A Nonlinear Feynman-Kay Formula with Application in Linearly Solvable Optimal Control

Tom Lefebvre and Guillaume Crevecoeur
{Tom.Lefebvre,Guillaume.Crevecoeur}@ugent.be are with the Dept. of Electromechanical, Systems and Metal Engineering, Ghent University, 9000 Ghent; and with the Core Lab EEDT-DC, Flanders Make.

Abstract
In this article we present a solution to a nonlinear relative of the parabolic differential equation that was tackled by Feynman and Kac in the late 1940s. For the proof we rely on continuous time stochastic calculus. Second we draw an interesting connection with a related recurrence relation affirming the presented result by collapsing onto the continuous time framework but only in the limit. The equation emerges in the context of infinite horizon discounted Linearly Solvable Optimal Control, which, as far as we are aware of, is untreated by the literature. The continuous time setting can be treated using our new result. As we will demonstrate the discrete time setting is intractable. Nevertheless we can provide close estimates based on the recurrence relation which also allows us to estimate the influence of time discretization errors. We demonstrate our solution treating a small case study.
1. Introduction

Ever since the work of Feynman and Kac in the late 1940’s, when both men were addressing what turned out to be essentially the same problem\(^1\) the Feynman-Kac formula has grown into a well established means of evaluating parabolic partial differential equations (PDEs) subject to, amongst others, terminal boundary constraints.

\[
\frac{\partial}{\partial t} Z(t, x) = -q(t, x)Z(t, x) + \nabla_x Z(t, x)^\top a(t, x) + \frac{1}{2} \text{tr} \left( \Sigma(t, x) \nabla_{xx} Z(t, x) \right) \\
Z(T, x) = Z_T(x)
\]

The Feynman-Kac formula states

\[
Z(t, x) = \mathbb{E}_{P(X(t \to T)|X(t)=x)} \left[ \exp \left( -\int_t^T q(X(\tau))d\tau \right) Z_T(x) \right]
\]

here \(X(\tau \geq t)\) is a diffusion process with drift \(dX = a(X)dt + \sigma(X)dW\) with initial condition \(X(t) = x\) and \(dW\) being a Wiener process. Other quantities are introduced in theorem 1.

The formula draws a marvellous connection between parabolic PDEs and stochastic Itô or diffusion processes. Essentially the formula implies that the PDE mentioned above can be evaluated by simulating random paths of a stochastic Itô process and paved way for practical methods of evaluation for problems that were otherwise deemed intractable. Major areas of interest are quantum mechanics [4], finances [5] and more recently also optimal control [6, 7, 8, 9].

The aim of this paper is to present a solution to the related differential equation introduced below, which can be interpreted reasonably as a non-linear (and stationary) relative of the problem studied by Feynman and Kac,\(^1\)Although targeting the problem from their own respective field of expertise. Feynman’s work was motivated by attempts to produce tractable solutions for the one-dimensional Schrödinger equation. Feynman developed the concept that is now known as Path Integral theory, where the state of a particle at time \(t\) is obtained by averaging over all possible paths starting at a fixed initial state [1]. When Kac got wind of his then colleague’s ideas, he showed that the same principles applied to the heat equation, reinterpreting the averaging operation probabilistically as a conditional expectation. He subsequently revealed the fundamental relationship between stochastic diffusion processes and certain partial differential equations in a series of papers giving rise to what is now known as the Feynman-Kac formula [2, 3].
that is, as opposed to the former, not subjected to any constraints

\[
\alpha Z(x) \log Z(x) = -q(x)Z(x) + \nabla_x Z(x)^\top a(x) + \frac{1}{2} \text{tr} \left( \Sigma(x) \nabla_{xx} Z(x) \right)
\]

Specifically, we shall illustrate that the solution to this equation is given by the following conditional expectation

\[
Z(x) = \lim_{T \to \infty} \mathbb{E}_{P(X(t \to T) | X(t) = x)} [I(T)]
\]

\[
I(s) = \exp \left( - \int_t^s e^{-\alpha(t-\tau)}q(X(\tau))d\tau \right)
\]

We provide two methods of proof. One involves a modified version of the modern method of proof used for the linear Feynman-Kac formula. The other is based on a stochastic recurrence relation that collapses onto the continuous time framework but only in the limit. To the best of our knowledge the presented results are strictly original.

The studied nonlinear PDE emerges in the context of discounted infinite horizon Linearly Solvable Optimal Control (LSOC). This is a particular subclass of stochastic optimal control problems that allow for explicit evaluation of the value function and hence the optimal control. In the second part of this contribution we treat its application in LSOC [9].

2. Derivation

This section contains two methods of proof for the central result. The proofs are organised in two theorems, one addresses the continuous time problem as such, the other establishes a connection with a closely related recurrence relation that collapses onto the continuous time framework in the limit.

2.1. Itô calculus

The derivation of our results relies on the Itô calculus [4, 5]. For the purpose of self containment, we briefly present notation and summarize the required results from that area for the reader to acquaint with before progressing further.

The key concept in Itô calculus is that of the Wiener process. A Wiener process is a mathematical entity that can be used to construct arbitrary stochastic processes.
Definition 1 (Wiener process). The Wiener process $W(t)$ is defined by means of the following properties

- for $\Delta t > 0$, the increment $W(t + \Delta t) - W(t)$ is Gaussian with mean zero and variance $\Delta t$
- the limit $\lim_{\Delta t \to 0} \frac{1}{\Delta t} (W(t + \Delta t) - W(t))$ is not defined

The mathematical construction of a signal exhibiting such properties is beyond that required to establish the result proposed here. It suffices to note that the Wiener process can be understood as the limit case of a discrete-time random walk.

These properties also imply that

\[ \mathbb{E}[dW] = 0 \]
\[ \mathbb{E}[dW^2] = dt \]

The concept of Brownian motion can be used to construct general stochastic processes. Specifically we will consider (Itô drift) diffusion or stochastic processes propelled by the stochastic differential equations of the following general form $dX = a(X(t))dt + \sigma(X(t))dW$ where $X \in \mathbb{R}^n$ and $dW \in \mathbb{R}^m$. Further we will denote the probability of a path spawned by such a diffusion process conditioned on the initial value $X(t) = x$ as $P(X(t \to T)|X(t) = x)$. Note that such a path is a stochastic variable and that by consequence the expectation $\mathbb{E}_{P(X(t \to T)|X(t) = x)}[\cdot]$ is defined.

Within this framework Itô’s lemma generalises the notion of the total derivative of a function $V(t, x)$ along paths spawned by any diffusion process

**Lemma 1** (Itô’s lemma). Let $\{V, a, \sigma\} : \mathbb{R} \times \mathbb{R}^n \mapsto \{\mathbb{R}, \mathbb{R}^n, \mathbb{R}^{n \times m}\}$ be functions of time and space where $V(t, x)$ is twice differentiable at least and let $X$ be a stochastic process that can be modelled as a Brownian motion with drift $a(t, x)$ and covariance $\Sigma(t, x) = \sigma(t, x)\sigma(t, x)^\top$ then we have that

\[
dV(t, X) \approx (\partial_t V(t, X) + \nabla_x V(t, X)^\top a(t, X) + \frac{1}{2} \text{tr}(\Sigma(t, X) \nabla_{xx} V(t, X)))dt + \nabla_x V(t, X)^\top \sigma(t, X)dW + \mathcal{O}(d^2)
\]

With major implication

\[
\lim_{dt \to 0} \frac{1}{dt} \mathbb{E}_{dX}[dV(t, X)] = \partial_t V(t, X) + \nabla_x V(t, X)^\top a(t, X) + \frac{1}{2} \text{tr}(\Sigma(t, X) \nabla_{xx} V(t, X))
\]
2.2. Main results

For notational brevity and rapid intuitive interpretation, henceforth we restrain from writing the functional argument as often as the context permits.

We provide the continuous time treatment first.

**Theorem 1.** Let \( \{Z, q, a, \sigma\} : \mathbb{R}^n \rightarrow \{\mathbb{R}, \mathbb{R}^n, \mathbb{R}^{n \times n}\} \) be scalar, vector and matrix functions of the spatial coordinate \( x \) respectively, so that \( \Sigma = \sigma \sigma^\top \succ 0 \) and let \( \alpha > 0 \) a positive constant. Now consider the PDE

\[
\alpha Z \log Z = -qZ + \nabla_x Z^\top a + \frac{1}{2} \text{tr} (\Sigma \nabla_{xx} Z)
\]

If \( X \) is a stochastic process propelled by the stochastic differential equation

\[
dX = a(X)dt + \sigma(X)dW
\]

where \( dW \) is a Wiener process then we have that

\[
Z(x) = \lim_{T \to \infty} \mathbb{E}_{P(X(t \to T)|X(t) = x)} [I(T)]
\]

\[
I(s) = \exp \left( -\int_t^s e^{-\alpha (\tau - t)} q(X(\tau)) d\tau \right)
\]

**Proof.** Consider the stochastic process

\[
Y(s) = Z(s) e^{-\alpha (s-t)} I(s)
\]

The product rule then implies that

\[
dY = d \left( Z^{-\alpha s'} \right) I + Z^{-\alpha s'} dI, \ s' = s - t
\]

and one easily verifies that

\[
dI = -e^{-\alpha s'} q I ds
\]

The first differential is trickier and involves the chain rule.

\[
d \left( Z^{-\alpha s'} \right) = \frac{\partial}{\partial Z} \left( Z^{-\alpha s'} \right) dZ + \frac{\partial}{\partial s} \left( Z^{-\alpha s'} \right) ds
\]

\[
= e^{-\alpha s'} Z^{-\alpha s' -1} dZ - \alpha e^{-\alpha s'} Z^{-\alpha s'} \log Z ds
\]

\[
= e^{-\alpha s'} Z^{-\alpha s' -1} \left( (\nabla_x Z^\top a + \frac{1}{2} \text{tr} (\Sigma \nabla_{xx} Z)) ds + \nabla_x Z^\top dW \right) - \alpha e^{-\alpha s'} Z^{-\alpha s'} \log Z ds
\]
Substitution of both results into the expression for $dY$ then yields

$$
dY = e^{-\alpha s'} Z e^{-\alpha s'} - 1 \left( \nabla_x Z^\top a + \frac{1}{2} \text{tr} (\Sigma \nabla_{xx} Z) \right) Ids
- \alpha e^{-\alpha s'} Z e^{-\alpha s'} I \log Z ds - e^{-\alpha s'} q Z e^{-\alpha s'} S ds
+ e^{-\alpha s'} Z e^{-\alpha s'} - 1 I \nabla_x Z^\top dW
$$

which can be rearranged into

$$
dY = e^{-\alpha s'} Z e^{-\alpha s'} - 1 \times
\left( -\alpha Z \log Z - qZ + \nabla_x Z^\top a + \frac{1}{2} \text{tr} (\Sigma \nabla_{xx} Z) \right) Ids
+ e^{-\alpha s'} Z e^{-\alpha s'} - 1 I \nabla_x Z^\top dW
$$

The first term vanishes as dictated by the PDE. We can also get rid of the second term. First integrating from $t$ to $T$, yielding

$$
Y(T) - Y(t) = \int_t^T e^{-\alpha s'} Z e^{-\alpha s'} - 1 S \nabla_x Z^\top ds dW
$$

We then note that the right-hand side depends on $dW$. Hence we take the expectation over the path $X(t \rightarrow T)$ conditioned on $X(t) = x$ and recognize an Itô integral so that as a result the term vanishes effectively.

$$
\mathbb{E}_{P(X(t \rightarrow T) | X(t) = x)} [Y(T)] - Y(t) =
\mathbb{E}_{P(X(t \rightarrow T) | X(t) = x)} \left[ \int_t^T e^{-\alpha s'} Z e^{-\alpha s'} - 1 I \nabla_x Z^\top ds dW \right] = 0
$$

On account of the conditionality on $X(t) = x$ and the definition of $Y$ it follows directly that $Y(t) = Z(x)$. Hence we can rearrange the result above into an expression for $Z(x)$.

Finally, we must get rid of $Z(X(T))$. It is easily verified that if we take the limit $T \rightarrow \infty$ the proof is completed.

$$
Y(t) = Z(x)
= \mathbb{E}_{P(X(t \rightarrow T) | X(t) = x)} [Y(T)]
= \mathbb{E}_{P(P(X(t \rightarrow T) | X(t) = x))} \left[ I(T) Z(X(T)) - e^{\alpha (T-t)} \right]
= \lim_{T \rightarrow \infty} \mathbb{E}_{P(X(t \rightarrow T) | X(t) = x)} [I(T)]
$$

We refer to this solution as a nonlinear or stationary Feynman-Kac formula. \hfill \Box
We now present a second alternative method of proof. This method relies on identification of a stochastic recurrence relation that collapses onto the continuous time framework but only in the limit.

One might recognize that the result in discrete time is stronger in the sense that the result holds for general discrete time transition probabilities $P(X' = y | X = x)$. However to establish the connection with the result in continuous time, we need to address a restrictive class of transition probabilities that can be classified effectively as random walks with drift. Second in contrast to the solution in continuous time, here the solution is only exact in the limit. This is unfortunate for the anticipated application in control as we will discuss later.

**Theorem 2.** Let \( \{Z, q, a, \sigma\} : \mathbb{R}^n \mapsto \{\mathbb{R}, \mathbb{R}^n, \mathbb{R}^n, \mathbb{R}^{n \times m}\} \) be scalar, vector and matrix functions of the spatial coordinate \( x \) respectively and \( 1 > \alpha > 0 \) a positive constant and so that \( \Sigma = \sigma \sigma^\top > 0 \). Now consider the stochastic nonlinear recurrence relation

\[
Z(x) = \mathbb{E}_{P(X' | X = x)} \left[ e^{-q(X)} Z(X')^\alpha \right]
\]

Then

\[
\tilde{Z}(x) \leq Z(x) \leq C \tilde{Z}(x)
\]

where \( C \geq 1 \) but bounded and

\[
\tilde{Z}(x) = \lim_{N \to \infty} \mathbb{E}_{P(X(0 \to N) | X(n) = x)} [S(N)]
\]

\[
S(s) = \exp \left( - \sum_{n=0}^{s} \alpha^n q(X(n)) \right)
\]

and if \( X \) is a stochastic process propelled by the random walk with drift \( X' = X + \Delta X, \Delta X = a(X) \Delta t + \sigma(X) \Delta W \) where \( \Delta W \sim \mathcal{N}(0, \Delta t I) \) then the stochastic recurrence relation collapses onto the stationary PDE from theorem \( \square \) in the limit \( \Delta t \to 0 \) and the solution collapses onto the corresponding nonlinear Feynman-Kac formula.

---

2Notation $X' | X = x$ indicates the value of the state after one time instant has passed conditioned on $X = x$. This implies that we can associate a transition probability density function $P(X' = y | X = x)$ expressing the probability of reaching $y$ from $x$. The probability of a stochastic discrete time path conditioned on $X(n) = x$ is denoted as $P(X(n \to N) | X(n) = x)$. 
Proof. First note that the random walk
\[
\Delta X = a(X) \Delta t + \sigma(X) \Delta W, \Delta W \sim N(0, \Delta t I)
\]
shares the characteristics of the continuous time diffusion process described in theorem [1] if \( \Delta t \to 0 \) and can be considered equivalent consequently.

Second we overload notation and substitute discrete time approximations of the continuous time function \( q \) and scalar \( \alpha \) for the associated entities in the discrete time setting.

Specifically we have that \( q \mapsto q \Delta t \) and \( \alpha \mapsto e^{-\alpha \Delta t} \). Substitution into the stochastic recurrence relation yields
\[
Z(x) = \mathbb{E}_{P(\Delta X|X=x)} \left[ e^{-q(X)\Delta t} Z(x + \Delta X)e^{-\alpha \Delta t} \right]
\]

With these preliminary results in place we can start manipulating the recurrence relation. Furthermore since \( e^{-q(x)\Delta t} Z(x)e^{-\alpha \Delta t} \) is not a stochastic variable we can introduce it into the expectation operators argument
\[
0 = e^{-q(x)\Delta t} \mathbb{E}_{P(\Delta X|X=x)} \left[ Z(x + \Delta x)e^{-\alpha \Delta t} - Z(x)e^{-\alpha \Delta t} \right]
\]
\[
+ e^{-q(x)\Delta t} Z(x)e^{-\alpha \Delta t} - Z(x)
\]

Now we approximate the argument by its differential representation, i.e.
\[
Z(x + \Delta x)e^{-\alpha \Delta t} - Z(x)e^{-\alpha \Delta t} \approx e^{-q(x)\Delta t} Z(x)e^{-\alpha \Delta t} - 1 \Delta Z.
\]
This approximation is exact for \( \Delta Z \to 0 \) or equivalently when \( \{\Delta X, \Delta t\} \to 0 \)
\[
0 = e^{-\alpha \Delta t} e^{-q(x)\Delta t} Z(x)e^{-\alpha \Delta t} - 1 \mathbb{E}_{P(\Delta X|X=x)} [\Delta Z] + e^{-q(x)\Delta t} Z(x)e^{-\alpha \Delta t} - Z(x)
\]

By definition of the transition probability it follows direct that the expectation is equal to
\[
\mathbb{E}_{P(\Delta X)}[\Delta Z] = (\nabla_x Z^\top a + \frac{1}{2} \text{tr}(\Sigma \nabla_{xx} Z)) \Delta t
\]

The former result can be substituted into the recurrence relation.
\[
0 = \lim_{\Delta t \to 0} \frac{1}{\Delta t} e^{-q(\Delta t)} Z^{-\alpha \Delta t} \left( (\nabla_x Z^\top a + \frac{1}{2} \text{tr}(\Sigma \nabla_{xx} Z))\Delta t + \lim_{\Delta t \to 0} \frac{1}{\Delta t} \left( e^{-q(\Delta t)} Z^{-\alpha \Delta t} - Z \right) \right)
\]

Evaluation of the first term is trivial and established the part of the PDE that can be associated to the diffusion process.
The second term can be recognized as the definition of the derivative with respect to $\Delta t$ evaluated at $\Delta t = 0$.

$$\lim_{\Delta t \to 0} \frac{1}{\Delta t} \left( e^{-q\Delta t} Z e^{-\alpha\Delta t} - Z \right) = \frac{d}{d\Delta t} \left( e^{-q\Delta t} Z e^{-\alpha\Delta t} \right) \bigg|_{\Delta t = 0}$$

$$= -qZ - \alpha Z \log Z$$

These terms produce the $\alpha Z \log Z$ and $qZ$ terms that were still missing.

Substituting the intermediary results back into the main equation recovers the stationary PDE. It follows that the discrete time stochastic recurrence relation collapses onto the differential equation in the limit.

$$\alpha Z \log Z = -qZ + \nabla_x Z^\top a + \frac{1}{2} \text{tr} (\Sigma \nabla_{xx} Z)$$

Now it is left to proof that the same holds true for the solution. First let us show that the discrete time solution indeed satisfies the bounds in the theorem. As is easily verified, on account of the power $\alpha$ the recursion resists explicit evaluation. To resolve this inconvenience we introduce the variable $\delta_n$. Since $(\cdot)^\alpha$ is concave when $0 < \alpha < 1$, Jensen’s inequality ascertains the lower bound. Then since $\delta_0 = 1$ and $\lim_{n \to \infty} \delta_n = 1$ the upper bound is a reasonable assumption.

$$1 \leq \delta_n(x) = \frac{\mathbb{E}_{P(X'\mid X=x)} [Z(X')^\alpha]^n}{\mathbb{E}_{P(X'\mid X=x)} [Z(X')^\alpha]^{n+1}} \leq C_n$$

Substituting $\delta_n(x)\mathbb{E}_{P(X'\mid X=x)} [Z(X')^\alpha]^{n+1}$ for $\mathbb{E}_{P(X'\mid X=x)} [Z(X')^\alpha]^n$ whenever such an expression emerges, allows to evaluate the recursion anyhow and produces the result

$$Z(x) = \mathbb{E}_{P(X(0\to N)\mid X(0)=x)} \left[ S(N) Z(X(N))^\alpha \delta(N) \right]$$

where

$$\delta(N) = \prod_{n=1}^{N} \delta_n(X(n))$$

On account of the bounds on $\delta_n$ we have that $1 \leq \lim_{N \to \infty} \delta(N) \leq C \leq \infty$ guaranteeing that the solution satisfies indeed

$$\tilde{Z}(x) \leq Z(x) \leq C \tilde{Z}(x)$$

To show that the bounds tighten when $\Delta t \to 0$ we simply note that $\lim_{\Delta t \to 0} \delta_n = 1$ with $\alpha \mapsto e^{-\alpha \Delta t}$.
Further since the discrete time dynamics collapse onto the diffusion process it follows that \( \lim_{\Delta t \to 0} P(X(n \to N)|X(n) = x) = P(X(t \to T)|X(t) = x) \). Finally we practice the definition of the Riemann integral in the expectation operators argument so to obtain

\[
Z(x) = \tilde{Z}(x) = \lim_{\Delta t \to 0} \lim_{N \to \infty} \mathbb{E} P(X(n \to N)|X(n) = x) [S'(N)] \\
= \lim_{T \to \infty} \mathbb{E} P(X(t \to T)|X(t) = x) [I(T)]
\]

where

\[
S'(s) = \exp \left(- \sum_{m=n}^{s} e^{-\alpha(m-n)\Delta t} q(X(m)) \Delta t \right)
\]

This finalises the connection between theorems 1 and 2.

The theorem has two interesting consequences. Firstly, it guarantees that we can estimate a lower bound for the solution \( Z(x) \), namely \( \tilde{Z}(x) \), by approximating the expectation using sampled paths. Secondly the alternative proof is valuable in its own right as it offers a tool to study the effect of time discretization schemes when evaluating the continuous solution numerically. Specifically for significantly small \( \Delta t \ll 1 \) the estimate will be reliable within a margin \([1, C]\) with \( C \geq 1 \). We shall demonstrate this empirically in our numerical experiment.

3. Linearly Solvable Optimal Control

In this section we introduce two problems that are rooted in stochastic optimal control which merit from the theorems in the previous section. In particular we will introduce two optimal control problems, either in a continuous be it discrete time setting for which we estimate the value function by sampling paths from the uncontrolled process. This class of control problems is known as LSOC [6]. On account of the Feynman-Kac formula it is possible to estimate the control based on uncontrolled sample paths.

Initiated by the work of Kappen — who first noticed the potential use of this peculiar subclass — this lead to the development of a series of control methods known as Path Integral Control (see [8, 7] and references therein) which has been connected to Entropy Regularized Optimal Control more recently [10]. As far as we are aware of the discounted infinite horizon LSOC setting is untreated by the literature.
3.1. Continuous Time Infinite Horizon

Let us begin our discussion by considering the continuous time discounted infinite horizon stochastic optimal control with control affine dynamics and control quadratic cost rate term.

**Definition 2.** Let \( q : \mathbb{R}^n \to \mathbb{R} \) be a strictly positive function, \( R \in \mathbb{R}^{m \times m} \) a strictly positive definite matrix and \( \alpha \) a strictly positive scalar. The continuous time infinite horizon stochastic optimal control for control affine dynamics and control quadratic cost rate is then defined as

\[
\begin{align*}
    u^* &= \arg \min_u C[u] \\
    &= \arg \min_u \lim_{T \to \infty} \mathbb{E}_{P(X(t \to T)|X(t) = x)} \left[ \int_t^T c(\tau) d\tau \right]
\end{align*}
\]

where \( c(\tau) = e^{-\alpha(\tau-t)} \left( q(X(\tau)) + \frac{1}{2} u(\tau)^\top Ru(\tau) \right) \) and \( X(\tau \geq t)|X(t) = x \) is governed by the control affine diffusion process

\[
dX = (a(X) + B(x)u) dt + \sigma(x)dW.
\]

It is well-known that the optimal control is governed by the following differential equation — also known as the stochastic Hamilton-Jacobi-Bellman (HJB) differential equation. Here \( V \) represents the so called *Value function* which quantifies the accumulated cost when the optimal control is continuously applied from initial state \( X(t) = x \)

\[
\alpha V = \min_u q + \frac{1}{2} u^\top Ru + \nabla_x V^\top (a + Bu) + \frac{1}{2} \text{tr}(\Sigma_{xx})
\]

It is easily verified that the minimizer satisfies

\[
u^* = -R^{-1}B^\top \nabla_x V
\]

Substitution into the original equation then produces the differential equation which we refer to as the optimal HJB equation

\[
\alpha V = q - \frac{1}{2} \nabla_x V^\top BR^{-1}B^\top \nabla_x V + \nabla_x V^\top a + \frac{1}{2} \text{tr}(\Sigma \nabla_{xx} V)
\]

This equation appears to be intractable on account of the quadratic term in \( \nabla_x V \). However a well known trick in physics is to introduce the following Value function transformation

\[
Z(x) = e^{-V(x)}
\]
so that

\[
\nabla_x V = -\frac{1}{2} \nabla_x Z, \quad \nabla_{xx} V = -\frac{1}{2} \nabla_{xx} Z + \frac{1}{2} \nabla_x Z \nabla_x Z^T
\]

and equivalently \( u^* = \frac{1}{2} R^{-1} B^T \nabla_x Z \).

This transformation will allow us to reduce the complexity of the PDE significantly and propose a solution in terms of theorem 1 for a limited problem subclass. Substitution of this intermediary result into the HJB equation reveals that

\[
\alpha Z \log Z = -qZ + \nabla_x Z^T a + \frac{1}{2} \text{tr} (\Sigma \nabla_{xx} Z) + \frac{1}{2} \frac{1}{Z} \nabla_x Z^T B R^{-1} B^T \nabla_x Z - \frac{1}{2} \frac{1}{Z} \nabla_x Z^T \Sigma \nabla_x Z
\]

Hence by choosing \( \Sigma = BR^{-1} B^T \) the last two terms cancel out and the nonlinear PDE treated in theorem 1 emerges

\[
\alpha Z \log Z = -qZ + \nabla_x Z^T a + \frac{1}{2} \text{tr} (\Sigma \nabla_{xx} Z)
\]

Thus by consequence it follows that

\[
Z(x) = \lim_{T \to \infty} \mathbb{E}_{P(X(t \to T)|X(t) = x)} [I(T)]
\]

where \( X(\tau \geq t)|X(t) = x \) is governed by the uncontrolled diffusion process

\[
dX = a(X)dt + \sigma(X)dW.
\]

This remarkable result implies that we can estimate the Value function from sample paths taken from the uncontrolled diffusion process and since the optimal control is expressed in terms of the Value function, subsequently also the policy.

### 3.2. Discrete Time Infinite Horizon

Analogous results can be established in a discrete time setting. Here the associated optimal control problem is known as Kullback-Leibler control or discrete time Linearly Solvable Optimal Control. As will soon prove to be the case, the discrete time stochastic optimal control cannot be solved directly by introducing a control quadratic cost rate and assuming control affine dynamics. Instead one introduces an information-geometric cost that penalizes the control indirectly \[9\]. Specifically the quadratic control cost is replaced by the Kullback-Leibler divergence between the generic controlled transition probability \( P_\mu \) and the uncontrolled transition probability \( P \) that is obtained for zero control.
We will show that for control affine dynamics this reduces effectively to an equivalent quadratic control cost.

\[
\mathbb{D}[P_u(X'|X = x)||P(X'|X = x)] = \mathbb{E}_{P_u}(X'|X = x) \left[ \log \frac{P_u(X'|X = x)}{P(X'|X = x)} \right]
\]

The discrete time infinite horizon stochastic optimal control for control affine dynamics and Kullback-Leibler control cost rate is established rigorously in the following definition.

**Definition 3.** Let \( q : \mathbb{R}^n \mapsto \mathbb{R} \) be a strictly positive function and \( \alpha \) a strictly positive scalar. The discrete time infinite horizon stochastic optimal control for control affine dynamics and Kullback-Leibler control cost rate is then defined as

\[
u^* = \arg\min_u C[u]
\]

\[
= \arg\min_u \lim_{N \to \infty} \mathbb{E}_{P_u(X(n \to N)|X(n) = x)} \left[ \sum_{m=n}^{N} c(m) \right]
\]

where \( c(m) = \alpha^{m-n} \left( q(X(m)) + \log \frac{P_u(X(m+1)|X(m))}{P(X(m+1)|X(m))} \right) \) and the transition probability is governed by the control affine random walk \( (\Delta W \sim \mathcal{N}(0, I)) \) \( X' = X + \Delta X, \Delta X = a(X) + B(x)u + \sigma(X)\Delta W. \)

First note that for a control affine random walk the Kullback-Leibler cost rate simplifies to a quadratic control cost rate. Hence it follows that an equivalent relation between the stochastic process and the control cost rate, which allowed us to solve the stochastic optimal control problem in continuous time, is inherent to the Kullback-Leibler control setting.

\[
\mathbb{D}[P_u(X'|X = x)||P(X'|X = x)] = \frac{1}{2} u^T \underbrace{B^T \Sigma^{-1} B}_{R} u + c
\]

Further it can be shown that the problem in definition 3 is equivalent to the following recursive optimization problem. Again \( V \) represents the Value function which is subject to a similar interpretation as in continuous time. In a generic setting this equation is referred to as the stochastic Bellman equation.

\[
V(x) = \min_{P_u \in \mathcal{P}} \mathbb{E}_{P_u(X'|X = x)} \left[ q(x) + \log \frac{P_u(X'|X = x)}{P(X'|X = x)} + \alpha V(X') \right]
\]

13
Specific to the discrete time Linearly Solvable Optimal Control setting is that the problem can be lifted from the control space to the controlled transition probability space, that is \( P_u(X' = y | X = x) \in \mathcal{P} \) where \( \mathcal{P} \) represents the space of all transition probabilities on \( \{ X' = y | X = x \} \) defined as \( \mathcal{P} = \{ P(X' = y | X = x) | \int P(X' = y | X = x) dy = 1 \} \). Hence the problem below can no longer be treated as an optimization problem but as a variational one.

We can solve this problem by introducing the Lagrangian

\[
L[P_u, \lambda] = \mathbb{E}_{P_u} \left[ q + \log \frac{P_u}{\mathbb{E}} + \alpha V' \right] + \lambda (\mathbb{E}_{P_u} [1] - 1)
\]

so that the optimal transition probability is governed by

\[
\partial_{P_u} L = \int (\log \frac{P_u}{\mathbb{E}} + \alpha V' + \lambda + 1) dX' = 0
\]

which is identically equal to zero if the integrand is, and so

\[
P^*_u(X' = y | X = x) \propto P(X' = y | X = x) e^{-\alpha V(y)}
\]

where the relation \( \propto \) implies that the left-hand side and right-hand side are equivalent up to a normalization constant which on its turn depends on the Lagrangian multiplier \( \lambda \).

Substitution of the optimal transition probability function into the Bellman equation reveals then that the Value function is governed by the following stochastic recurrence relation

\[
V(x) = q(x) - \log \mathbb{E}_{P(X'|X=x)} \left[ e^{-\alpha V(X')} \right]
\]

Reintroducing the Value function transformation \( Z(x) = e^{-V(x)} \) then produces

\[
Z(x) = e^{-q(x)} \mathbb{E}_{P(X'|X=x)} [Z(X')^\alpha]
\]

and by consequence of theorem \( 2 \) it then also follows that

\[
\tilde{Z}(x) = \lim_{N \to \infty} \mathbb{E}_{P(X(n \to N)|X(n) = x)} [S(N)]
\]

Equivalently we have that

\[
P^*_u(X' = y | X = x) \propto P(X' = y | X = x) Z(y)^\alpha
\]

\[
= \frac{P(X' = y | X = x) Z(y)^\alpha}{\mathbb{E}_{P(X'|X=x)} [Z(X')^\alpha]}
\]

14
These results determine a straightforward discrete time version of the continuous time optimal control discussed in the previous paragraph. On account of theorem 2, unfortunately here we can only approximate $Z(x)$ and hence the performance of the derived policy will deteriorate in practice.

The symmetry is completed by noting that we can abstract the optimal control for control affine dynamics remarking that

$$
\mathbb{E}_{P_u(x'|X=x)}[X'] = a(x) + B(x)u^*(x)
= \frac{\mathbb{E}_{P(X'|X=x)}[(a(x) + \sigma(x)\Delta W) Z(X')^a]}{\mathbb{E}_{P(X'|X=x)}[Z(X')^a]}
$$

so that

$$
u^*(x) = R^{-1} B^\top \Sigma^{-1} \frac{\mathbb{E}_{P(X'|X=x)}[\sigma(x)\Delta W Z(X')^a]}{\mathbb{E}_{P(X'|X=x)}[Z(X')^a]}
$$

The same result follows from the following projection strategy

$$
u^*(x) \approx \arg \min_u \mathbb{D} [P_u^*(X'|X = x)||P_u(X'|X = x)]
$$

3.3. Note on the transformation and solution scaling

Both in a continuous and discrete time setting the control problem is subject to a constraint relating the diffusion covariance matrix $\Sigma$ with the control penalty matrix $R$

$$
\Sigma = BR^{-1}B^\top \text{ or } R = B^\top \Sigma^{-1}B
$$

This constraint implies that the diffusion process and the admissible control are inherently balanced and the control designer cannot choose a different control weighing. Although the constraint makes sense from a control engineering perspective — since the larger the uncertainty is, the higher the admissible control can be — it poses a limiting restriction that practitioners would like to see resolved. The issue can be remedied easily by scaling the solely state dependent cost rate term $q$. Alternatively we could reconsider the Value function transformation. Specifically we could also have used the parametrized transformation $Z = e^{-\frac{1}{2}V}$. If one repeats the continuous time derivation substituting this transformation rather than the unparametrized transformation, we find

$$
\Sigma = \gamma BR^{-1}B^\top \text{ or } R = \gamma B\Sigma^{-1}B^\top
$$
which allows to effectively scale the relation. The discrete time constraint is obtained by weighting the Kullback-Leibler control cost rate with $\gamma$.

Finally note that in this case the solution is adapted to $Z(x) = \lim_{T \to \infty} \mathbb{E}_{P(X(t\to T)|X(t)=x)} \left[ I(T)^{\frac{1}{2}} \right]$ in the continuous time setting and $Z(x) = \lim_{N \to \infty} \mathbb{E}_{P(X(n\to N)|X(n)=x)} \left[ S(N)^{\frac{1}{2}} \right]$ in the discrete time setting.

This effectively illustrates that the parametrized value function transformation is equivalent to a rescaling of the state dependent control cost rate term $q$.

4. Example

In this section we demonstrate the proposed solution for a small case study and verify the influence of some parameters empirically. The system under study is the controlled van der Pol equation $\ddot{y} - 2\zeta\omega(1-y^2)\dot{y} + \omega^2 y = u$. Here we assume that the source of uncertainty operates solely in the input space, i.e. $dX = (a(X) + Bu)dt + B\sigma dW$ with $dW \sim \mathcal{N}(0, I)$ or equivalently $dX = (a(X) + Bu)dt + dW'$ where $dW' \sim \mathcal{N}(0, \Sigma), \Sigma = \sigma^2 BB^\top$. We choose $\omega = 1$, $\zeta = \frac{1}{4}$ and $\sigma^2 = \frac{1}{2}$. A discounted infinite horizon stochastic optimal control problem is defined with state penalty rate $q = \frac{1}{2} \|x\|^2$ and discount factor $\alpha = 1$. The conversion of this system to a state-space representation is trivial. In our simulations we use a discretised version of the continuous dynamics. The discretization is performed as in the proof of theorem 2 where we choose $\Delta t = 1 \times 10^{-3}$ which is several orders of magnitude smaller than the system’s time constant.

The goal is to determine the optimal policy. Therefore we first estimate the value function making use of the nonlinear Feynman-Kac formula from theorem 1. This is done by determining the value function over a sample grid of initial values (see figure 1). We approximate the associated function value using a Monte Carlo estimate of the expectation using $M$ sample paths and truncate the cost after time $T$. For sufficiently large $M$ and $T$ and small $\Delta t$ the estimate should be exact. We then approximate the surface using a Gaussian Process and determine the optimal policy as detailed in section 3.1.
Figure 1 illustrates the estimation procedure for a single coordinate, here \( x = (2, 2) \). The parameter \( \gamma \) is varied over the set \( \{4, 1, \frac{1}{3}\} \). Note that the same set of experiments can be used for every value of \( \gamma \). Figures 2 and 3 demonstrate the \( Z, V \) and policy response function estimation for \((T, M) = (1, 25)\) and \((T, M) = (25, 500)\) respectively. In figure 4 the effect of \( \gamma \) is illustrated on the estimates of \( Z \). Finally figure 5 illustrates the stabilizing effect of the policy on the dynamic landscape governing the van der Pol equation.

5. Conclusion

In this contribution we present a solution to a nonlinear PDE closely related to the parabolic PDE studied by Feynman and Kac. Our results have applications in Linearly Solvable Optimal Control. In particular our work extends the framework to incorporate the discounted infinite horizon setting which remained untreated by the literature.

References

[1] Richard Phillips Feynman. Space-time approach to non-relativistic quantum mechanics. In Feynman’s Thesis—A New Approach To Quantum Theory, pages 71–109. World Scientific, 2005.

[2] Mark Kac. On distributions of certain wiener functionals. Transactions of the American Mathematical Society, 65(1):1–13, 1949.
Figure 2: Solution for \((T, M, \gamma) = (1, 25, 4)\). From left to right: estimate of the function \(Z\), estimate of the function \(V\) and estimate of the optimal policy by approximating and differentiating \(V\) using a Gaussian Process.

Figure 3: Solution for \((T, M, \gamma) = (25, 500, 4)\).

Figure 4: Solution for varying \(\gamma\) and \((T, M) = (15, 250)\)
Figure 5: Illustration of optimal control. *From left to right:* streamline visualization of unactuated dynamic landscape, streamline visualization of actuated dynamic landscape ($\gamma = 4$ versus $\gamma = \frac{1}{4}$ actuated in blue or red respectively) and visualization of actuated stochastic rollouts.

[3] Mark Kac et al. On some connections between probability theory and differential and integral equations. In Proceedings of the second Berkeley symposium on mathematical statistics and probability. The Regents of the University of California, 1951.

[4] Brian C Hall. Quantum theory for mathematicians. Springer, 2013.

[5] Paolo Brandimarte. Numerical methods in finance and economics: a MATLAB-based introduction. John Wiley & Sons, 2013.

[6] Hilbert Kappen. Linear theory for control of nonlinear stochastic systems. Physical review letters, 95(20):200201, 2005.

[7] Evangelos Theodorou, Jonas Buchli, and Stefan Schaal. Reinforcement learning of motor skills in high dimensions: A path integral approach. In 2010 IEEE International Conference on Robotics and Automation, pages 2397–2403. IEEE, 2010.

[8] Dominik Thalmeier, Hilbert Kappen, Simone Totaro, and Vicenç Gómez. Adaptive smoothing for path integral control. Journal of Machine Learning Research, 21(191):1–37, 2020.

[9] Dvijotham Krishnamurthy and Todorov Emanuel. A unifying framework for linearly solvable control. In Proceedings of the Twenty-Seventh Conference on Uncertainty in Artificial Intelligence, UAI’11, page 179–186, Arlington, Virginia, USA, 2011. AUAI Press.
[10] Tom Lefebvre and Guillaume Crevecoeur. On entropy regularized path integral control for trajectory optimization. *Entropy, 22*(10):1120, 2020.