American Control Conference 2010

Stochastic Differential Dynamic Programming

Evangelos Theodorou, Yuval Tassa & Emo Todorov

***Abstract*— Although there has been a significant amount of work in the area of stochastic optimal control theory towards the development of new algorithms, the problem of how to control a stochastic nonlinear system remains an open research topic. Recent iterative linear quadratic optimal control methods iLQG [1], [2] handle control and state multiplicative noise while they are derived based on first order approximation of dynamics. On the other hand, methods such as Differential Dynamic Programming expand the dynamics up to the second order but so far they can handle nonlinear systems with additive noise.**

In this work we present a generalization of the classic Differential Dynamic Programming algorithm. We assume the existence of state and control multiplicative 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 that minimize future costs under the constraint of state space dynamics. In the continuous nonlinear case, *local methods* are one of the very few classes 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 [3], computes a quadratic approximation of the cost-to-go and correspondingly, a local linear-feedback controller. The state space dynamics are also quadratically approximated around a trajectory. As in the 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.

Past work on nonlinear optimal control allows the use of DDP in problems with state and control constraints [4],[5]. In this work, state and control constraints are expanded up to the first order and the KKT conditions are formulated resulting in a unconstrained quadratic optimization problem. The application of DDP in real robotic high dimensional control tasks created the need for extensions that relax the condition for accurate models. In this vain, in [6] DDP is extended for the case of min - max formulation. The goal for this formulation is to make DDP robust against the model uncertainties and hybrid dynamics in a biped robotic locomotion task. In [7] implementation improvements regarding the evaluation of the value function allow the use of DDP in a receding horizon mode. In all the past work related to DDP, the optimal control problem is considered to be 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 [1]. This latter case was addressed within the iterative-LQG framework [1], in which the optimal controller is derived based on the first order approximation of the dynamics and the second order approximation of the cost to go. However, for the iterative nonlinear optimal control algorithms such as DDP in which second order expansion of the dynamics is considered, the more general case of state and/or control multiplicative noise appears to have never been tackled.

In this paper, we derive the DDP algorithm for state and control multiplicative process noise. We find that despite the potential of cubic and quartic terms, these cancel out, allowing us to maintain the quadratic form of the approximation. Moreover we show how the new generalized formulation of Stochastic Differential Dynamic Programming (SDDP) recovers the standard DDP deterministic solution as well as the special cases in which only state multiplicative or control multiplicative noise is considered.

The remaining of this work is organized as follows: in the next section we provide the definition of the SDDP. In section III, the second order expansion of the cost to go is presented and in section IV the optimal controls are derived and the overall SDDP algorithm is presented. In addition in section IV we show how SDDP recovers the deterministic solution as well as the cases of only control multiplicative, only state multiplicative and only additive noise. Finally in sections V and VI simulation results and conclusions are discussed.

II. Stochastic Differential Dynamic Programming

We consider the class of nonlinear stochastic optimal control problems with cost

$$v^{\pi}(\mathbf{x}, t) = E\left[h(\mathbf{x}(T)) + \int_{t_0}^{T} \ell\left(\tau, \mathbf{x}(\tau), \pi(\tau, \mathbf{x}(\tau))\right) d\tau\right] \quad (1)$$

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

subject to the stochastic dynamics of the form:

$$d\mathbf{x} = f(\mathbf{x}, \mathbf{u})dt + F(\mathbf{x}, \mathbf{u})d\omega \tag{2}$$

where $\mathbf{x} \in \Re^{n\times 1}$ is the state, $\mathbf{u} \in \Re^{p\times 1}$ is the control and $d\omega \in \Re^{m\times 1}$ is brownian noise. The term $h(\mathbf{x}(T))$ in the cost function (1), is the terminal cost while the $\ell\left(\tau, \mathbf{x}(\tau), \pi(\tau, \mathbf{x}(\tau))\right)$ is the instantaneous cost rate which is a function of the state $\mathbf{x}$ and control policy $\pi(\tau, \mathbf{x}(\tau))$. The cost-to - go $v^{\pi}(\mathbf{x}, t)$ is defined as the expected cost accumulated over the time horizon $(t_0, ..., T)$ starting from the initial state $\mathbf{x}_t$ to the final state $\mathbf{x}(T)$.

To enhance the readability of our derivations we write the dynamics as a function $\Phi \in \Re^{n\times 1}$ of the state, control and instantiation of the noise:

$$\Phi(\mathbf{x}, \mathbf{u}, d\omega) \equiv f(\mathbf{x}, \mathbf{u})dt + F(\mathbf{x}, \mathbf{u})d\omega \tag{3}$$

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

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

Every element of the vector $\Phi(\mathbf{x}, \mathbf{u}, d\omega) \in \Re^{n\times 1}$ 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:

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

where $\mathbf{O}(\delta\mathbf{x}, \delta\mathbf{u}, d\omega) \in \Re^{n\times 1}$ contains all the second order terms in the deviations in states, controls and noise[1]. 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 $O^{(j)}(\delta\mathbf{x}, \delta\mathbf{u}, d\omega) \in \Re$ as:

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

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:

$$\nabla_{\mathbf{x}}\Phi = \nabla_{\mathbf{x}}f(\mathbf{x}, \mathbf{u})\delta t + \nabla_{\mathbf{x}}\left(\sum_{i=1}^{m} F_c^i \; d\omega_t^{(i)}\right)$$
$$\nabla_{\mathbf{u}}\Phi = \nabla_{\mathbf{u}}f(\mathbf{x}, \mathbf{u})\delta t + \nabla_{\mathbf{u}}\left(\sum_{i=1}^{m} F_c^i \; d\omega_t^{(i)}\right)$$

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

$$\nabla_{\mathbf{xx}}\Phi^{(j)} = \nabla_{\mathbf{xx}}f^{(j)}(\mathbf{x}, \mathbf{u})\delta t + \nabla_{\mathbf{xx}}\left(F_r^{(j)}(\mathbf{x}, \mathbf{u})d\omega_t\right)$$
$$\nabla_{\mathbf{uu}}\Phi^{(j)} = \nabla_{\mathbf{uu}}f^{(j)}(\mathbf{x}, \mathbf{u})\delta t + \nabla_{\mathbf{uu}}\left(F_r^{(j)}(\mathbf{x}, \mathbf{u})d\omega_t\right)$$
$$\nabla_{\mathbf{ux}}\Phi^{(j)} = \nabla_{\mathbf{ux}}f^{(j)}(\mathbf{x}, \mathbf{u})\delta t + \nabla_{\mathbf{ux}}\left(F_r^{(j)}(\mathbf{x}, \mathbf{u})d\omega_t\right)$$
$$\nabla_{\mathbf{xu}}\Phi^{(j)} = \left(\nabla_{\mathbf{ux}}\Phi^{(j)}\right)^{\mathsf{T}}$$

After expanding the dynamics up to the second order we can transition from continuous to discrete time. More precisely the discrete-time dynamics are formulated as:

$$\delta\mathbf{x}_{t+\delta t} =$$
$$\left(I_{n\times n} + \nabla_{\mathbf{x}}f(\mathbf{x}, \mathbf{u})\delta t + \nabla_{\mathbf{x}}\left(\sum_{i=1}^{m} F_c^{(i)} \; \xi_t^{(i)}\sqrt{\delta t}\right)\right)\delta\mathbf{x}_t$$
$$+ \left(\nabla_{\mathbf{u}}f(\mathbf{x}, \mathbf{u})\delta t + \nabla_{\mathbf{u}}\left(\sum_{i=1}^{m} F_c^{(i)} \; \boldsymbol{\xi}_t^{(i)}\sqrt{\delta t}\right)\right)\delta\mathbf{u}_t$$
$$+ F(\mathbf{x}, \mathbf{u})\sqrt{\delta t}\boldsymbol{\xi}_t + \mathbf{O}_d(\delta\mathbf{x}, \delta\mathbf{u}, \boldsymbol{\xi}, \delta t)$$

with $\delta t = t_{k+1} - t_k$ corresponding to a small discretization interval. Note that the term $\mathbf{O}_d$ is the equivalent of $\mathbf{O}$ but in discrete time and therefore it is now a function of δt. In fact, since $\mathbf{O}_d$ contains all the second order expansion terms of the dynamics it contains second order derivatives WRT state and control expressed as follows:

$$\nabla_{\mathbf{xx}}\Phi^{(j)} = \nabla_{\mathbf{xx}}f^{(j)}(\mathbf{x}, \mathbf{u})\delta t + \nabla_{\mathbf{xx}}\left(F_r^{(j)}(\mathbf{x}, \mathbf{u})\boldsymbol{\xi}_t\right)\sqrt{\delta t}$$
$$\nabla_{\mathbf{uu}}\Phi^{(j)} = \nabla_{\mathbf{uu}}f^{(j)}(\mathbf{x}, \mathbf{u})\delta t + \nabla_{\mathbf{uu}}\left(F_r^{(j)}(\mathbf{x}, \mathbf{u})\boldsymbol{\xi}_t\right)\sqrt{\delta t}$$
$$\nabla_{\mathbf{ux}}\Phi^{(j)} = \nabla_{\mathbf{ux}}f^{(j)}(\mathbf{x}, \mathbf{u})\delta t + \nabla_{\mathbf{ux}}\left(F_r^{(j)}(\mathbf{x}, \mathbf{u})\boldsymbol{\xi}_t\right)\sqrt{\delta t}$$
$$\nabla_{\mathbf{xu}}\Phi^{(j)} = \left(\nabla_{\mathbf{ux}}\Phi^{(j)}\right)^{\mathsf{T}}$$

The random variable $\boldsymbol{\xi} \in \Re^{m\times 1}$ is zero mean and Gaussian distributed with covariance $\Sigma = \sigma^2 I_{m\times m}$ 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:

$$\delta\mathbf{x}_{t+\delta t} =$$
$$A_t\delta\mathbf{x}_t + B_t\delta\mathbf{u}_t + \Gamma_t\boldsymbol{\xi}_t + \mathbf{O}_d(\delta\mathbf{x}, \delta\mathbf{u}, \boldsymbol{\xi}, \delta t) \tag{4}$$

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

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

with $\Gamma^{(i)} \in \Re^{n\times 1}$ 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 expresses 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:

[1] Not to be confused with “big-O”.

$$\Gamma_t = \Delta_t(\delta\mathbf{x}, \delta\mathbf{u}) + F(\mathbf{x}, \mathbf{u}) \tag{5}$$

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 $\bar{\mathbf{x}}$:

$$\begin{aligned} &V(\bar{\mathbf{x}} + \delta\mathbf{x}) = \\ &V(\bar{\mathbf{x}}) + V_{\mathbf{x}}^T \delta\mathbf{x} + \frac{1}{2}\delta\mathbf{x}^T V_{\mathbf{xx}} \delta\mathbf{x} \end{aligned} \tag{6}$$

Substitution of the discretized dynamics (4) in the second order Value function expansion (6) results in:

$$\begin{aligned} &V(\bar{\mathbf{x}}_{t+\delta t} + \delta\mathbf{x}_{t+\delta t}) = V(\bar{\mathbf{x}}_{t+\delta t}) \\ &+ V_{\mathbf{x}}^T \left(A_t \delta\mathbf{x}_t + B_t \delta\mathbf{u}_t + \Gamma_t \boldsymbol{\xi} + \mathbf{O}_d\right) \\ &+ \left(A_t \delta\mathbf{x}_t + B_t \delta\mathbf{u}_t + \Gamma_t \boldsymbol{\xi} + \mathbf{O}_d\right)^T \\ &\times V_{\mathbf{xx}} \left(A_t \delta\mathbf{x}_t + B_t \delta\mathbf{u}_t + \Gamma_t \boldsymbol{\xi} + \mathbf{O}_d\right) \end{aligned} \tag{7}$$

Next we will compute $E\left(V(\bar{\mathbf{x}}_{t+\delta t} + \delta\mathbf{x}_{t+\delta t})\right)$ 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:

$$E\left(V_{\mathbf{x}}^T \delta\mathbf{x}_{t+\delta t}\right) \tag{8}$$

and

$$E\left(\delta\mathbf{x}_{t+\delta t}^T V_{\mathbf{xx}} \delta\mathbf{x}_{t+\delta t}\right) \tag{9}$$

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

$$\delta\mathbf{x}_{t+\delta t} = A_t \delta\mathbf{x}_t + B_t \delta\mathbf{u}_t + \Gamma_t \boldsymbol{\xi} + \mathbf{O}_d \tag{10}$$

The analysis that follows in section III-A consist of the computation of the expectation of the four terms which result from the substitution of the linearized dynamics (10) into (8). In section III-B we compute the expectation of the 16 terms that result from the substitution of (10) into (9).

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

The expectation of the first order term results in:

$$\begin{aligned} E\left(V_{\mathbf{x}}^T \left(A_t \delta\mathbf{x}_t + B_t \delta\mathbf{u}_t + \Gamma_t \boldsymbol{\xi}_t + \mathbf{O}_d\right)\right) = \\ V_{\mathbf{x}}^T \left(A_t \delta\mathbf{x}_t + B_t \delta\mathbf{u}_t + E\left(\mathbf{O}_d\right)\right) \end{aligned} \tag{11}$$

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

$$\begin{aligned} &E\left(O^{(j)}(\delta\mathbf{x}, \delta\mathbf{u}, \boldsymbol{\xi}_t, \delta t)\right) = \\ &E\left(\frac{1}{2}\begin{pmatrix} \delta\mathbf{x} \\ \delta\mathbf{u} \end{pmatrix}^T \begin{pmatrix} \nabla_{\mathbf{xx}}\Phi^{(j)} & \nabla_{\mathbf{xu}}\Phi^{(j)} \\ \nabla_{\mathbf{ux}}\Phi^{(j)} & \nabla_{\mathbf{uu}}\Phi^{(j)} \end{pmatrix} \begin{pmatrix} \delta\mathbf{x} \\ \delta\mathbf{u} \end{pmatrix}\right) = \end{aligned} \tag{12}$$

$$= \frac{\delta t}{2}\begin{pmatrix} \delta\mathbf{x} \\ \delta\mathbf{u} \end{pmatrix}^T \begin{pmatrix} \nabla_{\mathbf{xx}} f^{(j)} & \nabla_{\mathbf{xu}} f^{(j)} \\ \nabla_{\mathbf{ux}} f^{(j)} & \nabla_{\mathbf{uu}} f^{(j)} \end{pmatrix} \begin{pmatrix} \delta\mathbf{x} \\ \delta\mathbf{u} \end{pmatrix} = \tilde{O}^j$$

Therefore we will have that:

$$E\left(V_{\mathbf{x}}^T \delta\mathbf{x}_{t+\delta t}\right) = V_{\mathbf{x}}^T \left(A_t \delta\mathbf{x}_t + B_t \delta\mathbf{u}_t + \tilde{\mathbf{O}}_d\right) \tag{13}$$

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

$$\tilde{\mathbf{O}}_d(\delta\mathbf{x}, \delta\mathbf{u}, \delta t) = \begin{pmatrix} \tilde{O}^{(1)}(\delta\mathbf{x}, \delta\mathbf{u}, \delta t) \\ \cdots \\ \cdots \\ \tilde{O}^{(n)}(\delta\mathbf{x}, \delta\mathbf{u}, \delta t) \end{pmatrix} \tag{14}$$

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

$$V_{\mathbf{x}}^T \tilde{\mathbf{O}}_d = \frac{1}{2}\delta\mathbf{x}^T \boldsymbol{\mathcal{F}} \delta\mathbf{x} + \frac{1}{2}\delta\mathbf{u}^T \boldsymbol{\mathcal{Z}} \delta\mathbf{u} + \delta\mathbf{u}^T \boldsymbol{\mathcal{L}} \delta\mathbf{x} \tag{15}$$

with

$$\boldsymbol{\mathcal{F}} = \left(\sum_{j=1}^{n} \nabla_{\mathbf{xx}} f^{(j)} V_{x_j}\right) \tag{16}$$

$$\boldsymbol{\mathcal{Z}} = \left(\sum_{j=1}^{n} \nabla_{\mathbf{uu}} f^{(j)} V_{x_j}\right) \tag{17}$$

$$\boldsymbol{\mathcal{L}} = \left(\sum_{j=1}^{n} \nabla_{\mathbf{ux}} f^{(j)} V_{x_j}\right) \tag{18}$$

From the analysis above we can see that the expectation $\nabla_x V^T \delta\mathbf{x}_{t+\delta t}$ 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 $\delta\mathbf{x}_{t+\delta t}^T \nabla_{xx} V^T \delta\mathbf{x}_{t+\delta t}$.

In this section we compute all the terms that appear due to the expectation of the second approximation of the value function $E\left(\delta\mathbf{x}_{t+\delta t}^T \nabla_{xx} V \delta\mathbf{x}_{t+\delta t}\right)$. The term $\delta\mathbf{x}_{t+\delta t}$ is given by the stochastic dynamics in (10). Substitution of (10) results in 16 terms. To make our analysis clear we classify these 16 terms terms above into five classes. More precisely we will have that:

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

where the terms $\boldsymbol{\mathcal{E}}_1, \boldsymbol{\mathcal{E}}_2, \boldsymbol{\mathcal{E}}_3, \boldsymbol{\mathcal{E}}_4$ and $\boldsymbol{\mathcal{E}}_5$ are defined as follows:

$$\begin{aligned}\boldsymbol{\mathcal{E}}_1 \; &= E\left(\delta\mathbf{x}_t^T A_t^T V_{\mathbf{xx}} A_t \delta\mathbf{x}_t\right) + E\left(\delta\mathbf{u}_t^T B_t^T V_{\mathbf{xx}} B_t \delta\mathbf{u}_t\right) \\ &+ E\left(\delta\mathbf{x}_t^T A_t^T V_{\mathbf{xx}} B_t \delta\mathbf{u}_t\right) + E\left(\delta\mathbf{u}_t^T B_t^T V_{\mathbf{xx}} A_t \delta\mathbf{x}_t\right)\end{aligned} \quad (20)$$

$$\begin{aligned}\boldsymbol{\mathcal{E}}_2 \; &= \; E\left(\boldsymbol{\xi}_t^T \Gamma_t^T V_{\mathbf{xx}} A_t \delta\mathbf{x}\right) + E\left(\boldsymbol{\xi}_t^T \Gamma_t^T V_{\mathbf{xx}} B_t \delta\mathbf{u}\right) \\ &+ \; E\left(\delta\mathbf{x}^T A_t^T V_{\mathbf{xx}} \Gamma_t \boldsymbol{\xi}_t\right) + E\left(\delta\mathbf{u}^T B_t^T V_{\mathbf{xx}} \Gamma_t \boldsymbol{\xi}_t\right) \\ &+ \; E\left(\boldsymbol{\xi}_t^T \Gamma_t^T V_{\mathbf{xx}} \Gamma_t \boldsymbol{\xi}_t\right)\end{aligned} \quad (21)$$

$$\boldsymbol{\mathcal{E}}_3 \; = \; E\left(\mathbf{O}_d^T V_{\mathbf{xx}} \Gamma_t \boldsymbol{\xi}_t\right) + E\left(\boldsymbol{\xi}_t^T \Gamma_t^T V_{\mathbf{xx}} \mathbf{O}_d\right) \quad (22)$$

$$\begin{aligned}\boldsymbol{\mathcal{E}}_4 = E\left(\delta\mathbf{x}_t^T A_t^T V_{\mathbf{xx}} \mathbf{O}_d\right) + E\left(\delta\mathbf{u}_t^T B_t^T V_{\mathbf{xx}} \mathbf{O}_d\right) + \\ E\left(\mathbf{O}_d^T V_{\mathbf{xx}} B_t \delta\mathbf{u}_t\right) + E\left(\mathbf{O}_d^T V_{\mathbf{xx}} A_t \delta\mathbf{x}_t\right)\end{aligned} \quad (23)$$

$$\boldsymbol{\mathcal{E}}_5 = E\left(\mathbf{O}_d^T V_{\mathbf{xx}} \mathbf{O}_d\right) \quad (24)$$

In the first category we have all these terms that depend neither on $\boldsymbol{\xi}_t$ and nor on $\mathbf{O}_d(\delta\mathbf{x}, \delta\mathbf{u}, \boldsymbol{\xi}_t, \delta t)$. These are the terms that define $\boldsymbol{\mathcal{E}}_1$. The second category $\boldsymbol{\mathcal{E}}_2$ includes terms that depend on $\boldsymbol{\xi}_t$ but not on $\mathbf{O}_d(\delta\mathbf{x}, \delta\mathbf{u}, \boldsymbol{\xi}_t, \delta t)$. In the third class $\boldsymbol{\mathcal{E}}_3$, there are terms that depends both on $\mathbf{O}_d(\delta\mathbf{x}, \delta\mathbf{u}, \boldsymbol{\xi}_t, \delta t)$ and $\boldsymbol{\xi}_t$. In the fourth class $\boldsymbol{\mathcal{E}}_4$, we have terms that depend on $\mathbf{O}_d(\delta\mathbf{x}, \delta\mathbf{u}, \boldsymbol{\xi}_t, \delta t)$. Finally in the fifth class $\boldsymbol{\mathcal{E}}_5$, we have all these terms that depend on $\mathbf{O}_d(\delta\mathbf{x}, \delta\mathbf{u}, \boldsymbol{\xi}_t, \delta t)$ 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 compute the expectations of all the terms in the $\boldsymbol{\mathcal{E}}_1$ class. More precisely we will have that:

$$\begin{aligned} E\left(\delta\mathbf{x}_t^T A_t^T V_{\mathbf{xx}} A_t \delta\mathbf{x}_t\right) &= \delta\mathbf{x}_t^T A_t^T V_{\mathbf{xx}} A_t \delta\mathbf{x}_t \\ E\left(\delta\mathbf{u}_t^T B_t^T V_{\mathbf{xx}} B_t \delta\mathbf{u}_t\right) &= \delta\mathbf{u}_t^T B_t^T V_{\mathbf{xx}} B_t \delta\mathbf{u}_t \\ E\left(\delta\mathbf{x}_t^T A_t^T V_{\mathbf{xx}} B_t \delta\mathbf{u}_t\right) &= \delta\mathbf{x}_t^T A_t^T V_{\mathbf{xx}} B_t \delta\mathbf{u}_t \\ E\left(\delta\mathbf{u}_t^T B_t^T V_{\mathbf{xx}} A_t \delta\mathbf{x}_t\right) &= \delta\mathbf{u}_t^T B_t^T V_{\mathbf{xx}} A_t \delta\mathbf{x}_t \end{aligned} \quad (25)$$

We continue our analysis by calculating all the terms in the class $\boldsymbol{\mathcal{E}}_2$. More presicely we will have:

$$\begin{aligned} E\left(\boldsymbol{\xi}_t^T \Gamma_t^T V_{\mathbf{xx}} A_t \delta\mathbf{x}\right) &= 0 \\ E\left(\boldsymbol{\xi}_t^T \Gamma_t^T V_{\mathbf{xx}} B_t \delta\mathbf{u}\right) &= 0 \\ E\left(\boldsymbol{\xi}_t^T \Gamma_t^T V_{\mathbf{xx}} A_t \delta\mathbf{x}\right)^T &= 0 \\ E\left(\boldsymbol{\xi}_t^T \Gamma_t^T V_{\mathbf{xx}} B_t \delta\mathbf{u}\right)^T &= 0 \end{aligned} \quad (26)$$

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

$$E\left(\boldsymbol{\xi}_t^T \Gamma_t^T V_{\mathbf{xx}} \Gamma_t \boldsymbol{\xi}_t\right) = \textbf{trace}\left(\Gamma_t^T V_{\mathbf{xx}} \Gamma_t \Sigma_\omega\right) \quad (27)$$

Since matrix Γ depends on variations in states and controls $\delta\mathbf{x}, \delta\mathbf{u}$ we can further massage the expressions above so that it can be expressed as quadratic functions in $\delta\mathbf{x}, \delta\mathbf{u}$.

$$\textbf{trace}\left(\Gamma_t^T V_{\mathbf{xx}} \Gamma_t \Sigma_\omega\right) = \sigma_{d\omega}^2 \delta t \quad (28)$$

$$\textbf{trace}\left(\left(\begin{array}{c} \Gamma^{(1)T} \\ \ldots \\ \ldots \\ \Gamma^{(m)T} \end{array}\right) V_{\mathbf{xx}} \left(\begin{array}{cccc} \Gamma^{(1)} & \ldots & \ldots & \Gamma^{(m)} \end{array}\right)\right) \quad (29)$$

$$= \sigma_{d\omega}^2 \delta t \sum_{i=1}^m \Gamma^{(i)T} V_{\mathbf{xx}} \Gamma^{(i)} \quad (30)$$

The last equation is written in the form:

$$\begin{aligned}\textbf{trace}\left(\Gamma_t^T V_{\mathbf{xx}} \Gamma_t \Sigma_\omega\right) = \delta\mathbf{x}^T \tilde{\boldsymbol{\mathcal{F}}} \delta\mathbf{x} + 2\delta\mathbf{x}^T \tilde{\boldsymbol{\mathcal{L}}} \delta\mathbf{u} + \delta\mathbf{u}^T \tilde{\boldsymbol{\mathcal{Z}}} \delta\mathbf{u} \\ + 2\delta\mathbf{u}^T \tilde{\boldsymbol{\mathcal{U}}} + 2\delta\mathbf{x}^T \tilde{\boldsymbol{\mathcal{S}}} + \gamma\end{aligned} \quad (31)$$

Where the terms $\tilde{\boldsymbol{\mathcal{F}}} \in \Re^{n\times m}, \tilde{\boldsymbol{\mathcal{L}}} \in \Re^{n\times p}, \tilde{\boldsymbol{\mathcal{Z}}} \in \Re^{p\times p}, \tilde{\boldsymbol{\mathcal{U}}} \in \Re^{p\times 1}, \tilde{\boldsymbol{\mathcal{S}}} \in \Re^{n\times 1}$ and $\gamma \in \Re$ are defined as follows:

$$\tilde{\boldsymbol{\mathcal{F}}} = \sigma^2 \delta t \sum_{i=1}^m \nabla_{\mathbf{x}} F_c^{(i)T} V_{\mathbf{xx}} \nabla_{\mathbf{x}} F_c^{(i)} \quad (32)$$

$$\tilde{\boldsymbol{\mathcal{L}}} = \sigma^2 \delta t \sum_{i=1}^m \nabla_{\mathbf{x}} F_c^{(i)T} V_{\mathbf{xx}} \nabla_{\mathbf{u}} F_c^{(i)} \quad (33)$$

$$\tilde{\boldsymbol{\mathcal{Z}}} = \sigma^2 \delta t \sum_{i=1}^m \nabla_{\mathbf{u}} F_c^{(i)T} V_{\mathbf{xx}} \nabla_{\mathbf{u}} F_c^{(i)} \quad (34)$$

$$\tilde{\boldsymbol{\mathcal{U}}} = \sigma^2 \delta t \sum_{i=1}^m \nabla_{\mathbf{u}} F_c^{(i)T} V_{\mathbf{xx}} F_c^{(i)} \quad (35)$$

$$\tilde{\boldsymbol{\mathcal{S}}} = \sigma^2 \delta t \sum_{i=1}^m \nabla_{\mathbf{x}} F_c^{(i)T} V_{\mathbf{xx}} F_c^{(i)} \quad (36)$$

$$\gamma = \sigma^2 \delta t \sum_{i=1}^m F_c^{(i)T} V_{\mathbf{xx}} F_c^{(i)} \quad (37)$$

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

$$\begin{aligned} E\left(\mathbf{O}_d^T V_{\mathbf{xx}} \Gamma_t \boldsymbol{\xi}_t\right) &= E\left(\textbf{trace}\left(V_{\mathbf{xx}} \Gamma_t \boldsymbol{\xi}_t \mathbf{O}_d^T\right)\right) \\ &= \textbf{trace}\left(V_{\mathbf{xx}} \Gamma_t E\left(\boldsymbol{\xi}_t \mathbf{O}_d^T\right)\right) \end{aligned} \quad (38)$$

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

$$\begin{aligned} &E\left(\mathbf{O}_d^T V_{\mathbf{xx}} \Gamma_t \boldsymbol{\xi}_t\right) = \quad (39) \\ &\textbf{trace}\left(V_{\mathbf{xx}} \Gamma_t E\left[\begin{array}{ccc} \boldsymbol{\xi}_t O^{(1)} & \ldots & \boldsymbol{\xi}_t O^{(n)} \end{array}\right]\right) \end{aligned}$$

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

$$E\left(\sqrt{\delta t}\boldsymbol{\xi}_t O^{(j)}\right) = \tag{40}$$
$$\frac{1}{2}E\left(\sqrt{\delta t}\boldsymbol{\xi}_t \delta\mathbf{x}^T \Phi^{(i)}_{\mathbf{xx}}\delta\mathbf{x}\right) + \frac{1}{2}E\left(\sqrt{\delta t}\boldsymbol{\xi}_t \delta\mathbf{u}^T \Phi^{(i)}_{\mathbf{uu}}\delta\mathbf{u}\right) +$$
$$E\left(\sqrt{\delta t}\boldsymbol{\xi}_t \delta\mathbf{u}^T \Phi^{(i)}_{\mathbf{ux}}\delta\mathbf{x}\right)$$

We first calculate the term:

$$E\left(\sqrt{\delta t}\boldsymbol{\xi}_t \delta\mathbf{x}^T \nabla_{\mathbf{xx}}\Phi^{(i)}\delta\mathbf{x}\right) \tag{41}$$
$$= E\left(\sqrt{\delta t}\boldsymbol{\xi}_t \delta\mathbf{x}^T \left(\nabla_{\mathbf{xx}}f^{(i)}\delta t + \nabla_{\mathbf{xx}}F_r^{(i)}\boldsymbol{\xi}_t\sqrt{\delta t}\right)\delta\mathbf{x}\right)$$
$$= E\left(\sqrt{\delta t}\boldsymbol{\xi}_t \delta\mathbf{x}^T \left(\nabla_{\mathbf{xx}}f^{(i)}\delta t\right)\delta\mathbf{x}\right)$$
$$+ E\left(\sqrt{\delta t}\boldsymbol{\xi}_t \delta\mathbf{x}^T \left(\nabla_{\mathbf{xx}}F_r^{(i)}\boldsymbol{\xi}_t\sqrt{\delta t}\right)\delta\mathbf{x}\right)$$

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

$$E\left(\sqrt{\delta t}\boldsymbol{\xi}_t \delta\mathbf{x}^T \nabla_{\mathbf{xx}}\Phi^{(i)}\delta\mathbf{x}\right) = \tag{42}$$
$$E\left(\sqrt{\delta t}\boldsymbol{\xi}_t \delta\mathbf{x}^T \nabla_{\mathbf{xx}}\left(F_r^{(i)}\boldsymbol{\xi}_t\sqrt{\delta t}\right)\delta\mathbf{x}\right)$$

Since the $\boldsymbol{\xi}_t = \left(\xi^{(1)},...,\xi^{(m)}\right)^T$ and $F_r^{(i)} = \left(F^{(i1)},....,F^{(im)}\right)$ we will have that:

$$E\left(\sqrt{\delta t}\boldsymbol{\xi}_t \delta\mathbf{x}^T \nabla_{\mathbf{xx}}\Phi^{(i)}\delta\mathbf{x}\right) = \tag{43}$$
$$E\left(\delta t\boldsymbol{\xi}_t \delta\mathbf{x}^T \nabla_{\mathbf{xx}}\left(\sum_{j=1}^{m} F^{(ij)}\xi^{(j)}\right)\delta\mathbf{x}\right)$$
$$E\left(\delta t\boldsymbol{\xi}_t \delta\mathbf{x}^T \left(\sum_{j=1}^{m} \nabla_{\mathbf{xx}}\left(F^{(ij)}\xi^{(j)}\right)\right)\delta\mathbf{x}\right)$$
$$E\left(\delta t\boldsymbol{\xi}_t \delta\mathbf{x}^T \left(\sum_{j=1}^{m} \xi^{(j)}\nabla_{\mathbf{xx}}\left(F^{(ij)}\right)\right)\delta\mathbf{x}\right)$$

By writing $\boldsymbol{\xi}_t$ in vector form we have that:

$$E\left(\sqrt{\delta t}\boldsymbol{\xi}_t \delta\mathbf{x}^T \nabla_{\mathbf{xx}}\Phi^{(i)}\delta\mathbf{x}\right) = \tag{44}$$
$$E\left(\delta t\begin{bmatrix}\xi^{(1)}\\ \cdots\\ \cdots\\ \xi^{(m)}\end{bmatrix}\delta\mathbf{x}^T\left(\sum_{j=1}^{m}\xi^{(j)}\nabla_{\mathbf{xx}}\left(F^{(ij)}\right)\right)\delta\mathbf{x}\right)$$

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

$$\begin{bmatrix}\delta t E\left(\xi^{(1)}\delta\mathbf{x}^T\left(\sum_{j=1}^{m}\xi^{(j)}\nabla_{\mathbf{xx}}\left(F^{(ij)}\right)\right)\delta\mathbf{x}\right)\\ \cdots\\ \cdots\\ \delta t E\left(\xi^{(m)}\delta\mathbf{x}^T\left(\sum_{j=1}^{m}\xi^{(j)}\nabla_{\mathbf{xx}}\left(F^{(ij)}\right)\right)\delta\mathbf{x}\right)\end{bmatrix} \tag{45}$$

Since $E\left(\xi^{(i)}\xi^{(i)}\right) = \sigma^2$ and $E\left(\xi^{(i)}\xi^{(j)}\right) = 0$ we can show that:

$$E\left(\sqrt{\delta t}\boldsymbol{\xi}_t \delta\mathbf{x}^T \nabla_{\mathbf{xx}}\Phi^{(i)}\delta\mathbf{x}\right) = \tag{46}$$
$$\sigma^2\delta t\begin{bmatrix}\delta\mathbf{x}^T\nabla_{\mathbf{xx}}F_r^{(i1)}\delta\mathbf{x}\\ \cdots\\ \cdots\\ \delta\mathbf{x}^T\nabla_{\mathbf{xx}}F_r^{(im)}\delta\mathbf{x}\end{bmatrix}$$

In a similar way we can show that:

$$E\left(\sqrt{\delta t}\boldsymbol{\xi}_t \delta\mathbf{x}^T \nabla_{\mathbf{uu}}\Phi^{(i)}\delta\mathbf{x}\right) = \tag{47}$$
$$\sigma^2\delta t\begin{bmatrix}\delta\mathbf{u}^T\nabla_{\mathbf{uu}}F_r^{(i1)}\delta\mathbf{u}\\ \cdots\\ \cdots\\ \delta\mathbf{u}^T\nabla_{\mathbf{uu}}F_r^{(im)}\delta\mathbf{u}\end{bmatrix}$$

and

$$E\left(\sqrt{\delta t}\boldsymbol{\xi}_t \delta\mathbf{u}^T \nabla_{\mathbf{xu}}\Phi^{(i)}\delta\mathbf{x}\right) = \tag{48}$$
$$\sigma^2\delta t\begin{bmatrix}\delta\mathbf{u}^T\nabla_{\mathbf{ux}}F_r^{(i1)}\delta\mathbf{x}\\ \cdots\\ \cdots\\ \delta\mathbf{u}^T\nabla_{\mathbf{ux}}F_r^{(im)}\delta\mathbf{x}\end{bmatrix}$$

Since we have calculated all the terms of expression (41) we can proceed with the computation of (38). According to the analysis above the term $E\left(\mathbf{O}_d^T V_{\mathbf{xx}}\Gamma_t\boldsymbol{\xi}_t\right)$ can be written as follows:

$$E\left(\mathbf{O}_d^T V_{\mathbf{xx}}\Gamma_t\boldsymbol{\xi}_t\right) = \tag{49}$$
$$\mathbf{trace}\left(V_{\mathbf{xx}}\Gamma_t\left(\boldsymbol{\mathcal{M}}+\boldsymbol{\mathcal{N}}+\boldsymbol{\mathcal{G}}\right)\right)$$

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

$$\boldsymbol{\mathcal{M}} = \tag{50}$$
$$\sigma^2\delta t\begin{bmatrix}\delta\mathbf{x}^T\nabla_{\mathbf{xx}}F_r^{(11)}\delta\mathbf{x} & ... & \delta\mathbf{x}^T\nabla_{\mathbf{xx}}F_r^{(1n)}\delta\mathbf{x}\\ ... & ... & ...\\ \delta\mathbf{x}^T\nabla_{\mathbf{xx}}F_r^{(m1)}\delta\mathbf{x} & ... & \delta\mathbf{x}^T\nabla_{\mathbf{xx}}F_r^{(mn)}\delta\mathbf{x}\end{bmatrix}$$

Similarly

$$\boldsymbol{\mathcal{N}} = \tag{51}$$
$$\sigma^2\delta t\begin{bmatrix}\delta\mathbf{x}^T\nabla_{\mathbf{xu}}F_r^{(1,1)}\delta\mathbf{u} & ... & \delta\mathbf{x}^T\nabla_{\mathbf{xu}}F_r^{(1,n)}\delta\mathbf{u}\\ ... & ... & ...\\ \delta\mathbf{x}^T\nabla_{\mathbf{xu}}F_r^{(m,1)}\delta\mathbf{u} & ... & \delta\mathbf{x}^T\nabla_{\mathbf{xu}}F_r^{(m,n)}\delta\mathbf{u}\end{bmatrix}$$

and

$$\boldsymbol{\mathcal{G}} = \qquad (52)$$
$$\sigma^2\delta t \begin{bmatrix} \delta\mathbf{u}^T\nabla_{\mathbf{uu}}F_r^{(1,1)}\delta\mathbf{u} & ... & \delta\mathbf{u}^T\nabla_{\mathbf{uu}}F_r^{(1,n)}\delta\mathbf{u} \\ ... & ... & ... \\ \delta\mathbf{u}^T\nabla_{\mathbf{uu}}F_r^{(m,1)}\delta\mathbf{u} & ... & \delta\mathbf{u}^T\nabla_{\mathbf{uu}}F_r^{(m,n)}\delta\mathbf{u} \end{bmatrix}$$

Based on (5) 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 $\boldsymbol{\mathcal{M}}$, $\boldsymbol{\mathcal{N}}$ and $\boldsymbol{\mathcal{G}}$ 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 $\boldsymbol{\mathcal{M}}$, $\boldsymbol{\mathcal{N}}$ and $\boldsymbol{\mathcal{G}}$ will result into 3th order terms that can be neglected. By neglecting these terms we can show that:

$$\begin{aligned} &E\left(\mathbf{O}_d^T V_{\mathbf{xx}}\Gamma_t\boldsymbol{\xi}_t\right) = \qquad (53)\\ &= \mathbf{trace}\left(V_{\mathbf{xx}}(\Delta+F)\left(\boldsymbol{\mathcal{M}}+\boldsymbol{\mathcal{N}}+\boldsymbol{\mathcal{G}}\right)\right)\\ &= \mathbf{trace}\left(V_{\mathbf{xx}}F\left(\boldsymbol{\mathcal{M}}+\boldsymbol{\mathcal{N}}+\boldsymbol{\mathcal{G}}\right)\right) \end{aligned}$$

Each element (i,j) of the product $\boldsymbol{\mathcal{C}} = V_{\mathbf{xx}}F$ can be expressed as $\boldsymbol{\mathcal{C}}^{(i,j)} = \sum_{r=1}^{n} V_{\mathbf{xx}}^{(i,r)}F^{(r,j)}$ where $\boldsymbol{\mathcal{C}} \in \Re^{n\times p}$. Furthermore the element (μ,ν) of the product $\boldsymbol{\mathcal{H}} = \boldsymbol{\mathcal{C}}\boldsymbol{\mathcal{M}}$ is formulated $\boldsymbol{\mathcal{H}}^{(\mu,\nu)} = \sum_{k=1}^{n}\boldsymbol{\mathcal{C}}^{(\mu,k)}\boldsymbol{\mathcal{M}}^{(k,\nu)}$ with $\boldsymbol{\mathcal{H}} \in \Re^{n\times n}$. Thus, the term $\mathbf{trace}\left(V_{\mathbf{xx}}F\boldsymbol{\mathcal{M}}\right)$ can be now expressed as:

$$\begin{aligned} \mathbf{trace}\left(V_{\mathbf{xx}}F\boldsymbol{\mathcal{M}}\right) &= \sum_{\ell=1}^{n}\boldsymbol{\mathcal{H}}^{(\ell,\ell)} \qquad (54)\\ &= \sum_{\ell=1}^{n}\sum_{k=1}^{m}\boldsymbol{\mathcal{C}}^{(\ell,k)}\boldsymbol{\mathcal{M}}^{(k,\ell)}\\ &= \sum_{\ell=1}^{n}\sum_{k=1}^{m}\left(\sum_{r=1}^{n}V_{\mathbf{xx}}^{(k,r)}F^{(r,\ell)}\right)\boldsymbol{\mathcal{M}}^{(k,\ell)} \end{aligned}$$

Since $\boldsymbol{\mathcal{M}}^{(k,\ell)} = \delta t\sigma_{d\omega_1}^2\delta\mathbf{x}^T\nabla_{\mathbf{xx}}F^{(k,\ell)}\delta\mathbf{x}$ 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:

$$\begin{aligned} &\mathbf{trace}\left(V_{\mathbf{xx}}F\boldsymbol{\mathcal{M}}\right) \qquad (55)\\ &= \sum_{\ell=1}^{n}\sum_{k=1}^{m}\left(\left(\sum_{r=1}^{n}V_{\mathbf{xx}}^{(k,r)}F^{(r,\ell)}\right)\sigma^2\delta t\delta\mathbf{x}^T\nabla_{\mathbf{xx}}F^{(k,\ell)}\delta\mathbf{x}\right)\\ &= \delta\mathbf{x}^T\sigma^2\delta t\sum_{\ell=1}^{n}\sum_{k=1}^{m}\left(\left(\sum_{r=1}^{n}V_{\mathbf{xx}}^{(k,r)}F^{(r,\ell)}\right)\nabla_{\mathbf{xx}}F^{(k,\ell)}\right)\delta\mathbf{x}\\ &= \delta\mathbf{x}^T\tilde{\mathbf{M}}\delta\mathbf{x} \end{aligned}$$

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

$$\tilde{\mathbf{M}} = \sigma^2\delta t\sum_{\ell=1}^{n}\sum_{k=1}^{m}\left(\left(\sum_{r=1}^{n}V_{\mathbf{xx}}^{(k,r)}F^{(r,\ell)}\right)\nabla_{\mathbf{xx}}F^{(k,\ell)}\right) \qquad (56)$$

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

$$\mathbf{trace}\left(V_{\mathbf{xx}}F\boldsymbol{\mathcal{N}}\right) = \delta\mathbf{x}^T\tilde{\mathbf{N}}\delta\mathbf{u} \qquad (57)$$

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

$$\tilde{\mathbf{N}} = \sigma^2\delta t\sum_{\ell=1}^{n}\sum_{k=1}^{m}\left(\left(\sum_{r=1}^{n}V_{\mathbf{xx}}^{(k,r)}F^{(r,\ell)}\right)\nabla_{\mathbf{xu}}F^{(k,\ell)}\right) \qquad (58)$$

and

$$\mathbf{trace}\left(V_{\mathbf{xx}}F\boldsymbol{\mathcal{G}}\right) = \delta\mathbf{u}^T\tilde{\mathbf{G}}\delta\mathbf{u} \qquad (59)$$

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

$$\tilde{\mathbf{G}} = \sigma^2\delta t\sum_{\ell=1}^{n}\sum_{k=1}^{m}\left(\left(\sum_{r=1}^{n}V_{\mathbf{xx}}^{(k,r)}F^{(r,\ell)}\right)\nabla_{\mathbf{uu}}F^{(k,\ell)}\right) \qquad (60)$$

Thus the term $E\left(\mathbf{O}_d^T V_{\mathbf{xx}}\Gamma_t\boldsymbol{\xi}_t\right)$ is formulated as:

$$E\left(\mathbf{O}_d^T V_{\mathbf{xx}}\Gamma_t\boldsymbol{\xi}_t\right) = \frac{1}{2}\delta\mathbf{x}^T\tilde{\mathbf{M}}\delta\mathbf{x} + \frac{1}{2}\delta\mathbf{u}^T\tilde{\mathbf{G}}\delta\mathbf{u} + \delta\mathbf{x}^T\tilde{\mathbf{N}}\delta\mathbf{u} \qquad (61)$$

Similarly we can show that:

$$E\left(\boldsymbol{\xi}_t^T\Gamma_t^T V_{\mathbf{xx}}\mathbf{O}_d\right) = \frac{1}{2}\delta\mathbf{x}^T\tilde{\mathbf{M}}\delta\mathbf{x} + \frac{1}{2}\delta\mathbf{u}^T\tilde{\mathbf{G}}\delta\mathbf{u} + \delta\mathbf{x}^T\tilde{\mathbf{N}}\delta\mathbf{u} \qquad (62)$$

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

$$\begin{aligned} E\left(\delta\mathbf{x}_t^T A_t^T V_{\mathbf{xx}}\mathbf{O}_d\right) &= \delta\mathbf{x}_t^T A_t^T V_{\mathbf{xx}}\tilde{\mathbf{O}}_d = 0\\ E\left(\delta\mathbf{u}_t^T B_t^T V_{\mathbf{xx}}\mathbf{O}_d\right) &= \delta\mathbf{u}_t^T B_t^T V_{\mathbf{xx}}\tilde{\mathbf{O}}_d = 0\\ E\left(\mathbf{O}_d^T V_{\mathbf{xx}}A_t\delta\mathbf{x}_t\right) &= \tilde{\mathbf{O}_d}^T V_{\mathbf{xx}}A_t\delta\mathbf{x}_t = 0\\ E\left(\mathbf{O}_d^T V_{\mathbf{xx}}B_t\delta\mathbf{u}_t\right) &= \tilde{\mathbf{O}_d}^T V_{\mathbf{xx}}B_t\delta\mathbf{u}_t = 0 \end{aligned} \qquad (63)$$

where the quantity $\tilde{\mathbf{O}_d}$ has been defined in (14). 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

$$\begin{aligned} \boldsymbol{\mathcal{E}}_5 &= E\left(\mathbf{O}_d^T V_{\mathbf{xx}}\mathbf{O}_d\right) \qquad (64)\\ &= E\left(\mathbf{trace}\left(V_{\mathbf{xx}}\mathbf{O}_d\mathbf{O}_d^T\right)\right)\\ &= \mathbf{trace}\left(V_{\mathbf{xx}}E\left(\mathbf{O}_d\mathbf{O}_d^T\right)\right)\\ &= \mathbf{trace}\left(V_{\mathbf{xx}}E\left(\begin{bmatrix} O^{(1)} \\ ... \\ O^{(n)} \end{bmatrix}\begin{bmatrix} O^{(1)} \\ ... \\ O^{(n)} \end{bmatrix}^T\right)\right) \end{aligned}$$

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 $\boldsymbol{\mathcal{E}}_5 = E\left(\mathbf{O}_d^T V_{\mathbf{xx}} \mathbf{O}_d\right)$ is equal to zero.

With the computation of the expectation of term that is quadratic WRT $\mathbf{O}_d$ we have calculated all the terms of the second order expansion of the cost to go function. In the next section we derive the optimal controls and we present the SDDP algorithm. Furthermore we show how SDDP recover the deterministic solution as well as the cases of only control multiplicative, only state multiplicative and only additive noise.

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 computed all the terms of expansion of the cost to go function $V(\mathbf{x}_t)$ at state$\mathbf{x}_t$ we show that its form remains quadratic WRT variations in state $\delta\mathbf{x}_t$ under the constraint of the nonlinear stochastic dynamics in (2). More precisely we have that:

$$\begin{aligned}
&V(\bar{\mathbf{x}}_{t+\delta t} + \delta\mathbf{x}_{t+\delta t}) = V(\bar{\mathbf{x}}_{t+\delta t})\\
&+ \nabla_x V^T A_t \delta\mathbf{x}_t + \nabla_x V^T B_t \delta\mathbf{u}_t\\
&+ \frac{1}{2}\delta\mathbf{x}^T \boldsymbol{\mathcal{F}} \delta\mathbf{x} + \frac{1}{2}\delta\mathbf{u}^T \boldsymbol{\mathcal{Z}} \delta\mathbf{u} + \delta\mathbf{u}^T \boldsymbol{\mathcal{L}} \delta\mathbf{x}\\
&+ \frac{1}{2}\delta\mathbf{x}_t^T A_t^T V_{\mathbf{xx}} A_t \delta\mathbf{x}_t + \frac{1}{2}\delta\mathbf{u}_t^T B_t^T V_{\mathbf{xx}} B_t \delta\mathbf{u}_t\\
&+ \frac{1}{2}\delta\mathbf{x}_t^T A_t^T V_{\mathbf{xx}} B_t \delta\mathbf{u}_t + \frac{1}{2}\delta\mathbf{u}_t^T B_t^T V_{\mathbf{xx}} A_t \delta\mathbf{x}_t\\
&+ \frac{1}{2}\delta\mathbf{x}^T \tilde{\boldsymbol{\mathcal{F}}} \delta\mathbf{x} + \delta\mathbf{x}^T \tilde{\boldsymbol{\mathcal{L}}} \delta\mathbf{u} + \frac{1}{2}\delta\mathbf{u}^T \tilde{\boldsymbol{\mathcal{Z}}} \delta\mathbf{u}\\
&+ \delta\mathbf{u}^T \tilde{\boldsymbol{\mathcal{U}}} + \delta\mathbf{x}^T \tilde{\boldsymbol{\mathcal{S}}} + \frac{1}{2}\gamma\\
&+ \frac{1}{2}\delta\mathbf{x}^T \tilde{\mathbf{M}} \delta\mathbf{x} + \frac{1}{2}\delta\mathbf{u}^T \tilde{\mathbf{G}} \delta\mathbf{u} + \delta\mathbf{x}^T \tilde{\mathbf{N}} \delta\mathbf{u}
\end{aligned} \tag{65}$$

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

$$Q(\mathbf{x}_k, \mathbf{u}_k) = \ell(\mathbf{x}_k, \mathbf{u}_k) + V(\mathbf{x}_{k+1}) \tag{66}$$

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

$$\begin{aligned}
&Q(\bar{\mathbf{x}} + \delta\mathbf{x}, \bar{\mathbf{u}} + \delta\mathbf{u}) = Q_0 + \delta\mathbf{u}^T Q_{\mathbf{u}} + \delta\mathbf{x}^T Q_{\mathbf{x}}\\
&\frac{1}{2}\begin{bmatrix} \delta\mathbf{x}^T & \delta\mathbf{u}^T \end{bmatrix} \begin{bmatrix} Q_{\mathbf{xx}} & Q_{\mathbf{xu}} \\ Q_{\mathbf{ux}} & Q_{\mathbf{uu}} \end{bmatrix} \begin{bmatrix} \delta\mathbf{x} \\ \delta\mathbf{u} \end{bmatrix}
\end{aligned} \tag{67}$$

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:

$$\begin{aligned}
Q_{\mathbf{x}} &= \ell_{\mathbf{x}} + A_t V_{\mathbf{x}} + \tilde{\boldsymbol{\mathcal{S}}}\\
Q_{\mathbf{u}} &= \ell_{\mathbf{u}} + A_t V_{\mathbf{x}} + \tilde{\boldsymbol{\mathcal{U}}}\\
Q_{\mathbf{xx}} &= \ell_{\mathbf{xx}} + A_t^T V_{\mathbf{xx}} A_t + \boldsymbol{\mathcal{F}} + \tilde{\boldsymbol{\mathcal{F}}} + \tilde{\mathbf{M}}\\
Q_{\mathbf{xu}} &= \ell_{\mathbf{xu}} + A_t^T V_{\mathbf{xu}} B_t + \boldsymbol{\mathcal{L}} + \tilde{\boldsymbol{\mathcal{L}}} + \tilde{\mathbf{N}}\\
Q_{\mathbf{uu}} &= \ell_{\mathbf{uu}} + B_t^T V_{\mathbf{uu}} B_t + \boldsymbol{\mathcal{Z}} + \tilde{\boldsymbol{\mathcal{Z}}} + \tilde{\mathbf{G}}
\end{aligned} \tag{68}$$

TABLE I

Pseudocode of the SDDP algorithm

```
• Given:
    – An immediate cost function $\ell(\mathbf{x}, \mathbf{u})$
    – A terminal cost term $\phi_{t_N}$.
    – The stochastic dynamics $d\mathbf{x} = f(\mathbf{x}, \mathbf{u})dt + F(\mathbf{x}, \mathbf{u})d\omega$
• Repeat until convergence:
    – Given a trajectory in states and controls $\bar{\mathbf{x}}, \bar{\mathbf{u}}$ find the quadratic
      approximations of the stochastic dynamics $A_t, B_t, \Gamma_t, \mathbf{O}_d$ and
      the quadratic approximation of the immediate cost function
      $\ell_o, \ell_{\mathbf{x}}, \ell_{\mathbf{xx}}, \ell_{\mathbf{uu}}, \ell_{\mathbf{ux}}$ around these trajectories.
    – Compute all the terms $Q_{\mathbf{x}}, Q_{\mathbf{u}}, Q_{\mathbf{xu}}$ and $Q_{\mathbf{uu}}$ according to
      equation (68).
    – Back-propagate the quadratic approximation of the value function
      based on the equations:
        * $V_0^{(k+1)} = V_0^{(k+1)} - Q_{\mathbf{u}} Q_{\mathbf{uu}}^{-1} Q_{\mathbf{u}}$
        * $V_{\mathbf{x}}^{(k+1)} = Q_{\mathbf{x}}^{(k+1)} - Q_{\mathbf{u}} Q_{\mathbf{uu}}^{-1} Q_{\mathbf{ux}}$
        * $V_{\mathbf{xx}}^{(k+1)} = Q_{\mathbf{xx}}^{(k+1)} - Q_{\mathbf{xu}} Q_{\mathbf{uu}}^{-1} Q_{\mathbf{ux}}$
    – Compute $\delta\mathbf{u}^* = -Q_{\mathbf{uu}}^{-1}\left(Q_{\mathbf{u}} + Q_{\mathbf{ux}}\delta\mathbf{x}\right)$
    – Update controls $\mathbf{u}^* = \mathbf{u}^* + \gamma \cdot \delta\mathbf{u}^*$
    – Get the new optimal trajectory $x^*$ by propagating the nonlinear
      dynamics $d\mathbf{x} = f(\mathbf{x}, \mathbf{u}^*)dt + F(\mathbf{x}, \mathbf{u}^*)d\omega$.
    – Set $\bar{\mathbf{x}} = \mathbf{x}^*$ and $\bar{\mathbf{u}} = \mathbf{u}^*$ and repeat.
```

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

$$\begin{aligned}
&\ell(\bar{\mathbf{x}} + \delta\mathbf{x}, \bar{\mathbf{u}} + \delta\mathbf{u}) = \ell_0 + \delta\mathbf{u}^T \ell_{\mathbf{u}} + \delta\mathbf{x}^T \ell_{\mathbf{x}}\\
&\frac{1}{2}\begin{bmatrix} \delta\mathbf{x}^T & \delta\mathbf{u}^T \end{bmatrix} \begin{bmatrix} \ell_{\mathbf{xx}} & \ell_{\mathbf{xu}} \\ \ell_{\mathbf{ux}} & \ell_{\mathbf{uu}} \end{bmatrix} \begin{bmatrix} \delta\mathbf{x} \\ \delta\mathbf{u} \end{bmatrix}
\end{aligned} \tag{69}$$

with $\ell_{\mathbf{x}} = \frac{\partial \ell}{\partial \mathbf{x}}$, $\ell_{\mathbf{u}} = \frac{\partial \ell}{\partial \mathbf{u}}$, $\ell_{\mathbf{xx}} = \frac{\partial^2 \ell}{\partial \mathbf{x}^2}$, $\ell_{\mathbf{uu}} = \frac{\partial^2 \ell}{\partial \mathbf{u}^2}$ and $\ell_{\mathbf{ux}} = \frac{\partial^2 \ell}{\partial \mathbf{u} \partial \mathbf{x}}$. The local variations in control $\delta\mathbf{u}^*$ that maximize the state action value function are expressed by the equation that follows:

$$\begin{aligned}
\delta\mathbf{u}^* &= \underset{\mathbf{u}}{\operatorname{argmax}}\, Q(\bar{\mathbf{x}} + \delta\mathbf{x}, \bar{\mathbf{u}} + \delta\mathbf{u})\\
&= -Q_{\mathbf{uu}}^{-1}\left(Q_{\mathbf{u}} + Q_{\mathbf{ux}}\delta\mathbf{x}\right)
\end{aligned} \tag{70}$$

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. The SDDP algorithm is provided in a pseudocode form on Table I.

For the special cases where the stochastic dynamics have only additive noise $F(\mathbf{u}, \mathbf{x}) = F$ then the terms $\tilde{\mathbf{M}}, \tilde{\mathbf{N}}, \tilde{\mathbf{G}}, \tilde{\boldsymbol{\mathcal{F}}}, \tilde{\boldsymbol{\mathcal{L}}}, \tilde{\boldsymbol{\mathcal{Z}}}, \tilde{\boldsymbol{\mathcal{U}}}, \tilde{\boldsymbol{\mathcal{S}}}$ will be zero since they are functions of $\nabla_{\mathbf{xx}} F$ and $\nabla_{\mathbf{xu}} F$ and $\nabla_{\mathbf{uu}} F$ and it holds that $\nabla_{\mathbf{xx}} F = 0$, $\nabla_{\mathbf{xu}} F = 0$ and $\nabla_{\mathbf{uu}} F = 0$. In such a type of systems the control does not depend on the statistical characteristics of the noise. In cases of deterministic systems again $\tilde{\mathbf{M}}, \tilde{\mathbf{N}}, \tilde{\mathbf{G}}, \tilde{\boldsymbol{\mathcal{F}}}, \tilde{\boldsymbol{\mathcal{L}}}, \tilde{\boldsymbol{\mathcal{Z}}}, \tilde{\boldsymbol{\mathcal{U}}}, \tilde{\boldsymbol{\mathcal{S}}}$ will be zero because these terms depend on the variance of the noise $\sigma_{d\omega_i} = 0,\ \forall i = 1, ..., m$. Finally if the noise is only control depended then $\tilde{\mathbf{M}}, \tilde{\mathbf{N}}, \tilde{\boldsymbol{\mathcal{L}}}, \tilde{\boldsymbol{\mathcal{F}}}, \tilde{\boldsymbol{\mathcal{S}}}$ will be zero since $\nabla_{\mathbf{xx}} F(\mathbf{u}) = 0, \nabla_{\mathbf{xu}} F(\mathbf{u}) = 0$ and $\nabla_{\mathbf{x}} F_c^{(i)}(\mathbf{x}) = 0$ while if it is state dependent then $\tilde{\mathbf{N}}, \tilde{\mathbf{G}}, \tilde{\boldsymbol{\mathcal{Z}}}, \tilde{\boldsymbol{\mathcal{L}}}, \tilde{\boldsymbol{\mathcal{U}}}$ will be zero since $\nabla_{\mathbf{xu}} F(\mathbf{x}) = 0, \nabla_{\mathbf{uu}} F(\mathbf{x}) = 0$ and $\nabla_{\mathbf{u}} F_c^{(i)}(\mathbf{x}) = 0$.

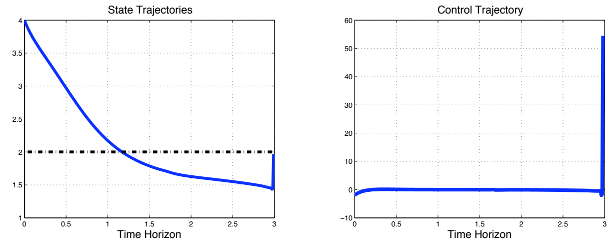

Fig. 1. The left plot illustrates the state trajectory for the the system $dx = \alpha cos(x)dt + u + x^2 d\omega$. The right plot corresponds to the optimal control.

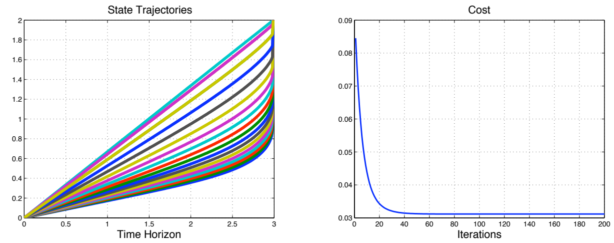

Fig. 2. Left plot illustrates the state space trajectories of the system $dx = \alpha x^2 dt + udt + x^2 d\omega$ for different values of noise. The right plot shows the stereotypical convergence behavior of SDDP.

V. Simulation Results

In this section we are testing the SDDP on two different one dimensional systems. Most precisely, we consider the one-dimensional stochastic nonlinear system of the form:

$$dx = cos(x)dt + udt + x^2 d\omega \tag{71}$$

The quadratic approximation is based on the matrices $A_t = 1 - sin(x)dt$, $B_t = dt$ and $O_{dt} = -cos(x)dt$. The system has only state depended noise and therefore $\tilde{\boldsymbol{\mathcal{M}}} = 2\sigma^2 dt V_{\mathbf{xx}} x^2$, $\tilde{\boldsymbol{\mathcal{F}}} = 4\sigma^2 dt V_{\mathbf{xx}} x^2$, $\tilde{\boldsymbol{\mathcal{S}}} = 2x^3 V_{\mathbf{xx}} \sigma^2 dt$ while the terms $\tilde{\boldsymbol{\mathcal{N}}} = \tilde{\boldsymbol{\mathcal{G}}} = \tilde{\boldsymbol{\mathcal{L}}} = \tilde{\boldsymbol{\mathcal{Z}}} = \tilde{\boldsymbol{\mathcal{U}}} = 0$. We apply the SDDP algorithm for the task of bringing the state from the initial $x_0 = 0$ to terminal $x(T) = p^* = 4$. The cost function is expressed as $v^\pi(\mathbf{x}, t) = E\left[h(\mathbf{x}(T)) + \int_{t_0}^T ru^2 d\tau\right]$ with $h(\mathbf{x}(T)) = w_p(x - p^*)^2$ where $w_p = 1$ and $r = 10^{-6}$. In this example, there is only a terminal cost for the state while during the time horizon the state dependent cost is zero and thus there is only control cost. In figure 1, the left plot illustrates the state trajectory as it approaches the target state $p^* = 4$ while the right plot illustrates the optimal control. Since there is only a terminal cost, the control over the time horizon is almost zero while at the very end of the time horizon the control is activated and the state reaches the target state.

The second system is given by the equation that follows:

$$dx = \alpha x^2 dt + udt + x^2 d\omega \tag{72}$$

The quadratic approximation is based on the matrices $A_t = 1 - 2\alpha x(t)dt$, $B_t = dt$ and $O_{dt} = -\alpha dt$. The parameter α controls the degree of the instability of the system. For our simulations we used $\alpha = 0.005$. The task is to bring the state $x(t)$ from the initial state $x_o = 0$ to target state $x(T) = p^* = 2$. The cost function has the same form as in the first example but with tuning parameters $w_p = 1$ and $r = 10^{-3}$. Furthermore, the system has only state depended noise and therefore $\tilde{\boldsymbol{\mathcal{M}}} = 2\sigma^2 dt V_{\mathbf{xx}} x^2$, $\tilde{\boldsymbol{\mathcal{F}}} = 4\sigma^2 dt V_{\mathbf{xx}} x^2$, $\tilde{\boldsymbol{\mathcal{S}}} = 2x^3 V_{\mathbf{xx}} \sigma^2 dt$ while the terms $\tilde{\boldsymbol{\mathcal{N}}} = \tilde{\boldsymbol{\mathcal{G}}} = \tilde{\boldsymbol{\mathcal{L}}} = \tilde{\boldsymbol{\mathcal{Z}}} = \tilde{\boldsymbol{\mathcal{U}}} = 0$.

In figure 2, the left plot illustrates state space trajectories for different values of the variance of the noise[1] while the right plot illustrates the convergence behavior of SDDP. An important observation is that curved trajectories correspond to cases with high variance noise. Furthermore, as the variance of the noise decreases the optimal trajectories become more and more straight.

VI. 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 further testing and evaluation of our generalized Stochastic DDP algorithm on multidimensional stochastic systems which are highly nonlinear and have noise that is control and state dependent. Biomechanical models belong in 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.

[1] In both examples, the noise is used only during the iterative optimization process of SDDP. When testing the optimal control, instead of running the controlled system many times for different realizations of noise and then calculate the mean, we just zero the noise and run the system only once.

References

[1] 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.

[2] W. Li. and E Todorov. Iterative optimal control and estimation design for stochastic nonlinear systems. In *Proceedings of Conference on Decision and Control*, 2006.

[3] D. H. Jacobson and D. Q. Mayne. *Differential Dynamic Programming*. Optimal Control. Elsevier Publishing Company, New York, 1970.

[4] Gregory Lantoine and Ryan P. Russell. A hybrid differential dynamic programming algorithm for robust low-thrust optimization. In *AAS/AIAA Astrodynamics Specialist Conference and Exhibit*, 2008.

[5] S. Yakowitz. The stagewise kuhn-tucker condition and differential dynamic programming. *IEEE Transactions on Automatic Control*, 31(1):25–30, 1986.

[6] Jun Morimoto and Chris Atkeson. Minimax differential dynamic programming: An application to robust biped walking. In *In Advances in Neural Information Processing Systems 15*. MIT Press, Cambridge, MA, 2002.

[7] Yuval Tassa, Tom Erez, and William Smart. Receding horizon differential dynamic programming. In J.C. Platt, D. Koller, Y. Singer, and S. Roweis, editors, *Advances in Neural Information Processing Systems 20*, pages 1465–1472. MIT Press, Cambridge, MA, 2008.