear policy, remains unchanged. In addition we have shown how the cases of deterministic and stochastic DDP with additive noise are sub-cases of our generalized formulation of Stochastic DDP.

Current and future research includes the testing and evaluation of our generalized Stochastic DDP algorithm on stochastic systems which are highly nonlinear and have noise that is control and state dependent. Biomechanical models belong to this class due to highly nonlinear and noisy nature of muscle dynamics. Moreover we are aiming to incorporate a second order Extended Kalman Filter that can handle observation as well as process noise. The resulting optimal controller-estimator scheme will be a version of iterative Quadratic Gaussian regulator that can handle stochastic dynamics expanded up to the second order with state and control dependent noise.

REFERENCES

- [1] D. H. Jacobson and D. Q. Mayne, *Differential Dynamic Programming*. Elsevier, 1970.
- [2] C. E. Garcia, D. M. Prett, and M. Morari, "Model predictive control: theory and practice," *Automatica*, vol. 25, pp. 335–348, 1989.
- [3] P. Abbeel, A. Coates, M. Quigley, and A. Y. Ng, "An application of reinforcement learning to aerobatic helicopter flight," in *Advances in Neural Information Processing Systems 19*, 2007.
- [4] E. Todorov and W. Li, "A generalized iterative LQG method for locally-optimal feedback control of constrained nonlinear stochastic systems," in *proceedings of the American Control Conference*, 2005, pp. 300–306.