The $\delta-$ sensitivity and its application to stochastic optimal control of nonlinear diffusions.

Evangelos A. Theodorou[1] and Emo Todorov[2]

***Abstract*— We provide optimal control laws by using tools from stochastic calculus of variations and the mathematical concept of $\delta-$sensitivity. The analysis relies on logarithmic transformations of the value functions and the use of linearly solvable Partial Differential Equations(PDEs). We derive the corresponding optimal control as a function of the $\delta-$sensitivity of the logarithmic transformation of the value function for the case of nonlinear diffusion processes affine in control and noise.**

I. INTRODUCTION

Stochastic optimal control for Markov diffusion process affine in controls and noise has been considered in different areas of science and engineering such as machine learning, control theory and robotics [1], [3]–[6], [9]–[11], [15], [16]. One of the common methodologies for solving such optimal control problems is based on the exponential transformations of value functions. The main idea is inspired by the fact that linear PDEs can be solved with the application of the Feynman-Kac lemma which provides a tractable sampling numerical scheme. Therefore, instead of working with the initial nonlinear PDE that is the Hamilton-Jacobi-Bellman (HJB) equation, one can use the exponential transformation of the initial value function to get the corresponding linear PDE and then apply the Feynman-Kac lemma to find stochastic representations of its solution.

Parallel to the work in control theory, machine learning and robotics, researchers in the domain of financial engineering developed tools based on stochastic calculus of variations to find sensitivities of expectations of payoff functions with respect to changes in the parameters of the underlying diffusion processes [2], [7], [8], [12]. These sensitivities are named *Greeks* because they are often denoted by Greek letters. More precisely, many financial applications involve the computation of cost function:

$$\Psi(\mathbf{x}) = E\bigg(\phi(\vec{\mathbf{x}})\bigg) \tag{1}$$

where $\vec{\mathbf{x}}$ is a sample path $\vec{\mathbf{x}} = \{\mathbf{x}_{t_1}, \mathbf{x}_{t_2}, ..., \mathbf{x}_{t_N}\}$ generated by forward sampling of the diffusion process:

$$d\mathbf{x} = \mathbf{b}(\mathbf{x})dt + \boldsymbol{\sigma}(\mathbf{x}(t))d\mathbf{w}(t) \tag{2}$$

The cost function $\Psi(\mathbf{x})$ can be computed with Monte-Carlo simulations. Despite the estimation of the cost based on sample paths, financial applications require the computation of its differential with respect to the initial condition $\mathbf{x}_{t_0}$(δ-*delta sensitivity*), drift $\mathbf{b}(\mathbf{x})$(ρ-*rho sensitivity*) and diffusion $\boldsymbol{\sigma}(\mathbf{x})$ (ν-*vega sensitivity*) of the stochastic dynamics in (2).

The contribution of this work is to show that the $\delta-$sensitivity appears in the computation of the optimal control under the logarithmic transformation of value function. In particular we provide explicit formulas for the stochastic optimal control of Markov diffusion processes affine in control and noise. The analysis relies on the stochastic calculus of variations and the concept of Malliavin derivative. Related work in this area traces back to [13] as well as more recent work on the application of Malliavin calculus to finance [7], [8]. Here we are revisiting the topic of stochastic calculus of variations and show its applicability to find optimal control laws. In addition we provide all the underlying assumptions regarding smoothness and differentiability conditions of the stochastic dynamics under consideration.

The paper is organized as follows. In section III we review basic theorem of the stochastic calculus of variations which will be used for computing the δ- sensitivity. In section V we present the derivation of stochastic optimal control based on the logarithmic transformation of the value function. In the last section VI we conclude.

II. NOTATION

We denote the norm $x \in L^2(P)$ as $||x||_{L^2(P)} = (E[x^2])^{1/2} = (\int x^2(\omega)P(d\omega))^{1/2}$. Let $\mathbf{x} \in \Re^n$ and the function $h(\mathbf{x}) : \Re^n \rightarrow \Re$ then $\nabla h(\mathbf{x}) = (\frac{\partial h(\mathbf{x})}{\partial \mathbf{x}_1}, ..., \frac{\partial h(\mathbf{x})}{\partial \mathbf{x}_n})^T$. Let $\mathbf{b}(\mathbf{x}) : \Re^n \rightarrow \Re^n$ then $J_{\mathbf{b}(\mathbf{x})} : \Re^n \rightarrow \Re^n \times \Re^n$ is the Jacobian of $\mathbf{b}(\mathbf{x})$ with respect to $\mathbf{x}$ defined as $J^T_{\mathbf{b}(\mathbf{x})} = [\nabla \mathbf{b}_i(\mathbf{x}), ..., \nabla \mathbf{b}_N(\mathbf{x})]$. Let $\mathbf{x} = (x_1, ..., x_n)$ then the norm $||\mathbf{x}||_2 = \sqrt{x_1^2 + x_2^2 + ... + x_n^2}$. We denote the Banach spaces $\mathbb{D}_{1,2}$ equipped with the corresponding norm $||F||_{1,2}$ that is defined in the next section.

III. STOCHASTIC CALCULUS OF VARIATIONS.

In this section we provide the basic definitions and properties of the stochastic calculus of variations in the form of propositions and definitions.

Definition 1-(Malliavin Derivative): *Let* $\{\mathbf{w}(t), 0 \leq t \leq T\}$ *be a n-dimensional brownian motion on a complete probability space* $(\Omega, \mathcal{F}, \mathcal{P})$. *Let the set* $\mathcal{G}$ *of random variables* F *of the form:*

$$F = f\left(\int_0^\infty h_1(t)d\mathbf{w}(t), ... \int_0^\infty h_n(t)d\mathbf{w}(t)\right) \tag{3}$$

[1] Evangelos A. Theodorou is a Postdoctoral Research Associate with the Department of Computer Science and Engineering,University of Washington, Seattle. etheodor@cs.washington.edu

[2] Emo Todorov is Associate professor with the Departments of Computer Science and Engineering, and Applied Math, University of Washington, Seattle. todorov@cs.washington.edu

with $f \in \mathcal{G}(\Re^n)$. The set $\mathcal{G}(\Re^n)$ consists of infinitely differentiable and rapidly decreasing function on $\Re^n$ and $h_1, h_2, ..., h_n \in L^2(\Omega \times \Re_+)$. For $F \in \mathcal{G}$ the Malliavin derivative DF of f is defined as the process $\{D_t F, 0 \leq t\}$ with $D_t F : L^2(\Omega \times \Re_+) \rightarrow L^2(\Re_+)$ and expressed by the equation:

$$D_t F = \sum_1^n \frac{\partial f}{\partial x_i} \left(\int_0^\infty h_1(t) d\mathbf{w}(t), ... \int_0^\infty h_n(t) d\mathbf{w}(t) \right) h_i(t) \tag{4}$$

for $t \geq 0$.

Definition 2: *Let the norm $||F||_{1,2}$ denoted as:*

$$||F||_{1,2} = E(F^2)^{1/2} + \left(E\left(\int_0^\infty (D_t F)^2 dt \right) \right)^{1/2} \tag{5}$$

then $\mathbb{D}_{1,2}$ denotes the Banach space that is the completion of $\mathcal{G}$ with respect to the norm $||\ ||_{1,2}$. The stochastic derivative operator D_t is the a closed linear operator defined in $\mathbb{D}_{1,2}$ that takes values in $L^2(\Omega \times \Re_+)$. .

Next we provide a set of propositions regarding the use of the Malliavin derivative.

Proposition1-(chain rule): *Let $q(\mathbf{x}) : \Re^n \rightarrow \Re$ be a continuously differentiable function with bounded partial derivatives and $F = (F_1, F_2, ..., F_n)$ a random vector whose components $F_i \in \mathbb{D}_{1,2}$ then $q(F) \in \mathbb{D}_{1,2}$ and its Malliavin derivative is defined as:*

$$D_t q(F) = \sum_{i=1}^n \frac{\partial q}{\partial x_i}(F)\ D_t F \tag{6}$$

Proposition 2-(The fundamental theorem of Calculus) *Let $u = u(s), s \in [0,T]$ be a stochastic processes such that $E \int_0^T u^2(s) ds < \infty$ and assume that $\forall s \in \in [0,T], u(s) \in \mathbb{D}_{1,2}$ then $\int_0^T u(s) \delta w(s)$ is well defined and belongs to $\mathbb{D}_{1,2}$ and*

$$D_t \bigg(\int_0^T u(s) \delta w(s) \bigg) = \int_0^T D_t u(s) \delta w(s) + u(t) \tag{7}$$

Proposition 3-(Duality Formula): *Let $F \in \mathbb{D}_{1,2}$ be $\mathcal{F}_T$ measurable and let u be an $\mathbb{F}$-adapted process with:*

$$E\left(\int_0^{t_N} u^2 dt \right) < \infty$$

Then:

$$E\bigg(F \int u(t) dW(t) \bigg) = E\bigg(\int u(t) D_t F dt \bigg) \tag{8}$$

Next we provide an important proposition that shows that for the case of Markov diffusion processes the stochastic derivative operator is related to the derivative of the processes with respect to the initial condition.

Proposition 4-(Malliavin Derivative and Diffusions Processes): *Let $\mathbf{x}(t), t \geq 0$ be an $\Re^n$ valued Itô process whose dynamics are driven by the stochastic differential equation:*

$$d\mathbf{x} = \mathbf{b}(\mathbf{x}) dt + \boldsymbol{\sigma}(\mathbf{x}(t)) d\mathbf{w}(t) \tag{9}$$

where $b(\mathbf{x})$ and $\sigma(\mathbf{x})$ are continuously differentiable functions with bounded derivatives. Let $\boldsymbol{\Upsilon}(t) = \frac{d\mathbf{x}(t)}{d\mathbf{x}(t_0)}, t \geq 0$ be the associated first variation process defined by the stochastic differential equation:

$$d\boldsymbol{\Upsilon}(t) = J_{\mathbf{b}}(\mathbf{x}(t)) \boldsymbol{\Upsilon}(t) dt + \sum_{i=1}^n J_{\boldsymbol{\sigma}_i}(\mathbf{x}(t)) \boldsymbol{\Upsilon}(t) dw^i(t) \tag{10}$$

where the initial condition $\boldsymbol{\Upsilon}(0) = I_n$ and $J_{\mathbf{b}}(\mathbf{x}(t))$ and $J_{\boldsymbol{\sigma}_i}(\mathbf{x}(t))$ the jacobians of $\mathbf{b}$ and $\boldsymbol{\sigma}_i$. I_n is the identity matrix of $\Re^n$, with $\boldsymbol{\sigma}_i(\mathbf{x}(t))$ denoting the $i-th$ column vector $\boldsymbol{\sigma}$. Then the process $\{\mathbf{x}(t), t \geq 0\}$ belongs to $\mathbb{D}_{1,2}$ and its Malliavin derivative is given by:

$$D_s \mathbf{x}(t) = \boldsymbol{\Upsilon}(t) \boldsymbol{\Upsilon}(s)^{-1} \boldsymbol{\sigma}(\mathbf{x}(s)) 1_{s \leq t}, \quad s \geq 0 \tag{11}$$

Hence, if $\psi \in C_b^1(\Re^n)$ then we have:

$$D_s \psi(\mathbf{x}_T) = \nabla \psi(\mathbf{x}_T) \boldsymbol{\Upsilon}(t) \boldsymbol{\Upsilon}(s)^{-1} \boldsymbol{\sigma}(\mathbf{x}(s)) 1_{s \leq t} \tag{12}$$

and also:

$$D_s \int_0^T \psi(\mathbf{x}_t) dt = \int_s^T \nabla \psi(\mathbf{x}_t) \boldsymbol{\Upsilon}(t) \boldsymbol{\Upsilon}(s)^{-1} \boldsymbol{\sigma}(\mathbf{x}(s)) dt \tag{13}$$

Proof: We consider the stochastic dynamics

$$d\mathbf{x} = \mathbf{b}(\mathbf{x}) dt + \boldsymbol{\sigma}(\mathbf{x}(t)) d\mathbf{w}(t)$$

written in the form:

$$\mathbf{x}(t) - \mathbf{x}(0) = \int_0^t \mathbf{b}(\mathbf{x}) d\tau + \int_0^t \boldsymbol{\sigma}(\mathbf{x}(t)) d\mathbf{w}(\tau) \tag{14}$$

or

$$\mathbf{x}(t) - \mathbf{x}(0) = \int_0^t \mathbf{b}(\mathbf{x}) d\tau + \int_0^t \sum_{j=1}^N \boldsymbol{\sigma}_j(\mathbf{x}(\tau)) dw_j(\tau)$$

We take the derivative with respect to the initial state:

$$\frac{d\mathbf{x}(t)}{d\mathbf{x}(0)} - I_n = \int_0^t \frac{d}{d\mathbf{x}(0)} \mathbf{b}(\mathbf{x}) d\tau + \int_0^t \frac{d}{d\mathbf{x}(0)} \sum_{j=1}^N \boldsymbol{\sigma}_j(\mathbf{x}(\tau)) dw_j(\tau)$$

we will have that:

$$\boldsymbol{\Upsilon}(t) - I_n = \int_0^t \frac{d\mathbf{b}(\mathbf{x})}{d\mathbf{x}(\tau)} \frac{d\mathbf{x}(\tau)}{d\mathbf{x}(0)} d\tau + \int_0^t \sum_{j=1}^N \frac{d\boldsymbol{\sigma}_j(\mathbf{x}(\tau))}{d\mathbf{x}(\tau)} \frac{d\mathbf{x}(\tau)}{d\mathbf{x}(0)} dw_j(\tau) \tag{15}$$

where the term $\boldsymbol{\Upsilon}(t) \equiv \frac{d\mathbf{x}(t)}{d\mathbf{x}(t_0)}$. In a more compact the equation above is written as:

$$\boldsymbol{\Upsilon}(t) - I_n = \int_0^t J_{\mathbf{b}(\mathbf{x})} \boldsymbol{\Upsilon}(\tau) d\tau + \int_0^t \sum_{j=1}^N J_{\boldsymbol{\sigma}_j(\mathbf{x})} \boldsymbol{\Upsilon}(\tau) dw_j(\tau)$$

or in a more compact form:

$$d\boldsymbol{\Upsilon}(t) = J_{\mathbf{b}(\mathbf{x})} \boldsymbol{\Upsilon}(t) dt + \sum_{j=1}^N J_{\boldsymbol{\sigma}_j(\mathbf{x})} \boldsymbol{\Upsilon}(t) dw_j(t), \quad \boldsymbol{\Upsilon}(0) = I. \tag{16}$$

Similarly we define $\mathcal{Z}(t) \equiv D_s \mathbf{x}(t)$and $\forall t \geq s$ so we will have from (14):

$$D_s \mathbf{x}(t) = D_s \int_s^t \mathbf{b}(\mathbf{x}) dt + D_s \int_s^t \boldsymbol{\sigma}(\mathbf{x}(t)) d\mathbf{w}(t)$$

which, after applying the fundamental theorem of calculus from (7), can be further written as:

$$\mathcal{Z}(t) = \int_s^t J_{\mathbf{b}(\mathbf{x})} \mathcal{Z}(\tau) d\tau + \int_s^t D_s \boldsymbol{\sigma}(\mathbf{x}(\tau)) d\mathbf{w}(\tau) + \boldsymbol{\sigma}(\mathbf{x}(s))$$

or

$$d\mathcal{Z}(t) = J_{\mathbf{b}(\mathbf{x})} \mathcal{Z}(t) dt + \sum_j^N J_{\boldsymbol{\sigma}_j(\mathbf{x}(t))} \mathcal{Z}(t) dw_j(t) \quad (17)$$

$$= \Big(J_{\mathbf{b}(\mathbf{x})} + \sum_j^N J_{\boldsymbol{\sigma}_j(\mathbf{x}(t))} dw_j(t) \Big) \mathcal{Z}(t) \quad (18)$$

with initial condition $\mathcal{Z}(s) = \boldsymbol{\sigma}(\mathbf{x}(s))$. From the equation (16) and (17) we will have that

$$\mathcal{Z}(t) = D_s \mathbf{x}(t) = \boldsymbol{\Upsilon}(t) \boldsymbol{\Upsilon}(s)^{-1} \boldsymbol{\sigma}(\mathbf{x}(s)) 1_{s \leq t}, \quad s \geq 0 \quad (19)$$

To see that we solve the equation above with respect to $\boldsymbol{\Upsilon}(t)$, $\boldsymbol{\Upsilon}(t) = \mathcal{Z}(t) \boldsymbol{\sigma}(\mathbf{x}(s))^{-1} \boldsymbol{\Upsilon}(s) 1_{s \leq t}$ and then substitute into (16). More precisely we write (16) starting from the initial time $\tau = s$ and we will have:

$$\boldsymbol{\Upsilon}(t) - \boldsymbol{\Upsilon}(s) = \int_s^t J_{\mathbf{b}(\mathbf{x})} \boldsymbol{\Upsilon}(\tau) d\tau + \int_s^t \sum_{j=1}^N J_{\boldsymbol{\sigma}_j(\mathbf{x})} \boldsymbol{\Upsilon}(\tau) dw_j(\tau)$$

substitution $\mathbf{Y}(t)$ results in:

$$\begin{aligned} &\Big(\mathcal{Z}(t) \boldsymbol{\sigma}(\mathbf{x}(s))^{-1} - I \Big) \boldsymbol{\Upsilon}(s) = \\ &\Big(\int_s^t J_{\mathbf{b}(\mathbf{x})} \mathcal{Z}(\tau) d\tau + \int_s^t \sum_{j=1}^N J_{\boldsymbol{\sigma}_j(\mathbf{x})} \mathcal{Z}(\tau) dw_j(\tau) \Big) \\ &\times \boldsymbol{\sigma}(\mathbf{x}(s))^{-1} \boldsymbol{\Upsilon}(s) \end{aligned}$$

Multiplication of both sides of the equation above with $\boldsymbol{\Upsilon}(s)^{-1} \boldsymbol{\sigma}(\mathbf{x}(s))$ will provide us with (17). Equations 12 and 11 can be trivially proved by application of the chain rule as defined in (6).

■

Next we derive another identity that will be used later in our derivations. More precisely let $a(t) \in [0 \quad T]$ be a deterministic function such that:

$$\int_{t_0}^T a(t) dt = 1$$

Then from (11) we have that for $t = t_i$ the expression that follows:

$$\mathcal{Z}(t_i) = D_s \mathbf{x}(t_i) = \boldsymbol{\Upsilon}(t_i) \boldsymbol{\Upsilon}(s)^{-1} \boldsymbol{\sigma}(\mathbf{x}(s)) 1_{s \leq t_i}, \quad (20)$$

and therefore we can write the following expression:

$$\boldsymbol{\Upsilon}(t_i) = \mathcal{Z}(t_i) \boldsymbol{\sigma}(\mathbf{x}(s))^{-1} \boldsymbol{\Upsilon}(s) 1_{s \leq t_i}$$

we multiply from both sides with $\alpha(s)$ and then integrate:

$$\int \alpha(s) \boldsymbol{\Upsilon}(t_i) ds = \int \alpha(s) \mathcal{Z}(t_i) \boldsymbol{\sigma}(\mathbf{x}(s))^{-1} \boldsymbol{\Upsilon}(s) 1_{s \leq t_i} ds$$

The equation above can be written as:

$$\boldsymbol{\Upsilon}(t_i) \int_0^T \alpha(s) ds = \int_0^T \alpha(s) \mathcal{Z}(t_i) \boldsymbol{\sigma}(\mathbf{x}(s))^{-1} \boldsymbol{\Upsilon}(s) 1_{s \leq t_i} ds$$

By making use of (20) we have the expression:

$$\boldsymbol{\Upsilon}(t_i) = \int_0^{t_N} \alpha(s) \mathcal{Z}(t_i) \boldsymbol{\sigma}(\mathbf{x}(s))^{-1} \boldsymbol{\Upsilon}(s) 1_{s \leq t_i} ds$$

and for $t_i = t_N$ the final result is:

$$\boldsymbol{\Upsilon}(t_N) = \int_0^{t_N} \alpha(s) \mathcal{Z}(t_N) \boldsymbol{\sigma}(\mathbf{x}(s))^{-1} \boldsymbol{\Upsilon}(s) ds$$

In the remaining of the analysis in this paper we assume that: **Assumption 1:** *The diffusion matrix $\boldsymbol{\sigma}$ satisfies the uniform ellipticity condition:* $\epsilon > 0, \quad \xi^T \boldsymbol{\sigma}^T \boldsymbol{\sigma} \xi \geq \epsilon ||\xi||_2^2, \quad \forall \xi, \mathbf{x} \in \Re^n$.

IV. The δ- sensitivity.

The goal in this section is to to find the gradient of the expectation of the form:

$$E\Big(\mathcal{J}(\vec{\mathbf{x}}) \Big| \mathbf{x}_{t_i} \Big) = E\Big(\int_t^{t_N} q(\mathbf{x}) dt \Big| \mathbf{x}_{t_i} \Big) \quad (21)$$

with respect to the changes in the initial state $\mathbf{x}_t$ of the diffusion process:

$$d\mathbf{x} = \mathbf{b}(\mathbf{x}) dt + \boldsymbol{\sigma}(\mathbf{x}(t)) d\mathbf{w}(t) \quad (22)$$

The trajectories $\vec{\mathbf{x}}$ are defined as $\vec{\mathbf{x}} = (\mathbf{x}_{t_{i+1}}, \mathbf{x}_{t_{i+2}}, ..., \mathbf{x}_{t_N})$. We also denote by ∇_i the partial derivative with respect to the i-th argument and we introduce the set Γ_m as follows:

$$\Gamma_N = \{ \alpha \in L^2([0.T]) | \int_0^{t_i} \alpha dt = 1, \quad \forall i = 1, 2, ..., m \} \quad (23)$$

Lemma 1: *Under the assumption 1, and for any $\mathbf{x} \in \Re^n$ and for any $\alpha \in \Gamma_N$ we have that:*

$$\nabla_{\mathbf{x}_{t_i}} E\Big(\mathcal{J}(\vec{\mathbf{x}}) \Big| \mathbf{x}_{t_i} \Big) = E\Big(\sum_{j=i+1}^N \nabla_{\mathbf{x}_{t_j}} \mathcal{J}(\vec{\mathbf{x}})^T \boldsymbol{\Upsilon}(t_j) \Big| \mathbf{x}_{t_i} \Big)$$

where $\boldsymbol{\Upsilon}(t)$ is the stochastic flow defined as:

$$d\boldsymbol{\Upsilon}(t) = J_{\mathbf{b}}(\mathbf{x}(t)) \boldsymbol{\Upsilon}(t) dt + \sum_{i=1}^n J_{\boldsymbol{\sigma}_i}(\mathbf{x}(t)) \boldsymbol{\Upsilon}(t) dw^i(t) \quad (24)$$

with the initial condition $\boldsymbol{\Upsilon}(0) = I_n$.

Proof: We assume that the functional $\mathcal{J}(\vec{\mathbf{x}})$ is continuously differentiable and therefore we will have that:

$$\begin{aligned}\nabla_{\mathbf{x}_{t_i}} E\Big(\mathcal{J}(\vec{\mathbf{x}})\Big|\mathbf{x}_{t_i}\Big)^T &= E\Big(\nabla_{\mathbf{x}_{t_i}}\mathcal{J}(\vec{\mathbf{x}})\Big|\mathbf{x}_{t_i}\Big)^T \\ &= E\Big(\sum_{j=i+1}^{N}\nabla_{\mathbf{x}_{t_j}}\mathcal{J}(\vec{\mathbf{x}})^T\frac{d\mathbf{x}_{t_j}}{d\mathbf{x}_{t_i}}\Big|\mathbf{x}_{t_i}\Big) \\ &= E\Big(\sum_{j=i+1}^{N}\nabla_{\mathbf{x}_{t_j}}\mathcal{J}(\vec{\mathbf{x}})^T\boldsymbol{\Upsilon}(t_j)\Big|\mathbf{x}_{t_i}\Big)\end{aligned}$$

■

For the case where $\mathcal{J}(\vec{\mathbf{x}}) = \phi(\mathbf{x}_T)$ and therefore $\mathcal{J}(\vec{\mathbf{x}})$ depends on the last state of the state trajectory $\vec{\mathbf{x}} = \{\mathbf{x}_{t_1}, ..., \mathbf{x}_{t_N}\}$ we have the following proposition:

Proposition 3: *Under the assumption 1, and for any* $\mathbf{x} \in \Re^n$ *and for any* $\alpha \in \Gamma_N$ *we have that:*

$$\begin{aligned}&\nabla_{\mathbf{x}_{t_i}} E\Big(\phi(\mathbf{x}_{t_N})\Big|\mathbf{x}_{t_i}\Big) \\ &= E\Big(\phi(\mathbf{x}_{t_N})\int_{t_i}^{t_N}\alpha(t)\Big(\boldsymbol{\sigma}(\mathbf{x})^{-1}\boldsymbol{\Upsilon}(t)\Big)^T d\mathbf{w}(t)\Big)\end{aligned}$$

where $\boldsymbol{\Upsilon}(t)$ *is the stochastic flow defined as:*

$$d\boldsymbol{\Upsilon}(t) = J_{\mathbf{b}}(\mathbf{x}(t))\boldsymbol{\Upsilon}(t)dt + \sum_{i=1}^{n} J_{\boldsymbol{\sigma}_i}(\mathbf{x}(t))\boldsymbol{\Upsilon}(t)^{-1}dw^i(t) \tag{25}$$

with the initial condition $\boldsymbol{\Upsilon}(0) = I_n$.

Proof: Since the stochastic flow is defined as $\boldsymbol{\Upsilon}(t_j) = \int_0^T D_t\mathbf{x}(t_i)\alpha(s)\boldsymbol{\sigma}(\mathbf{x}(s))^{-1}\boldsymbol{\Upsilon}(s)1_{s\leq t_i}ds$ we will have that:

$$\begin{aligned}&\nabla_{\mathbf{x}_{t_i}} E\Big(\phi(\mathbf{x}_{t_N})\Big|\mathbf{x}_{t_i}\Big)^T = \\ &= E\Big(\sum_{j=i+1}^{N}\nabla_{\mathbf{x}_{t_j}}\phi(\mathbf{x}_{t_N})^T\boldsymbol{\Upsilon}(t_N)\Big) \\ &= E\Big(\nabla_{\mathbf{x}_{t_N}}\phi(\mathbf{x}_{t_N})^T\boldsymbol{\Upsilon}(t_N)\Big) \\ &= E\Big(\nabla_{\mathbf{x}_{t_N}}\phi(\mathbf{x}_{t_N})^T\int_0^{t_N}D_t\mathbf{x}(t_N)\alpha(s)\boldsymbol{\sigma}(\mathbf{x}(s))^{-1}\boldsymbol{\Upsilon}(s)ds\Big) \\ &= E\Big(\int_0^{t_N}\nabla_{\mathbf{x}_{t_N}}\phi(\mathbf{x}_{t_N})^T D_t\mathbf{x}(t_N)\alpha(t)\boldsymbol{\sigma}(\mathbf{x}(t))^{-1}\boldsymbol{\Upsilon}(t)dt\Big) \\ &= E\Big(\int_0^{t_N}D_t\phi(\mathbf{x}_{t_N})^T\ \alpha(t)\boldsymbol{\sigma}(\mathbf{x}(t))^{-1}\boldsymbol{\Upsilon}(t)dt\Big) \\ &= E\Big(\phi(\mathbf{x}_{t_N})\int_0^T\alpha(t)\Big(\boldsymbol{\sigma}(\mathbf{x}(t))^{-1}\boldsymbol{\Upsilon}(t)\Big)d\mathbf{w}\Big)\end{aligned}$$

where in the last line we make use of (8). Thus we have the final result:

$$\nabla_{\mathbf{x}_{t_i}} E\Big(\phi(\mathbf{x}_{t_N})\Big|\mathbf{x}_{t_i}\Big) = E\Big(\phi(\mathbf{x}_{t_N})v(t)\Big)$$

where the term $v(t) = \int_0^T \alpha(t)\Big(\boldsymbol{\sigma}(\mathbf{x}(t))^{-1}\boldsymbol{\Upsilon}(t)\Big)^T d\mathbf{w}(t)$ is the forms the stochastic derivative. ■

The derivation for the case where $\mathcal{J}(\vec{\mathbf{x}})$ is not continuously differentiable can be found in [8].

V. Stochastic Optimal Control

In this section we will show how $\delta-$sensitivity is used for computing the optimal control in feedback form. More precisely, we consider stochastic optimal control in the classical sense, as a constrained optimization problem, with the cost function under minimization given by the mathematical expression:

$$\begin{aligned}V(\mathbf{x}) &= \min_{\mathbf{u}} E^{(1)}\Big[J(\mathbf{x},\mathbf{u})\Big] \\ &= \min_{\mathbf{u}} E^{(1)}\Big[\phi(\mathbf{x}(t_N)) + \int_{to}^{t_N}\boldsymbol{\mathcal{L}}(\mathbf{x},\mathbf{u},t)dt\Big]\end{aligned} \tag{26}$$

subject to the nonlinear stochastic dynamics:

$$\mathbf{dx} = \mathbf{F}(\mathbf{x},\mathbf{u})dt + \mathbf{B}(\mathbf{x})d\mathbf{w} \tag{27}$$

with $\mathbf{x} \in \Re^{n\times 1}$ denoting the state of the system, $\mathbf{u} \in \Re^{p\times 1}$ the control vector and $d\mathbf{w} \in \Re^{p\times 1}$ brownian noise. The function $\mathbf{F}(\mathbf{x},\mathbf{u})$ is a nonlinear function of the state $\mathbf{x}$ and affine in controls $\mathbf{u}$ and therefore is defined as $\mathbf{F}(\mathbf{x},\mathbf{u}) = \mathbf{f}(\mathbf{x}) + \mathbf{G}(\mathbf{x})\mathbf{u}$. The matrix $\mathbf{G}(\mathbf{x}) \in \Re^{n\times p}$ is the control matrix, $\mathbf{B}(\mathbf{x}) \in \Re^{n\times p}$ is the diffusion matrix and $\mathbf{f}(\mathbf{x}) \in \Re^{n\times 1}$ are the passive dynamics. The cost function $J(\mathbf{x},\mathbf{u})$ is a function of states and controls. Under the optimal controls $\mathbf{u}^*$ the cost function is equal to the value function $V(\mathbf{x})$. The term $\boldsymbol{\mathcal{L}}$(**x**,**u**,t) is the running cost and it is expressed as:

$$\boldsymbol{\mathcal{L}}(\mathbf{x},\mathbf{u},t) = q_0(\mathbf{x},t) + q_1(\mathbf{x},t)\mathbf{u} + \frac{1}{2}\mathbf{u}^T\mathbf{R}\mathbf{u} \tag{28}$$

Essentially, the running cost has three terms, the first $q_0(\mathbf{x}_t, t)$ is a state-dependent cost, the second term depends on states and controls and the third is the control cost with the term $\mathbf{R} > 0$ the corresponding weight. The stochastic HJB equation [5], [14] associated with this stochastic optimal control problem is expressed as follows:

$$-\partial_t V = \min_{\mathbf{u}}\Big(\boldsymbol{\mathcal{L}} + (\nabla_{\mathbf{x}}V)^T\mathbf{F} + \frac{1}{2}tr\left((\nabla_{\mathbf{xx}}V)\mathbf{B}\mathbf{B}^T\right)\Big) \tag{29}$$

To find the minimum, the cost function (26) is inserted into (29) and the gradient of the expression inside the parenthesis is taken with respect to controls $\mathbf{u}$ and set to zero. The corresponding optimal control is given by the equation:

$$\mathbf{u}(\mathbf{x}_t) = -\mathbf{R}^{-1}\Big(q_1(\mathbf{x},t) + \mathbf{G}(\mathbf{x})^T\nabla_{\mathbf{x}}V(\mathbf{x},t)\Big) \tag{30}$$

These optimal controls will push the system dynamics in the direction opposite that of the gradient of the value function $\nabla_{\mathbf{x}}V(\mathbf{x},t)$. The value function satisfies nonlinear, second-order PDE:

$$-\partial_t V = \tilde{q} + (\nabla_{\mathbf{x}} V)^T \tilde{\mathbf{f}} - \underline{\frac{1}{2}(\nabla_{\mathbf{x}} V)^T \mathbf{G}\mathbf{R}^{-1}\mathbf{G}^T(\nabla_{\mathbf{x}} V)} + \frac{1}{2} tr\left((\nabla_{\mathbf{xx}} V)\mathbf{B}\mathbf{B}^T\right) \quad (31)$$

with $\tilde{q}(\mathbf{x}, t)$ and $\tilde{\mathbf{f}}(\mathbf{x}, t)$ defined as

$$\tilde{q}(\mathbf{x}, t) = q_0(\mathbf{x}, t) - \frac{1}{2} q_1(\mathbf{x}, t)^T \mathbf{R}^{-1} q_1(\mathbf{x}, t) \quad (32)$$

and

$$\tilde{\mathbf{f}}(\mathbf{x}, t) = \mathbf{f}(\mathbf{x}, t) - \mathbf{G}(\mathbf{x}, t)\mathbf{R}^{-1} q_1(\mathbf{x}, t) \quad (33)$$

and the boundary condition $V(\mathbf{x}_{t_N}) = \phi(\mathbf{x}_{t_N})$. Given the exponential transformation $V(\mathbf{x}, t) = -\lambda \log \Psi(\mathbf{x}, t)$ and the assumption $\lambda \mathbf{G}(\mathbf{x})\mathbf{R}^{-1}\mathbf{G}(\mathbf{x})^T = \mathbf{B}(\mathbf{x})\mathbf{B}(\mathbf{x})^T = \mathbf{\Sigma}(\mathbf{x}_t) = \mathbf{\Sigma}$ the resulting PDE is formulated as follows:

$$-\partial_t \Psi = -\frac{1}{\lambda}\tilde{q}\Psi + \tilde{\mathbf{f}}^T(\nabla_{\mathbf{x}}\Psi) + \frac{1}{2} tr\left((\nabla_{\mathbf{xx}}\Psi)\mathbf{\Sigma}\right) \quad (34)$$

with boundary condition: $\Psi(\mathbf{x}(t_N)) = \exp\left(-\frac{1}{\lambda}\phi(\mathbf{x}(t_N))\right)$. By applying the Feynman-Kac lemma to the Chapman-Kolmogorov PDE (34) yields its solution in form of an expectation over system trajectories. This solution is mathematically expressed as:

$$\Psi\left(\mathbf{x}_{t_i}\right) = E^{(0)}\left[\exp\left(-\int_{t_i}^{t_N} \frac{1}{\lambda}\tilde{q}(\mathbf{x})dt\right)\Psi(\mathbf{x}_{t_N})\right] \quad (35)$$

The expectation $E^{(0)}$ is taken on sample paths $\vec{\mathbf{x}}_i = (\mathbf{x}_i, ..., \mathbf{x}_{t_N})$ generated with the forward sampling of the uncontrolled diffusion equation:

$$\mathbf{dx} = \tilde{f}(\mathbf{x}_t)dt + \mathbf{B}(\mathbf{x})d\mathbf{w} \quad (36)$$

The optimal controls are specified as:

$$\mathbf{u}_{PI}(\mathbf{x}, t) = -\mathbf{R}^{-1}\left(q_1(\mathbf{x}, t) - \lambda \mathbf{G}(\mathbf{x})^T \frac{\nabla_{\mathbf{x}}\Psi(\mathbf{x}, t)}{\Psi(\mathbf{x}, t)}\right) \quad (37)$$

Since, the initial value the function $V(\mathbf{x}, t)$ is the minimum of the expectation of the objective function $J(\mathbf{x}, \mathbf{u})$ subject to controlled stochastic dynamics in (27), it can be trivially shown that:

$$V(\mathbf{x}, t_i) = -\lambda \log E^{(0)}\left[\exp\left(-\int_{t_i}^{t_N} \frac{1}{\lambda}\tilde{q}(\mathbf{x})dt\right)\Psi(\mathbf{x}_{t_N})\right] \leq E^{(1)}\Big(J(\mathbf{x}, \mathbf{u})\Big) \quad (38)$$

The expectation $E^{(1)}$ is taken on sample paths $\vec{\mathbf{x}}_i = (\mathbf{x}_i, ..., \mathbf{x}_{t_N})$ generated with the forward sampling of the controlled diffusion equation in (27). Closer look into the feedback control law in (37) reveals the use of the $\delta-$ sensitivity. In particular, the feedback control involves the term $\nabla_{\mathbf{x}}\Psi(\mathbf{x}, t)$ which correspond to the $\delta-$ Greek sensitivity of the functions $\Psi(\mathbf{x}, t)$ with respect to the initial state of the diffusion in (36). The analysis above is summarized by the following theorem:

Theorem 1: *Consider the stochastic optimal control problem with performance criterion and stochastic dynamics expressed as in (26),(28) and (27) with the terms* $q_0(\mathbf{x}, t), q_1(\mathbf{x}, t), \mathbf{f}(\mathbf{x}), \mathbf{B}(\mathbf{x}), \mathbf{G}(\mathbf{x})$ *be continuously differentiable,* $\mathbf{f}(\mathbf{x})$, $\mathbf{B}(\mathbf{x}), \mathbf{G}(\mathbf{x})$ *have bounded Lipschitz derivatives and* $\tilde{q}(\mathbf{x}, t)$ *and* $\tilde{\mathbf{f}}(\mathbf{x}, t)$ *are given by (32) and (33). Under the assumption of twice continuous differentiabilty of the value function* $V(\mathbf{x})$*, the optimal control law is expressed as:*

$$\mathbf{u}_{PI}(\mathbf{x}_{t_i}, t_i) = -\mathbf{R}^{-1} q_1(\mathbf{x}_{t_i}, t) + \mathbf{R}^{-1}\frac{\lambda \mathbf{G}(\mathbf{x}_{t_i})^T}{\Psi(\mathbf{x}_{t_i}, t_i)} E\Bigg(\sum_{j=i+1}^{N} \mathbf{\Upsilon}(t_j)^T \nabla_{\mathbf{x}(t_j)}\mathcal{J}(\vec{\mathbf{x}})\Bigg)\Bigg]$$

The term $\mathcal{J}(\vec{\mathbf{x}})$ *is defined as* $\mathcal{J}(\vec{\mathbf{x}}) = \exp\left(-\frac{1}{\lambda}\int_{t_i}^{t_N}\tilde{q}(\mathbf{x}(s))ds\right)$ *while* $\mathbf{\Upsilon}(t) = \frac{d\mathbf{x}(t)}{d\mathbf{x}(t_i)}$ $\forall t > t_i$ *is the stochastic flow:*

$$d\mathbf{\Upsilon}(t) = J_{\tilde{\mathbf{f}}(\mathbf{x}(t_i))}\mathbf{\Upsilon}(t)dt + \sum_{k=1}^{n} J_{\mathbf{B}_k}(\mathbf{x}(t_i))\mathbf{\Upsilon}(t)dw^k(t)$$

that corresponds to the stochastic differential equation:

$$\mathbf{dx} = \tilde{f}(\mathbf{x}_t)dt + \mathbf{B}(\mathbf{x})d\mathbf{w}$$

VI. Conclusions

We derived the optimal control law by using the $\delta-$sensitivity. Consider the case of estimating the expectation of a cost function over state paths generated by sampling of a diffusion process. Essentially $\delta-$sensitivity corresponds to the differential of the expected cost function with respect to the initial state of the underlying diffusion process. Future research will involve the development of efficient algorithms for applications to dynamical systems in robotics and biology. Another line of research involves the use of stochastic calculus of variations for optimal control of Markov jump diffusion processes.

References

[1] Paolo Dai Pra, Lorenzo Meneghini, and Wolfgang Runggaldier. Connections between stochastic control and dynamic games. *Mathematics of Control, Signals, and Systems (MCSS)*, 9(4):303–326, 1996-12-08.

[2] Mark H. A. Davis and Martin P. Johansson. Malliavin monte carlo greeks for jump diffusions. *Stochastic Processes and their Applications*, 116(1):101–129, January 2006.

[3] W. H. Fleming and W. M. McEneaney. Risk-sensitive control on an infinite time horizon. *SIAM J. Control Optim.*, 33:1881–1915, November 1995.

[4] W. H. Fleming and H. Mete Soner. *Controlled Markov processes and viscosity solutions*. Applications of mathematics. Springer, New York, 1nd edition, 1993.

[5] W. H. Fleming and H. Mete Soner. *Controlled Markov processes and viscosity solutions*. Applications of mathematics. Springer, New York, 2nd edition, 2006.

[6] W.H. Fleming. Exit probabilities and optimal stochastic control. *Applied Math. Optim*, 9:329–346, 1971.

[7] Eric Fournié, Jean-Michel Lasry, Pierre-Louis Lions, and Jrme Lebuchoux. Applications of malliavin calculus to monte-carlo methods in finance. ii. *Finance and Stochastics*, 5(2):201–236, 2001.

[8] Eric Fournié, Jean-Michel Lasry, Pierre-Louis Lions, Jrme Lebuchoux, and Nizar Touzi. Applications of malliavin calculus to monte carlo methods in finance. *Finance and Stochastics*, 3(4):391–412, 1999.

[9] H. J. Kappen. Path integrals and symmetry breaking for optimal control theory. *Journal of Statistical Mechanics: Theory and Experiment*, 11:P11011, 2005.

[10] H. J. Kappen. An introduction to stochastic control theory, path integrals and reinforcement learning. In J. Marro, P. L. Garrido, and J. J. Torres, editors, *Cooperative Behavior in Neural Systems*, volume 887 of *American Institute of Physics Conference Series*, pages 149–181, February 2007.

[11] Sanjoy K. Mitter and Nigel J. Newton. A variational approach to nonlinear estimation. *SIAM J. Control Optim.*, 42(5):1813–1833, May 2003.

[12] Giulia Di Nunno, Bernt Oksental, and Frank Proske. *Maliavin Calculus for Lévy Processes with Applications to Finance*. Springer, 1nd edition, 2009.

[13] D. Ocone and I. Karatzas. A generalized clark representation formula, with applications to optimal portfolios. *Stochastics*, 37:187–220., 1991.

[14] Robert F. Stengel. *Optimal control and estimation*. Dover books on advanced mathematics. Dover Publications, New York, 1994.

[15] E. Todorov. Linearly-solvable markov decision problems. In B. Scholkopf, J. Platt, and T. Hoffman, editors, *Advances in Neural Information Processing Systems 19 (NIPS 2007)*, Vancouver, BC, 2007. Cambridge, MA: MIT Press.

[16] E. Todorov. Efficient computation of optimal actions. *Proc Natl Acad Sci U S A*, 106(28):11478–83, 2009.