Feynman-Kac formula: Group project

K. Kowalczuk, M. Majocha, A. Mirecki, P. Schmidt, M. Napiórkowski July 30, 2025

Contents

Introduction

The Feynman-Kac formula represents one of the most elegant and profound bridges between the fields of deterministic analysis and probabilistic theory. At its core, the formula connects certain partial differential equations (PDEs) with expectations over stochastic processes—most notably, Brownian motion. Its historical development is deeply rooted in the evolution of both quantum mechanics and probability theory in the 20th century, reflecting the interplay between physics and mathematics.

The story begins with Richard Feynman, a theoretical physicist who, in the 1940s, introduced a new formulation of quantum mechanics known as the path integral formulation. In contrast to the operator-based approaches of Schrödinger and Heisenberg, Feynman's formulation described the evolution of quantum systems as an infinite sum (integral) over all possible classical paths a particle could take. Each path was weighted by a complex exponential involving the classical action functional. Though conceptually revolutionary, this method was formal and lacked rigorous mathematical underpinnings, particularly because path integrals, as envisioned by Feynman, were difficult to define within conventional mathematical frameworks.

Meanwhile, in the realm of mathematics, Mark Kac, a prominent figure in probability theory, was exploring the connections between stochastic processes and differential equations. In 1949, Kac developed a method to interpret certain path integrals probabilistically. He observed that expectations over Brownian motion—random continuous paths—could provide a rigorous foundation for expressions resembling Feynman's formal integrals, at least in the context of diffusion-type problems. His results led to a probabilistic representation of the solution to certain linear parabolic partial differential equations, now famously known as the Feynman-Kac formula.The significance of the Feynman-Kac formula extends far beyond its original motivations. In mathematical physics, it provided a bridge between quantum

mechanics and diffusion processes. In mathematical finance, it became central to the pricing of financial derivatives, where asset prices are modeled by stochastic differential equations and option pricing problems are reduced to solving PDEs. In control theory and statistical mechanics, the formula helped unify diverse methods of analysis.

Over the decades, the formula has inspired numerous extensions and generalizations. These include versions involving stochastic differential equations with jumps (e.g., Lévy processes), infinite-dimensional systems such as those arising in quantum field theory, and nonlinear analogues appearing in the study of backward stochastic differential equations (BSDEs). Nonetheless, the original idea remains both conceptually and practically important: that solutions to PDEs can often be interpreted as expectations over random paths.

Thus, the Feynman-Kac formula is not merely a technical tool, but a conceptual synthesis of ideas from physics, probability, and analysis. It illustrates how probabilistic methods can provide deep insight into deterministic problems and vice versa—a theme that has become increasingly important in modern applied mathematics.

1 Physical interpretation

1.1 Time evolution operator and the titular formula

This section is mainly based on [1]. Let $H = -\Delta + V$ be the quantum Hamiltonian, where V is a potential and $-\Delta$ is the Laplacian. Let us assume that $V \in L^2(\mathbb{R}) + L^\infty(\mathbb{R})$, that is, V can be written as a sum of a square-integrable function and a bounded function (it is not essential to understand the equation, but it is important in the proof). The solution for the Schrödinger equation for such Hamiltonian is equal to:

$$\psi = e^{-it(-\Delta+V)}\psi_0.$$

Everything's great, but what actually is $e^{-it(-\Delta+V)}$? Here comes the "Feynman-Kac" formula:

Theorem 1 [Feynman-Kac formula no. 1]: Let $H = -\Delta + V$ be defined as above. Then:

$$(e^{-itH}\psi)(x_0) = \lim_{n\to\infty} \int_{\mathbb{R}^3} \dots \int_{\mathbb{R}^3} e^{iS_n(x_0,\dots,x_n,t)}\psi(x_n)dx_n \dots dx_1, \tag{1}$$

where:

$$S_n(x_0, x_1 \dots x_{n-1}, x_n, t) = \sum_{i=1}^{n} \frac{t}{n}\left[\frac{1}{4}\left(\frac{|x_i - x_{i-1}|}{t/n}\right)^2 - V(x_i)\right].$$

Looks quite confusing, but it has a meaningful physical interpretation, which we'll discuss next.

1.2 Physics behind it – path integrals

Let the particle with mass m move along the curve $\gamma : I \longrightarrow \mathbb{R}^3$. From Lagrangian mechanics we know we can define the **action** to be:

$$S = \int_0^t \left(\frac{m}{2}|\dot{\gamma}(t)|^2 - V(\gamma(t))\right) \mathrm{d}t.$$

Briefly speaking, the particle chooses the path that extremalizes the action. Let $m = \frac{1}{2}$. If we assume that the particle moves along the path that consists of finitely many straight lines, and on each of them the velocity is constant and the particle spends equal amount of time on every one of them, the action is described by:

$$S = \sum_{i=1}^{n} \frac{1}{4}\left(\frac{t}{n}\right)\left(\frac{|x_i - x_{i-1}|}{t/n}\right)^2 - \int_0^t V(\gamma(t))\mathrm{d}t.$$

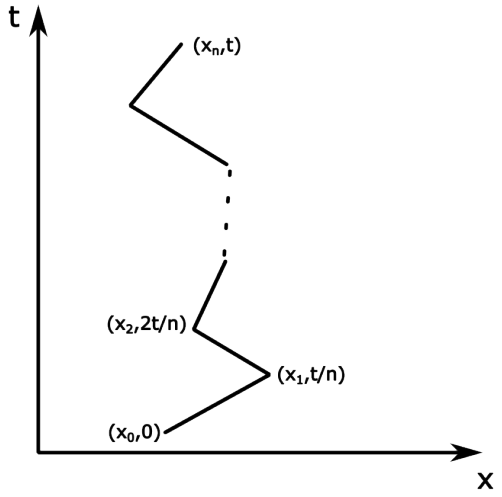

Figure 1: Path of the particle

Therefore, if we approximate the potential, we get:

$$S = \sum_{i=1}^{n} \frac{t}{n} \left[\frac{1}{4} \left(\frac{|x_i - x_{i-1}|}{t/n} \right)^2 - V(x_i) \right].$$

This formula is exactly the same as $S(x_0, x_1, \ldots, x_n, t)$. Thus, we can say that the integral

$$\int_{\mathbb{R}^3} \ldots \int_{\mathbb{R}^3} e^{iS_n(x_0,\ldots,x_n,t)} \psi(x_n) dx_n \ldots dx_1$$

corresponds to taking the integral of $e^{iS_n}\psi$ along all possible polygonal paths.

There exists an amazing interpretation of this equation. To obtain it, first one has to define **the kernel** of an operator.

Definition We say that $\mathscr{K}(t, t'; x, x')$ is a kernel of $e^{-i(t-t')H}$ if:

$$(e^{-i(t-t')H}\psi)(x) = \int_{\mathbb{R}} \mathscr{K}(x, x'; t, t')\psi(x')dx' \ .$$

From now on we will denote $\mathscr{K}(t, x, x') \equiv \mathscr{K}(t, 0; x, x')$. Because $e^{-i(t_1+t_2)H} = e^{-it_1H}e^{-it_2H}$, we get that

$$(e^{-i(t_1+t_2)H}\psi)(x) = e^{-it_1H}\left(\int_{\mathbb{R}^3} \mathscr{K}(t_2, x, x')\psi(x')dx' \right) = \int_{\mathbb{R}^3}\int_{\mathbb{R}^3} \mathscr{K}(t_1, x, x'')\mathscr{K}(t_2, x'', x')\psi(x')dx'dx''$$

and on the other hand

$$(e^{-i(t_1+t_2)H}\psi)(x) = \int_{\mathbb{R}^3} \mathscr{K}(t_1 + t_2, x, x')\psi(x')dx',$$

therefore

$$\mathscr{K}(t_1 + t_2, x, x') = \int_{\mathbb{R}^3} \mathscr{K}(t_1, x, x'')\mathscr{K}(t_2, x'', x')dx''.$$

Analogously, it is easy to show that

$$\mathscr{K}(t, x_0, x_{n+1}) = \int_{\mathbb{R}^3} \ldots \int_{\mathbb{R}^3} \prod_{j=i}^{n+1} \mathscr{K}(t_j - t_{j-1}, x_{j-1}, x_j)dx_n \ldots dx_1, \tag{2}$$

where $t_0 = 0, t_{n+1} = t$ and $0 < t_1 < \ldots < t_n < t$.

From this property, we can conclude that e^{iS_n} behaves like a kernel of time evolution. The kernel can be interpreted as a weight that tells us how important each of the previous positions of a state is. The stationary action principle suggests that a kernel should be larger for points near the stationary action

curve and smaller for points far from it. So maybe we can take all the curves and assign a proper weight for each of them? e^{iS_n} is a great candidate for this! The Feynman's idea was that

$$\mathscr{K}(t, x_0, x) \sim \int_{\Gamma_{x_0}} e^{iS(\gamma)/\hbar} d\gamma,$$

where Γ_{x_0} is the set of all curves that start at x_0. We have added $\hbar$ to the equation to give the Hamiltonian the proper unit ($\hbar$ is present in the Schrödinger equation but it was neglected previously to shorten the equations). $\hbar$ is extremely small compared to other quantities, so any path that diverge from the stationary path is canceled by the path that is really close to it because $e^{iS/\hbar}$ changes rapidly. If the curve is close to the stationary path though the phase factor is negligible and the factors are adding up instead of canceling. This is exactly what we wanted! It also gives us a powerful intuition: the particle is moving along all possible curves but only those that are close to the stationary path are interfering constructively and the rest of them cancel each other out.

Unfortunately, the equation above isn't well defined and there are some other problems. Luckily, there exist very similar formula where the 'i' is missing. In this case we get:

Theorem 2 [Feynman-Kac formula no. 2]
Let $\Gamma(0, t; x)$ be a set of all curves that begin at x for $t = 0$. Then

$$(e^{-tH}\psi)(x) = \int_{\Gamma(0,t;x)} e^{-\int_0^t V(\gamma(s))ds} \psi(\gamma(t))\, \mathrm{d}W_x^t(\gamma),$$

where W_x^t is the Wiener measure defined in the next section.

2 Functional approach to the Wiener measure

In this section we'll define $W_{x,x'}^t$ and study the consequences.

2.1 Solving the heat equation

Let $u(x, t)$ be a temperature function. Heat equation tells us that for some $\kappa \geq 0$

$$\partial_t u + \kappa \Delta u = 0.$$

There are multiple ways to derive it. We'll do so by assuming

$$\phi(x, t) = -k \nabla u(x, t)$$

for some k, where ϕ is the heat flux. Assuming that heat is proportional to temperature, we have

$$\int_{\partial V} (\phi | \hat{n}) = \nu \int_V \partial_t u \ ,$$

for some ν. However

$$\int_{\partial V} (\phi | \hat{n}) = \int_V \mathrm{div}\phi,$$

so we get that

$$\mathrm{div}\phi - \nu \partial_t u = 0 \ .$$

By substituting the expression for ϕ we obtain the heat equation. We will find a solution for it. Without loss of generality $\kappa = \frac{1}{2}$. Let $u_0 := u(\cdot, 0)$. We can see that this equation is analogous to the Schrödinger equation (but without i). We can write a solution as

$$u(\cdot, t) = e^{-tH_0} u_0,$$

where $H_0 = -\frac{1}{2}\Delta$. Now we'd like to find a way to write the solution in a more meaningful way (because right now it tells us nothing). To do so, we'll define a kernel of the operator

Definition 1 Let $\mathbb{R} \ni t \longmapsto U(t) \in \mathscr{B}(L_2(\mathbb{R}^3))$ be a group homomorphism. Then we say that $\mathscr{K}(t, \cdot, \cdot)$ is a kernel of $U(t)$ if $\forall_{\psi \in L_2(\mathbb{R})}\ \forall_{t,x \in \mathbb{R}}$

$$(U(t)\psi)(x) = \int_{\mathbb{R}^3} \mathscr{K}(t, x, x')\psi(x')dx'$$

and denote:

$$\ker(U(t)) := \mathscr{K}(t,\cdot,\cdot)\ .$$

Now let us find $\ker(e^{-tH_0})$. We will denote the unitary Fourier transform by $\mathcal{F}$. Using its elementary properties one can check that

$$-\Delta = \mathcal{F}^{-1}p^2\mathcal{F},$$

where p^2 - multiplication by $\|p\|^2$. Therefore we have:

$$(e^{-tH_0}\psi)(x) = \left(\mathcal{F}^{-1}\exp\left[\frac{-tp^2}{2}\right]\mathcal{F}\psi\right)(x) =$$

$$= \frac{1}{(2\pi)^3}\int_{\mathbb{R}^3\times\mathbb{R}^3} e^{ip\cdot(x-x')}e^{-t\frac{\|p\|^2}{2}}\psi(x')\mathrm{d}p\,\mathrm{d}x' = \frac{1}{(2\pi)^3}\int_{\mathbb{R}^3}\left(\int_{\mathbb{R}^3} e^{-t\frac{\|p\|^2}{2}+ip\cdot(x-x')}\mathrm{d}p\right)\psi(x')\,\mathrm{d}x' =$$

$$\frac{1}{(2\pi)^3}\int_{\mathbb{R}^3}\left(\int_{\mathbb{R}^3} e^{-\frac{t}{2}\left\|p-i\frac{x-x'}{t}\right\|^2}\mathrm{d}p\right)e^{-\frac{1}{2t}\|x-x'\|^2}\psi(x')dx'\ .$$

It is then straightforward to show that:

$$\int_{\mathbb{R}^3} e^{-\frac{t}{2}\left\|p-i\frac{x-x'}{t}\right\|^2}\mathrm{d}p = \int_{\mathbb{R}^3} e^{-\frac{t}{2}\|p\|^2}\mathrm{d}p = \left(\frac{2\pi}{t}\right)^{3/2},$$

and because of it

$$(e^{-tH_0}\psi)(x) = \frac{1}{(2\pi t)^{3/2}}\int_{\mathbb{R}^3} e^{-\frac{1}{2t}\|x-x'\|^2}\psi(x')dx'\ ,$$

and thus finally we get

$$\ker(e^{-tH_0})(x,x') = \left(\frac{1}{2\pi t}\right)^{3/2} e^{-\frac{1}{2t}\|x-x'\|^2}\ .$$

2.2 Wiener measure (Heuristically)

Before we start, we have to understand what we want to achieve. From the previous subsection we know that

$$(e^{-tH_0}\psi)(x) = \int_{\mathbb{R}^3}\mathscr{K}(t,x,x')\psi(x')\mathrm{d}x'\ ,$$

where $\mathscr{K}(t,x,x') = \ker(e^{-tH_0})(x,x') = (2\pi t)^{-3/2}\exp\left(-\frac{\|x-x'\|^2}{2t}\right)$. We strive to construct a measure with the following interpretation: the value of ψ in x after time t is 'a weighted average' of values of ψ on each curve that start at this point. Moreover, the weight should be determined by $\mathscr{K}(t,\cdot)$ from the previous equation. One could write it as

$$(e^{-tH_0}\psi)(x) = \int_{\text{'curves staring at } x\text{'}}\psi(\omega(t))\,\mathrm{d}\omega\ .$$

The hint is given by the following property

$$\mathscr{K}\left(t,x_0,x_{n+1}\right) = \int_{\mathbb{R}^3}\ldots\int_{\mathbb{R}^3}\prod_{j=i}^{n+1}\mathscr{K}(t_j-t_{j-1},x_{j-1},x_j)dx_n\ldots dx_1\ , \tag{3}$$

where $t_0 = 0, t_{n+1} = t$ and $0 < t_1 < \ldots < t_n < t$. One can prove it analogously to the proof in the first section. Additionally, it is easy to check that

$$\int_{\mathbb{R}^3}\mathscr{K}(t,x,x')\mathrm{d}x' = 1\ .$$

Now let us denote a set of all curves $[0,t]\ni t\longmapsto\gamma(t)\in\mathbb{R}^3$ such that $\gamma(0)=x$ by $\Gamma(0,t;x)$. We can define a measure on this set as follows

Definition 2 Let

$$\mathcal{C}(t_1,t_2,\ldots,t_n,I_1,\ldots,I_n) := \left\{\gamma\in\Gamma(0,t;x)\mid\forall_{i\in\{1,\ldots n\}}\ \gamma(t_i)\in I_i\right\}$$

and

$$\mathcal{C} = \left\{ \mathcal{C}(t_1, \ldots, t_n, I_1, \ldots, I_n) \mid n \in \mathbb{N},\ 0 < t_1 < \ldots < t_n < t \text{ and } \forall_{i \in \{1,\ldots n\}}\ I_i \subseteq \mathbb{R}^3 \right\}\ ;$$

we define

$$w_x^t : C \ni \mathcal{C}(t_1, t_2, \ldots, t_n, I_1, \ldots, I_n) \longmapsto \int_{I_1} \ldots \int_{I_n} \prod_{j=i}^{n} \mathscr{K}(t_j - t_{j-1}, x_{j-1}, x_j) dx_n \ldots dx_1\ ,$$

where $x_0 = x$.

It is possible to extend w_x^t to all Borel sets, such that the extension will be a measure.

Theorem 2 There exists an unique extension of measure w_x^t to all Borel sets of Γ

$$W_x^t : \beta(\Gamma(0, t; x)) \longrightarrow \mathbb{R}^+\ .$$

This measure is called **Wiener measure** or **Wiener conditional measure**.

The theorem is left without a proof, however we'll discuss how one can approach it later on in this section. First of all, let us see what are the properties of this measure.

Properties of W_x^t

1. $\int_{\Gamma(0,t;x)} \mathrm{d}W_x^t(\gamma) = 1$,
2. $\int_{\Gamma(0,t;x)} f(\gamma(t))\, \mathrm{d}W_x^t(\gamma) = \int_{\mathbb{R}^3} f(x') \mathscr{K}(t, x, x') dx'$,
3. For $0 < t_1 < \ldots < t_n < t$

$$\int_{\Gamma(0,t;x)} F(\gamma(t_1), \ldots, \gamma(t_n))\, \mathrm{d}W_x^t(\gamma) =$$
$$= \int_{(\mathbb{R}^3)^{\times n}} F(x_1, \ldots, x_n) \mathscr{K}(t_1, x, x_1) \mathscr{K}(t_2 - t_1, x_1, x_2) \ldots \mathscr{K}(t - t_n, x_{n-1}, x_n) dx_1 \ldots \mathrm{d}x_n\ .$$

Proof [Sketch]

1.

$$\int_{\Gamma(0,t;x)} \mathrm{d}W_x^t = W_x^t(\Gamma(0, t; x)) = \int_{\mathbb{R}^3} \mathscr{K}(t, x, x')\, \mathrm{d}x' = 1$$

2. We will show it for a simple function $f(x) = \sum_{j=1}^{k} c_j \chi_{A_j}(x)$, where $A_j \subseteq \mathbb{R}^3$ and $c_j \in \mathbb{C}$. We have

$$\int_{\Gamma(0,t;x)} f(\gamma(t))\, \mathrm{d}W_x^t(\gamma) = \sum_{j=1}^{k} c_j \int_{\Gamma(0,t;x)} \chi_{A_j}(\gamma(t))\, \mathrm{d}W_x^t(\gamma) = \sum_{j=1}^{k} c_j \int_{C(t,A_j)} \mathrm{d}W_x^t(\gamma) =$$
$$= \sum_{j=1}^{k} c_j \int_{A_j} \mathscr{K}(t, x, x') \mathrm{d}x' = \int_{\mathbb{R}^3} \left(\sum_{j=1}^{k} c_j \chi_{A_j}(x') \right) \mathscr{K}(t, x, x') \mathrm{d}x' = \int_{\mathbb{R}^{\not 3}} f(x') \mathscr{K}(t, x, x') \mathrm{d}x'\ .$$

One can conclude that it should follow for the wider families of functions by approximating them by simple functions.

3. It can be shown analogously to 2. (One can take $f(x_1, \ldots, x_n) = \sum_{j_1}^{k_1} \cdots \sum_{j_n}^{k_n} \left(\prod_{s=1}^{n} c_{j_s} \chi_{A_{j_s}}(x_s) \right)$.

□

Equipped with this measure we can prove Theorem 2. Before we do it, we need just one more thing. According to the first Feynman-Kac formula, we have that:

$$\left(e^{-tH}\psi\right)(x) = \lim_{m \to \infty} \int_{\mathbb{R}^3} \ldots \int_{\mathbb{R}^3} \exp\left(-\frac{t}{m} \sum_{j=1}^{m} V(x_j) \right) \psi(x_m) \mathscr{K}\left(\frac{t}{m}, x, x_1\right) \mathscr{K}\left(\frac{t}{m}, x_1, x_2\right) \ldots$$
$$\ldots \mathscr{K}\left(\frac{t}{m}, x_{m-1}, x_m\right) \mathrm{d}x_1 \ldots \mathrm{d}x_n \quad (4)$$

Now we are ready for the proof:

Proof of Theorem 2:
We will prove the theorem assuming V is continuous and has compact support. Using (4) and the 3rd property for

$$F(x_1,\ldots,x_m)=\exp\left(-\frac{t}{m}\sum_{j=1}^{m}V(x_j)\right)\psi(x_m)$$

we get that:

$$\left(e^{-tH}\psi\right)(x)=\int_{\Gamma(0,t;x)}\exp\left[-\frac{t}{m}\sum_{j=1}^{m}V\left(\gamma\left(\frac{jt}{m}\right)\right)\right]\psi(\gamma(t))\,\mathrm{d}W_x^t(\gamma)\ .$$

For continuous functions with a compact support, the Riemann sums are converging to the integrals, so:

$$\lim_{m\to\infty}\left[-\frac{t}{m}\sum_{j=1}^{m}V\left(\gamma\left(\frac{jt}{m}\right)\right)\right]=\int_0^t V(\gamma(s))\,\mathrm{d}s\ .$$

Additionally, exp is continuous, so almost everywhere:

$$\lim_{m\to\infty}\left\{\exp\left[-\frac{t}{m}\sum_{j=1}^{m}V\left(\gamma\left(\frac{jt}{m}\right)\right)\right]\psi(\gamma(t))\right\}=e^{\int_0^t V(\gamma(s))\,\mathrm{d}s}\psi(\gamma(t))\ .$$

Now because:

$$\exp\left[-\frac{t}{m}\sum_{j=1}^{m}V\left(\gamma\left(\frac{jt}{m}\right)\right)\right]\psi(\gamma(t))\le e^{\|V\|_\infty t}\psi(\gamma(t))\ ,$$

we get

$$\int_{\Gamma(0,t;x)}e^{\|V\|_\infty t}\psi(\gamma(t))\,\mathrm{d}W_x^t(\gamma)\le e^{\|V\|_\infty t}\int_{\mathbb{R}^3}\mathscr{K}(t,x,x')\psi(x')\,\mathrm{d}x'<\infty\ ,$$

and finally, the Dominated Convergence Theorem yields:

$$\left(e^{-tH}\psi\right)(x)=\int_{\Gamma(0,t;x)}\exp\left(\int_0^t V(\gamma(s))\,\mathrm{d}s\right)\psi(\gamma(t))\,\mathrm{d}W_x^t(\gamma)$$

□

3 Trotter product formula

3.1 Lie product formula

To make our way into understanding the Trotter product formula easier, it will be helpful to consider the proof of Lie product formula, which is essentially the same equation - but instead of operators we will be working with finite matrices A and B.
Lie product formula:

$$\exp(A+B)=\lim_{n\to\infty}\left[\exp\left(\frac{A}{n}\right)\exp\left(\frac{B}{n}\right)\right]^n \tag{5}$$

Proof:
Let $S_n:=\exp\left[\frac{1}{n}(A+B)\right]$, $T_n:=\exp\left(\frac{A}{n}\right)\exp\left(\frac{B}{n}\right)$, using identity: $a^n-b^n=\sum_m^{n-1}a^m(a-b)b^{n-1-m}$ (which is true for numbers, as well as matricies and operators, since it does not mess with the multiplication order, and can be easily proven using a straightforward induction argument),
we can write:

$$\begin{gathered}\|S_n^n-T_n^n\|=\left\|\sum_{m=0}^{n-1}S_n^m(S_n-T_n)T_n^{n-1-m}\right\|\le\|S_n-T_n\|\cdot\left\|\sum_{m+0}^{n-1}S_n^mT_n^{n-1-m}\right\|\le\\ \le\|S_n-T_n\|\cdot n\cdot\left(\max\left\{\|T_n\|,\|S_n\|\right\}\right)^{n-1}\end{gathered} \tag{6}$$

If the last inequality is confusing, one may consider the following argument, which can be easily understood by introducing a new matrix M_n, defined as such: $\|M_n\|=\max\left\{\|T_n\|,\|S_n\|\right\}$:

$$\left\|\sum_{m=0}^{n-1}S_n^mT_n^{n-1-m}\right\|\le\left\|nM_n^{n-1}\right\|=n\left\|M_n^{n-1}\right\|\le n\|M_n\|^{n-1}\equiv n\left(\max\left\{\|T_n\|,\|S_n\|\right\}\right)^{n-1}$$

Now, knowing from where last term in 6 came, let us have a look at what we can do with it:

- for $\|S_n\| \leq \|T_n\|$:

$$(\max\{\|T_n\|, \|S_n\|\})^{n-1} = \|T_n\|^{n-1} = \left\|\exp\left(\frac{A}{n}\right)\exp\left(\frac{B}{n}\right)\right\|^{n-1} \leq \left[\left\|\exp\left(\frac{A}{n}\right)\right\| \cdot \left\|\exp\left(\frac{B}{n}\right)\right\|\right]^{n-1} \leq$$
$$\leq \left[\exp\left(\frac{n-1}{n}\|A\|\right) \cdot \exp\left(\frac{n-1}{n}\|B\|\right)\right] \leq [\exp(\|A\|) \cdot \exp(\|B\|)]$$

- for $\|S_n\| \geq \|T_n\|$:

$$(\max\{\|T_n\|, \|S_n\|\})^{n-1} = \|S_n\|^{n-1} = \left\|\exp\left[\frac{1}{n}(A+B)\right]\right\| = \left\|\sum_{m=0}^{\infty}\frac{1}{n^m}\frac{(A+B)^m}{m!}\right\|^{n-1} \leq$$
$$\leq \left[\sum_{m=0}^{\infty}\frac{1}{n^m}\frac{\|(A+B)\|^m}{m!}\right]^{n-1} = \exp\left[\frac{1}{n}\|(A+B)\|\right]$$

Thus, in both cases we can estimate:

$$(\max\{\|T_n\|, \|S_n\|\})^{n-1} \leq \exp\left(\frac{n-1}{n}\|A+B\|\right) \leq \exp(\|A+B\|) \tag{7}$$

Now, let's go back to the 6:

$$\|S_n^n - T_n^n\| = ... \leq n\|S_n - T_n\|\exp(\|A+B\|)$$

Which, in the limit of $n \to \infty$ is equal to 0, since $\|S_n - T_n\|$ behaves as $\frac{C}{n^2}$:

$$\|S_n - T_n\| = \left\|\sum_{m=0}^{\infty}\frac{1}{m!}\left(\frac{A+B}{n}\right)^m - \left(\sum_{m=0}^{\infty}\frac{1}{m!}\left(\frac{A}{n}\right)^m\right)\left(\sum_{m=0}^{\infty}\frac{1}{m!}\left(\frac{B}{n}\right)^m\right)\right\| =$$
$$= \left\|1 + \frac{A+B}{n} + \frac{(A+B)^2}{2n^2} + O(n^{-3}) - \left(1 + \frac{A}{n} + \frac{A^2}{2n^2} + O(n^{-3})\right)\left(1 + \frac{B}{n} + \frac{B^2}{2n^2} + O(n^{-3})\right)\right\| \approx \frac{C}{n^2}$$

3.2 Trotter product formula

Trotter product formula is essentially the same formula as the presented above Lie product formula, but now A and B, instead of finite matricies, are semi-bounded operators on H. In the context of Feynman-Kac formula, we don't need Trotter formula itself, but slightly different formula (that arises when one attempts to prove the Trotter product formula).

Theorem

Let A and B be self-adjoint operators on $\mathcal{H}$ and suppose that $A+B$ is self-adjoint on $D := D(A) \cap D(B)$, then:

$$s\lim_{n\to\infty}\left[e^{it\frac{A}{n}}e^{it\frac{B}{n}}\right] = e^{it(A+B)} \tag{8}$$

Proof

The full proof of the above formula lies beyond the scope of this presentation, however, it will be illuminating to recall all crucial points of the proof, and how its construction resembles the proof of the Lie product formula presented above.
Let us start with defining operator $K(s)$, $s \in \mathbb{R}$:

$$K(s) = \frac{1}{s}\left(e^{isA}e^{isB} - e^{is(A+B)}\right), \tag{9}$$

one can see, that $K(s)\psi \to 0$ as $s \to 0$ for each ψ.
Since $A + B$ is self-adjoint, it follows from Stone's theorem and the uniform boundness theorem that $K(s)$ are uniformly bounded:

$$\|K(s)\psi\| \leq C\|\psi\|_{A+B} \text{ for all } s \in \mathbb{R} \text{ and } \psi \in \mathcal{D}, \tag{10}$$

where norm $\|.\|_{A+B}$ is defined as: $\|\psi\|_{A+B} = \|(A+B)\psi\| + \|\psi\|$, and set $\left\{e^{is(A+B)}\psi | s \in [-1,1]\right\}$ is a $\|.\|_{A+B}$ compact set in D for each ψ.
Crucial part of the proof is the statement that:

$$\underbrace{\frac{1}{t}\left(e^{itA}e^{itB} - e^{it(A+B)}\right)}_{K(s=t)}\underbrace{e^{it(A+B)}\psi}_{\widetilde{\psi}\in D} \to 0 \text{ as } t \to 0, \text{ uniformely for } s \in [-1,1] \tag{11}$$

Now, after mimicking steps in the Lie formula proof, one can prove that norm of the expression:

$$\left[\left(e^{\frac{it}{n}A}e^{\frac{it}{n}B}\right)^n - \left(e^{\frac{it}{n}(A+B)}\right)^n\right]\psi =: \Xi \tag{12}$$

can be estimated by:

$$\|\Xi\| \leq \|t\| \max_{|s|<t}\left|\left|\underbrace{\left(\frac{t}{n}\right)^{-1}\left(e^{\frac{it}{n}(A+B)} - e^{\frac{it}{n}A}e^{\frac{it}{n}B}\right)}_{K(s=\frac{t}{n})}\underbrace{e^{is(A+B)}\psi}_{\widetilde{\psi}\in D}\right|\right|, \tag{13}$$

Since we already argued that $K(s)\psi \to 0$ as $s \to 0$ for all $\psi \in D$, the inequality above completes the proof that $\left(e^{\frac{it}{n}A}e^{\frac{it}{n}B}\right)^n \overset{\psi}{\to} e^{it(A+B)\psi}$ as $n \to \infty$.

4 Probabilistic glossary and the Wiener Process

In the following sections we'll lay the groundwork needed to establish the promised connection between operator theoretic approach and probability theory framework, and in turn explain the Feynman-Kac formula for Stochastic Differential Equations (to which from now on we'll simply refer to as SDEs). To do so, first we need to introduce basic definitions from probability, and the succeeding subsection can be skipped if the reader is already familiar with them. This entire section is based heavily on [2] (unfortunately, available only in Polish).

4.1 Random variables, stochastic processes and martingales

This subsection is rather dry, lacking any examples for the listed definitions (although we provide some succinct motivations and intuitions) – we recommend consulting a good introductory probability theory textbook to fully grasp the subject, but at the same time we hope this overview suffices to understand the rest of the text.

Definition: A **probability space** is a triple $(\Omega, \mathcal{F}, \mathbb{P})$, where

1. Ω is a nonempty set (we call the sample space),
2. $\mathcal{F}$ is a σ-algebra of subsets of Ω,
3. $\mathbb{P} : \mathcal{F} \to [0,1]$ is a measure satisfying $\mathbb{P}(\Omega) = 1$.

It's customary to say that something is true **almost surely** instead of almost everywhere in the probability context. Similarly:

Definition: A **random variable** is just a measurable function

$$X : (\Omega, \mathcal{F}) \longrightarrow (E, \mathcal{E}),$$

i.e. for every set $e \subseteq E$, the preimage $\{\omega \in \Omega : X(\omega) \in e\}$ belongs to $\mathcal{E}$. We call $(E, \mathcal{E})$ the **state space**, and later on we will only consider it to be $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ for some n. Next:

Definition: If $X : (\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}$ is integrable (i.e. $\int_\Omega |X| \, \mathrm{d}\mathbb{P} < \infty$), its **expectation** (or expected value) is simply

$$\mathbb{E}[X] \; = \; \int_\Omega X \, \mathrm{d}\mathbb{P}.$$

For now, we've only introduced new names for already known concepts (expected value of a random variable is it's average, i.e. its integral, and it's an intuitive notion, e.g. expected value for a 6 dice roll

is 3.5). Let's say we wanted some gauge of how "spread out" a given random variable is. We could try to define it as $\mathbb{E}[X - \mathbb{E}[X]$, but (check it!) it's always 0. That's why **variance** is defined in the following way:

Definition: For a random variable that is both integrable and square-integrable, we define its variance as

$$\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2.$$

The last equality is a result of a simple calculation, and we encourage the reader to see it for themselves.

Here's another ,,new" probabilistic concept: let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $X_1, \ldots, X_n \colon \Omega \to (E, \mathcal{E})$ be random variables. They are said to be **(mutually) independent**, if for every choice of measurable sets $A_1, \ldots, A_n \in \mathcal{E}$,

$$\mathbb{P}\big(X_1 \in A_1,\ X_2 \in A_2,\ \ldots,\ X_n \in A_n\big) \ = \ \prod_{i=1}^{n} \mathbb{P}\big(X_i \in A_i\big).$$

Equivalently, the σ–algebras $\sigma(X_i) = \{\, X_i^{-1}(A) : A \in \mathcal{E}\}$ are independent, meaning that for any choice of $G_i \in \sigma(X_i)$,

$$\mathbb{P}\big(\bigcap_{i=1}^{n} G_i\big) = \prod_{i=1}^{n} \mathbb{P}(G_i).$$

The most important thing about this definition is that (in non pathological cases) it aligns with our intuitions (the reader can toy with some simple examples involving, let's say, multiple dice rolls).

For the purposes of the next subsection, we ought to define what's a distribution of a random variable, but we believe that at this level intuitive notions are enough. Instead we'll already steer into the direction of stochastic processes:

Definition: Let T be a totally ordered index set (usually $\mathbb{N}$ or $[0, \infty)$). A **filtration** on $(\Omega, \mathcal{F})$ is a family $\{\mathcal{F}_t\}_{t \in T}$ of sub-σ-algebras of $\mathcal{F}$ such that

$$\mathcal{F}_s \ \subseteq \ \mathcal{F}_t \quad \text{whenever } s \leq t.$$

We will interpret filtration as a ,,bookkeeping" device representing the increasing knowledge of an ongoing process, e.g. in the case of a series of coin-flips (with discrete time), $\mathcal{F}_0$ could be a trivial σ-algebra, as we know nothing as of $t = 0$, then in $t = 1$ we know the result of the first coin-flip, so $\mathcal{F}_1$ is bigger, as it separates possible events with respect to the result of the first coin-flip, and $\mathcal{F}_2$ is even more ,,atomised", and so on and so forth. We are now ready to get down to the nitty-gritty:

Definition: Given a measurable state space $(E, \mathcal{E})$ and an index set T, a **stochastic process** is a family of random variables

$$\{X_t\}_{t \in T}, \quad X_t : (\Omega, \mathcal{F}) \to (E, \mathcal{E}).$$

For each $\omega \in \Omega$, the map

$$X_\bullet(\omega) \ : \ T \ \longrightarrow \ E, \quad t \mapsto X_t(\omega)$$

is called the **trajectory** or **sample path** of the process. We say $\{X_t\}$ is **adapted to a filtration** $\{\mathcal{F}_t\}$ if each X_t is $\mathcal{F}_t$-measurable. Two stochastic processes are said to be **equal in law** if they have the same distributions (not the same trajectories, as they can be defined on different probability spaces!).

Unless stated otherwise, from now on for every stochastic process we'll assume its **natural filtration**, i.e. the family of sub-σ-algebras $\{\mathcal{F}_t^X\}_{t \in T}$, where $\mathcal{F}_t^X$ is the smallest σ-algebra with respect to which all the random variables X_s, for $s \leq t$, are measurable.

Examples of stochastic processes include the aforementioned coin-flips, a random walk on the real line or on a grid, and it is also common to model population dynamics in biology or even stock prices as stochastic processes.

We could, in principle, already discuss the Wiener process, but to better understand it, we need two more concepts... The following definition looks rather intimidating at first, but it has very neat interpretations that we'll try to familiarize the reader with after a brief discussion of its consequences.

Definition: Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $\mathcal{G} \subseteq \mathcal{F}$ a sub-σ-algebra, and X be an integrable random variable. A random variable Y is called the **conditional expectation** of X given $\mathcal{G}$, denoted by

$$Y = \mathbb{E}[\, X \mid \mathcal{G}\,],$$

if

1. Y is $\mathcal{G}$-measurable,
2. For every $G \in \mathcal{G}$,
$$\int_G Y \, dP = \int_G X \, dP.$$

We won't prove this, however one can show that such a variable always exists and is defined uniquely up to values on a set with probability zero. Now, also without a proof, we'll list some useful properties of conditional expectation:

Proposition: Conditional expectation satisfies:

1. Linearity: For integrable X, Y and constants a, b:
$$\mathbb{E}[\, aX_1 + bX_2 \mid \mathcal{G}\,] = a\,\mathbb{E}[X_1 \mid \mathcal{G}] \; + \; b\,\mathbb{E}[X_2 \mid \mathcal{G}].$$
2. If X is $\mathcal{F}$-measurable, then $\mathbb{E}[X \mid \mathcal{F}] = X$ almost surely.
3. $\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X \mid \mathcal{G}]]$.
4. $|\mathbb{E}[X \mid \mathcal{G}]| \leq \mathbb{E}[|X| \mid \mathcal{F}]$ almost surely.
5. The ,,Tower Property": If $\mathcal{H} \subseteq \mathcal{G} \subseteq \mathcal{F}$, then
$$\mathbb{E}\big[\mathbb{E}[X \mid \mathcal{G}] \,\big|\, \mathcal{H}\big] = \mathbb{E}[X \mid \mathcal{H}].$$
6. Monotonicity: If X, Y are integrable and $X \leq Y$ almost surely, then
$$\mathbb{E}[X \mid \mathcal{G}] \leq \mathbb{E}[Y \mid \mathcal{G}] \quad \text{almost surely.}$$
7. If Z is $\mathcal{G}$-measurable and X is integrable, then
$$\mathbb{E}[\, ZX \mid \mathcal{G}\,] = Z\,\mathbb{E}[X \mid \mathcal{G}].$$

With all that in mind, we can discuss how to understand conditional expectations: one could say that the expected value of a random variable is its simplest approximation; one can also readily see that the conditional expected value is a more detailed approximation of a random variable (in some cases it would be simply a function that is constructed by averaging out the values of the original function on consecutive intervals of the real line...) – in finance it's used to model ,,realistic" approximations of some ,,ideal" random variable, which are necessary due to limited knowledge of the probability space. If we restrict ourselves to L^2 spaces (which are Hilbert spaces), then $\mathbb{E}[\, X \mid \mathcal{G}\,]$ is an orthogonal projection of X onto the subspace of $\mathcal{G}$-measurable functions! There's more to this story, but it would require introducing the notion of conditional probability, which we won't need further down the line.

Instead, we'll use conditional expectation to define a class of stochastic processes with an additional useful and fairly intuitive property:

Definition: Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in T}, \mathbb{P})$ be a filtered probability space. A stochastic process $\{M_t\}_{t\in T}$ with each $M_t \in L^1(\Omega, \mathcal{F}, P)$ is a *martingale* w.r.t. $(\mathcal{F}_t)$ if:

1. M_t is $\mathcal{F}_t$-measurable for all t,
2. For all $s \leq t$,
$$\mathbb{E}[\, M_t \mid \mathcal{F}_s\,] = M_s \quad \text{almost surely.}$$

Using one of the properties (namely 3.) of conditional expectation listed above, one immediately finds that for a martingale the expected values remain constant, i.e. $\mathbb{E}[M_t] = \mathbb{E}[M_0] \quad \forall\, t$. That's why they are said to be modeling ,,fair games" (one can show that no betting strategy, which is formalized in the notion of stopping time, can beat the game, i.e. guarantee a net gain), but they also prove to be extremely powerful theoretical tools. This concludes our review of basic probability theory.

4.2 Wiener process

We are going to define (and discuss) perhaps the most important stochastic process of them all, used throughout all the natural and social sciences to model a wide array of phenomena.

Definition: We call stochastic process $\{W_t\}_{t\in[0,\infty)}$ a (standard) **Wiener process** if it satisfies the following properties:

1. $W_0 = 0$ almost surely.
2. The increments of W are independent, i.e. $W_{t+u} - W_t$ is independent of past values W_s for all $t, u > 0$ and $s < t$.
3. The increments are Gaussian, i.e. $W_{t+u} - W_t \sim \mathcal{N}(0, u)$ (meaning $W_{t+u} - W_t$ has normal distribution with mean 0 and variance u), again for all $t, u > 0$.
4. $W_t(\omega)$ is continuous as a function of t for a fixed $\omega \in \Omega$.

It is not at all obvious whether such process exists. There are many ways to construct it, but the details of the proofs are rather technical and lengthy. For our intents, it suffices to refer to the simplest construction, as a certain limit of infinitesimal random walks, which also explains why the Wiener process describes Brownian motion (or can be a simple model for stock prices, or why it's so prevalent in general, and also why the normal distribution pops up). Consider a series of i.i.d. Rademacher's variables, i.e. $X_i, i \in \mathbb{N}$, with $\mathbb{P}(X_i = \pm 1) = \frac{1}{2}$. For $t \in [0, 1]$ let

$$W_n(t) := \frac{1}{\sqrt{n}} \sum_{1\leq i\leq \lfloor nt \rfloor} X_i,$$

so that for any ω in the sample space (representing a single trajectory), $W_n(t)$ is equal to a sum of $\lfloor nt \rfloor$ random -1s and 1s (,,a new one" every $\frac{1}{n}$ seconds, one could say), scaled by $\frac{1}{\sqrt{n}}$. By Donsker's theorem (which is a generalisation of the central limit theorem), as $n \to \infty$, $W_n \to W$ – the Wiener process for $t \in [0, 1]$. Having such process, with some trial and error, it is easy to extend it to a process for $t \in [0, \infty)$ (essentially, it's possible thanks to the fractal nature of the trajectories, which we'll soon state rigorously): just take $W'_t = (1 + t)(W_{t/1+t} - \frac{1}{1+t} W_t)$. It is also obvious how to construct multidimensional Wiener processes. Now an intrigued reader can try to prove the following:

Theorem: A Wiener process exists.

□

Aside from the sole existence, to understand further discussions, we'll need only the following two properties of the Wiener process: **a trajectory is almost surely nowhere differentiable and almost surely has unbounded variation on every arbitrarily small interval** (this pathological behavior is in fact expected from our construction). Nonetheless, we'll list some other interesting properties, and again, we won't prove any of them (though some are almost immediate, while other require rather technical auxiliary lemmas):

Proposition: A Wiener process satisfies the following properties:

1. $\mathbb{E}[W_t] = 0$ for all $t \geq 0$.
2. $\mathrm{Var}[W_t] = t$
3. It's a martingale (combining this fact with the Optional Stopping Theorem, which we don't have the time to cover, would yield us many fascinating and powerful results).
4. $\mathbb{P}(W_t > r) = \frac{1}{2}\mathbb{P}(\sup_{s\leq t} W_s > r)$ for $r \geq 0$, meaning that the probability of arriving at some point r at time t is equal to $\frac{1}{2}$ the probability that we've arrived at it previously, because from that moment we could've equally well went either ,,up" or ,,down".
5. During an arbitrarily small time interval $[0, \epsilon]$, the trajectory was almost surely both below and above 0.
6. Here are a few processes that are equal to it in law: a time scaled process $X_t = \frac{\pm 1}{\sqrt{c}} W_t(ct)$, an inverted process $Y_t = tW_{\frac{1}{t}}$ and a shifted process $Z_t = W_{T+t} - W_T$ (with $c, T > 0$).

Moreover, it is crucial for our understanding of the relationship between previously discussed matter (Section 2.2) and the subject at hand, that **the Wiener measure is the pushforward of the measure induced by the Wiener process onto the space of paths** defined before. This is fairly intuitive (and strongly indicated by how those objects are called), but to further cement our intuitions and conclude this section with some nice pictures, here are few sample paths of a Wiener process, taken from Wikipedia:

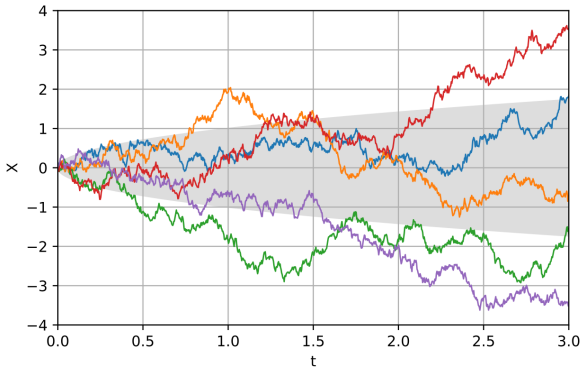

Figure 2: Sample trajectories of a standard Wiener process X_t

5 Introductory Itô calculus

Being fluent in probability theory language, we are now prepared to talk about (non-trivial) SDEs. Warning: we'll deal with only one among many inequivalent approaches to SDEs; on top of that, we won't discuss the theory in it's most general form, but it'll be enough to finally arrive at the Feynman-Kac formula. This section is based on [3], where the reader can find further development of the showcased ideas and tons of references to literature, that covers them more thoroughly.

5.1 Basic SDEs and Itô's integral

Observe that we can get a new interesting stochastic process (called Brownian motion with drift) from the Wiener process by letting

$$X_t := \mu t + \sigma W_t$$

(modifying Brownian motion is very common to e.g. model noise). Using the known properties of W_t we readily see that $\mathbb{E}[X] = \mu t$ and $\mathrm{Var}[X_t] = \sigma^2 t$. Based on previous discussions and Figure 2, it's easy to imagine what are the properties of trajectories. We could also say that for every path of X_t the following equation is satisfied:

$$\mathrm{d}X_t = \mu \mathrm{d}t + \sigma \mathrm{d}W_t,$$

if we are thinking of infinitesimal incremets; SDEs are a way to formalize exactly this notion, of which we now have a gist.

We want to consider equations of the form

$$\mathrm{d}X_t = \mu(X_t, t)\mathrm{d}t + \sigma(X_t, t)\mathrm{d}W_t, \tag{14}$$

where W_t is an m-dimensional Wiener process, $\mu : \mathbb{R}^d \times \mathbb{R}_{\geq 0} \to \mathbb{R}^d$ and $\sigma : \mathbb{R}^d \times \mathbb{R}_{\geq 0} \to \mathbb{R}^{d\times m}$ are the usual measurable, ,,deterministic" functions – we see that in the previous example they were just constant (so we already know the solution to the simplest possible equation!) – and we want a solution X_t, which can then be called a **diffusion process**, because it satisfies the so called **diffusion equation** of the form 14; we call μ the drift term, and σ the diffusion term. We also impose an initial condition $X_0 = x$, x being a random variable. Because the solution depends on the initial condition, we'll usually stress this fact by denoting it X_t^x. As it turns out, a proper way to interpret such equation is in integral form:

$$X_t^x = x + \int_0^t \mu(X_t^x, t)\,\mathrm{d}t + \int_0^t \sigma(X_t^x, t)\,\mathrm{d}W_t,$$

with the integral being... well, it depends – here comes the ambiguity. In fact, there are even uncountably many inequivalent ways to define stochastic integrals, but the central idea is the same, and it's based on Riemann–Stieltjes-like integral.

The first integral is straight forward, but what about the second? Because of unbounded variation of paths, we can't just define the integral $\int_0^T H_t \mathrm{d}W_t$ to be Riemann–Stieltjes integral with respect to Brownian motion on every path, because we'd run into ambiguities. Instead, we'll assume that our integrand H is adopted to the Wiener process' natural filtration, and then follow two simple steps:

First, we assume H_t to be a simple process, i.e. there exists a partition

$$0 = t_0 < t_1 < \cdots < t_n = T$$

and $\mathcal{F}_{t_k}$–measurable random variables H_{t_k}, such that

$$H_t(\omega) = \sum_{k=0}^{n-1} H_{t_k}(\omega)\, \mathbf{1}_{[t_k, t_{k+1})}(t).$$

For such H, we simply define

$$\int_0^T H_t \,\mathrm{d}W_t \;:=\; \sum_{k=0}^{n-1} H_{t_k} \left(W_{t_{k+1}} - W_{t_k}\right).$$

Next, let

$$\mathcal{H}^2([0,T]) := \Big\{ H_t \text{ adapted to the filtration } \Big|\; \mathbb{E}\big[\int_0^T H_t^2 \,\mathrm{d}t\big] < \infty \Big\}.$$

It's easy to check that or any simple H,

$$\mathbb{E}\Big[\Big(\int_0^T H_t \,\mathrm{d}W_t\Big)^2\Big] = \mathbb{E}\Big[\int_0^T H_t^2 \,\mathrm{d}t\Big],$$

and since simple processes are dense (a result similar to the standard measure-theoretic propositions about density) in $\mathcal{H}^2$ under the norm $\|H\|^2 = \mathbb{E}[\int_0^T H_t^2 \,\mathrm{d}t]$, the mapping $H \mapsto \int_0^T H_t \,\mathrm{d}W_t$ extends uniquely by continuity to all of $\mathcal{H}^2([0,T])$. Analogous arguments also show that the resulting random variable is a martingale. Our obtained integral is called **Itô's integral**. Had we chosen the integral for simple processes to be

$$\sum_{k=0}^{n-1} \tfrac{1}{2}\big(H_{t_k} + H_{t_{k+1}}\big) \left(W_{t_{k+1}} - W_{t_k}\right),$$

we would've ended up with **Stratonovich integral**, which is the second most commonly used type, and has differing ,,nice" properties (including a formula for transforming it into Itô's integral and vice versa), but we won't discuss is it this text; similarly, other choices of sums produce other integrals, but we'll just stick to the Itô's definition.

Having constructed a stochastic integral, we can now define what we mean by

Definition: a **strong solution** to a SDE as in eq. 14 along with the initial condition, is a process X_t with continous paths and adopted to the Wiener's filtration, such that:

1. For all $T \geq 0$, $\mu(\bullet, X_\bullet) \in L^1([0,T])$ and $\sigma(\bullet, X_\bullet) \in L^2([0,T])$ almost surely (so that integrals in the next condition are well defined).
2. For every $t \geq 0$, the equations: $X_0 = x$ and

$$X_t = x + \int_0^t \mu(X_t, t)\,\mathrm{d}t + \int_0^t \sigma(X_t, t)\,\mathrm{d}W_t$$

hold almost surely.

Moreover, under two additional assumptions, namely that there exists a constant C such that:

$$|\mu(x,t)| + |\sigma(x,t)|_F \leq C(1 + |x|), \tag{15}$$

and

$$|\mu(x,t) - \mu(y,t)| + |\sigma(x,t) - \sigma(y,t)|_F \leq C|x - y| \tag{16}$$

for all $x, y \in \mathbb{R}^d$ and $t \in [0,T]$ (this is somewhat reminiscent of Lipschitz condition for ODEs), where $|\bullet|_F$ denotes the Frobenius matrix norm (and if the reader is unfamiliar with it, they are probably fortunate not to have had to suffer through a course on computational mathematics), there holds a uniqueness

Theorem 4.1: under the above two assumptions, and additionally the initial value x is independent of W_ts, with $\mathbb{E}[x^2] < \infty$, then the SDE has a unique global strong solution satisfying

$$\mathbb{E}\left[\int_0^t |X_s|^2 ds\right] < \infty$$

for all $t > 0$ (uniqueness means that if Y_t is another solution, then $X_t = Y_t$ almost surely).

We won't prove this; it's solely for completeness of our discussion of basics of SDEs. Instead, we'll dive into one interesting property of diffusion processes, that'll lead us to a tool with strong connections to operator theory...

5.2 Markov property and generators

Intuitively, if a diffusion process is governed by a differential equation, then its future state doesn't depend on the past, only on the present. Formalization of this notion is called Markov property. To define it we first need:

Definition: Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. For an event $A \in \mathcal{F}$ and sub-σ-algebra $\mathcal{G}$ of $\mathcal{F}$, we define the **conditional probability**

$$\mathbb{P}(A \,|\, \mathcal{G}) := \mathbb{E}[1_A \,|\, \mathcal{G}] \ .$$

If X is a random variable, then

$$\mathbb{P}(A \,|\, X) := \mathbb{E}[1_A \,|\, \sigma(X)] \ .$$

We want the reader to rest assured that this definition fits their intuitions nicely (and, along with some simple lemmas for which we don't have the time, can be useful for calculating probabilities). Now:

Definition: Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $\{X_t\}_{t\in T}$ be a stochastic process with values in a measurable state space $(E, \mathcal{E})$. We say $\{X_t\}$ is a **Markov process** if for all $s, t \in T$ with $s < t$ and all $A \in \mathcal{E}$

$$P\big(X_t \in A \ \big|\ \mathcal{F}_s\big) \ = \ P\big(X_t \in A \ \big|\ X_s\big) \ ,$$

where $\mathcal{F}_s = \sigma\big(X_u : u \le s\big)$ is the natural filtration. The **transition function** of the process is a family of the so called probability kernels (which are very similar to kernels defined in Section 1):

$$P_{s,t}(x, A) := \mathbb{P}(X_t \in A \mid X_s = x),$$

defined for all $0 \le s \le t$, $x \in E$, and $A \in \mathcal{E}$. The process (X_t) is called **time-homogeneous** if the transition function depends only on the time difference $t - s$, i.e.,

$$P_{s,t}(x, A) = P_{t-s}(x, A) \quad \text{for all } 0 \le s < t,\ x \in E,\ A \in \mathcal{E}.$$

Showing the following statement is true is a bit tricky and we'll omit it, as it is fairly intuitive:

Proposition: A diffusion process is Markovian. The diffusion process is a time-homegeneous Markov process iff its diffusion equation is time-independent.

□

For time-homogeneous Markov processes we can define a semi-group of operators P which completely governs its evolution and satisfies $P_0 = I$ (identity), $P_t P_s = P_{t+s}$ for all $t, s \ge 0$. Then we can formally write $P_t = e^{t\mathcal{L}}$ and $X_t = P_t X_0$, where $\mathcal{L}$ is the so called **infinitesimal generator** of the process. For now, we'll only flash this definition of an operator semi-group, to signal yet another connection to our previous topics, and immediately go back to precise definitions which will prove useful in studying SDEs (in fact, the link between semi-groups and stochastic analysis runs much deeper than we could possibly have covered, see [4]). For $f \in C_b(\mathbb{R}^d)$, define

$$(P_t f)(x) := \mathbb{E}[f(X_t) \,|\, X_0 = x].$$

Then the above hold for

$$\mathcal{L}f := \lim_{t\to 0} \frac{P_t f - f}{t},$$

whenever this limit exists. As it turns out, the infinitesimal generator of the Wiener process is the Laplacian (and one more fan fact: its backward Kolmogorov equation, a sort of fundamental equation for Markov processes, is precisely the heat equation!).

Now here comes a little trick: consider a time-inhomogeneous SDE of the form (14). To transform it into a time-homogeneous SDE, we define an augmented process $(Y_t)_{t\geq 0} \subset \mathbb{R}^{d+1}$ by

$$Y_t := (t, X_t).$$

Let us denote $Y_t = (Y_t^0, Y_t^1, \ldots, Y_t^d)$, where $Y_t^0 = t$, and $Y_t^i = X_t^i$ for $i = 1, \ldots, d$. Then Y_t satisfies the time-homogeneous SDE:

$$dY_t^0 = dt, dY_t^i = \mu_i(Y_t^0, Y_t^1, \ldots, Y_t^d)\, dt + \sum_{j=1}^{m} \sigma_{ij}(Y_t^0, Y_t^1, \ldots, Y_t^d)\, dW_t^j, \quad i = 1, \ldots, d.$$

This let's us define the generator: let $f : \mathbb{R}^{d+1} \to \mathbb{R}$ be a function for which the necessary limit exists. Then one can show the infinitesimal generator $\mathcal{L}$ of the process Y_t is given by:

$$\mathcal{L}f(t,x) = \frac{\partial f}{\partial t}(t,x) + \sum_{i=1}^{d} \mu_i(t,x)\, \frac{\partial f}{\partial x_i}(t,x) + \frac{1}{2} \sum_{i,j=1}^{d} (\sigma(t,x)\sigma(t,x)^\top)_{ij}(t,x)\, \frac{\partial^2 f}{\partial x_i \partial x_j}(t,x),$$

and it's a result we'll use later on.

5.3 Itô's lemma

We'll briefly cover one last piece of the puzzle needed for the proof of Feynman-Kac formula, which is at the same time sometimes considered the most crucial result in SDEs theory. Ideally, we'd like to be able to write down every solution of a diffusion process as a function of time (that's obvious) and Brownian motion (because we know a lot about it). Here's a step in that direction: recall the second property of the Brownian motion that we've listed, namely the ,,quadratic variation", i.e. $\mathbb{E}[W_t^2] = t$ – now knowing that expectation of W_t at any time is zero and that increments are independent, we could say that (for now only heuristically) $(\mathrm{d}W_t)^2 = \mathrm{d}t$. That's really the cornerstone of Itô's calculus, and we can supplement this relation between $\mathrm{d}t$ and $\mathrm{d}W_t$ with $\mathrm{d}t\, \mathrm{d}W_t = 0 = \mathrm{d}W_t\, \mathrm{d}t$ and classic $(\mathrm{d}t)^2 = 0$. In the usual calculus, one can show that differentiable function with bounded derivative has quadratic variation equal to 0, and that's what makes Ito's calculus different – objects in question are fundamentally distinct. We'll see this in the appropriate (one could say extended) version of the Leibniz rule.

Suppose we have a (sufficiently nice) ,,deterministic" function $f : \mathbb{R} \to \mathbb{R}$ and want to consider $f(W_t)$. What's $\mathrm{d}f$? We will see that because $f(W_t)$ is itself a diffusion process, it can have non zero quadratic variation, and so it's not simply $f'(W_t)\mathrm{d}W_t$. Let's ,,Taylor expand" f: $f(W_{t+u}) - f(W_t) = f'(W_t)(W_{t+u} - W_t) + \frac{1}{2} f''(W_t)(W_{t+u}^2 - W_t^2) + \ldots$ where what follows are terms of higher order, which we can ignore if be believe in the relationships listed above. From this, we infer that

$$\mathrm{d}f(W_t) = f'(W_t)\mathrm{d}W_t + \frac{1}{2} f''(W_t)\mathrm{d}t\ . \tag{17}$$

The formula above is the simplest version of the so called **Itô's lemma**, which has a lengthy and technical proof, which we won't copy down. Instead, now that we have a grasp of where it comes from, here's a full version of Itô lemma:

Theorem 4.2: under the assumptions of Theorem 4.1, if X_t is a solution of diffusion equation of the form and $f \in C^{1,2}([0,T] \times \mathbb{R}^d)$, then the process $f(X_t)$ satisfies

$$f(t, X_t) = f(0, X_0) + \int_0^t \frac{\partial f}{\partial s}(s, X_s)\mathrm{d}s + \int_0^t \mathcal{L}f(s, X_s)\mathrm{d}s + \int_0^t \langle \nabla f(s, X_s), \sigma(X_s)\mathrm{d}W_s \rangle \tag{18}$$

□

As a corollary, we get the (often more useful) differential version:

$$\mathrm{d}f\big(t, X_t\big) = \frac{\partial f}{\partial t}\big(t, X_t\big)\, \mathrm{d}t + \sum_{i=1}^{d} \frac{\partial f}{\partial x_i}\big(t, X_t\big)\, \mu_i\big(t, X_t\big)\, \mathrm{d}t + \sum_{i=1}^{d}\sum_{j=1}^{m} \frac{\partial f}{\partial x_i}\big(t, X_t\big)\, \sigma_{ij}\big(t, X_t\big)\, \mathrm{d}W_t^j + \frac{1}{2} \sum_{i=1}^{d}\sum_{j=1}^{d} \Big(\sigma\sigma^\top\Big)_{ij}(t, X_t)\, \frac{\partial^2 f}{\partial x_i\, \partial x_j}\big(t, X_t\big)\, \mathrm{d}t. \tag{19}$$

Important thing to note is that $f(t, X_t)$ is itself a diffusion process!

We can toy with our new lemma to get our first stochastic integral: simply take $f(t,x) = x^2$. It's then easy to check that what follows from the above is that

$$\int_0^t W_s \mathrm{d}W_s = \frac{1}{2}W_t^2 - \frac{1}{2}t \ .$$

Cool!

6 The titular formula again

6.1 Probabilistic proof

Equipped with the Itô calculus machinery we're finally ready to prove our titular formula with probabilistic approach.

Theorem 1 [Feynman-Kac formula no. 2]: Let X_t^x be a diffusion process with some drift $\mu(\cdot)$, diffusion $\Sigma(\cdot) = \sigma\sigma^T(\cdot)$ and generator $\mathcal{L}$, such that $X_0^x = x$, and let $f \in \mathcal{C}_0^2(\mathbb{R}^d)$ and $V \in \mathcal{C}(\mathbb{R})$ bounded from below. Then the solution to initial value problem:

$$\frac{\partial u}{\partial t} = (\mathcal{L} - V)u,$$

$$u(0,x) = f(x)$$

, is given by:

$$u(x,t) = \mathbb{E}(e^{-\int_0^t V(X_s^x)ds} f(X_t^x)).$$

Proof: We know that SDE equation for X_t^x is:

$$dX_t^x = \mu(X_t^x)dt + \sigma(X_t^x)dW_t$$

,

$$X_0^x = x.$$

Let introduce a new variable $Y_t = e^{-\int_0^t V(X_s^x)ds}$, with SDE:

$$dY_t^x = -V(X_t^x)Y_t^x dt,$$

$$Y_0^x = 1$$

Now we can consider process $Z_t := Y_t f(X_t) = \psi(X_t, Y_t)$, where $\psi(x,y) = yf(x)$. Our goal now is to find the differential equation for $\mathbb{E}[Z_t]$, we start by using Itô formula:

$$dZ_t = d\psi(X_t, Y_t) = \sum_{i=1}^d \frac{\partial \psi}{\partial x_i} dX_t^i + \frac{\partial \psi}{\partial y} dY_t + \frac{1}{2}\sum_{i,j=1}^d \frac{\partial^2 \psi}{\partial x_i \partial x_j} dX_t^i dX_t^j$$

We put in derivatives of ψ and use more compact notation:

$$dZ_t = Y_t \nabla f(X_t)^T dX_t + f(X_t)dY_t + \frac{1}{2}Y_t \sum_{i,j=1}^d \frac{\partial^2 f}{\partial x_i \partial x_j} dX_t^i dX_t^j$$

Now we can use the fact that:

$$dX_t^i dX_t^j = [\sigma\sigma^T]_{ij}(X_t)dt =: a_{ij}(X_t)dt$$

and get:

$$dZ_t = Y_t \nabla f(X_t)^T dX_t + f(X_t)dY_t + \frac{1}{2}Y_t tr[a(X_t)D^2 f(X_t)]dt$$

and by putting in SDE for Y_t and X_t:

$$dZ_t = Y_t \nabla f(X_t)^T b(X_t) + Y_t \nabla f(X_t)^T \sigma(X_t)dW_t - V(X_t)Y_t f(X_t)dt + \frac{1}{2}Y_t tr[a(X_t)D^2 f(X_t)]dt$$

by gathering terms with dt and dW_t we can finaly write:

$$dZ_t = Y_t[\mathcal{L}f(X_t) - V(X_t)f(X_t)]dt + Y_t \nabla f(X_t)^T \sigma(X_t)dW_t$$

The stochastic integral

$$\int_0^t Y_s \nabla f(X_s)^T \sigma(X_s)\, dW_s$$

is a martingale, because - as noted in Section 4.1 — the Itô integral with respect to Brownian motion is a martingale for square-integrable processes. This condition is satisfied due to the regularity of the function f, the boundedness of Y_s, and standard assumptions on σ Now we can take expected value of equation, we get:

$$\frac{d}{dt}\mathbb{E}[Z_t] = \mathbb{E}[(\mathcal{L}f(X_t) - V(X_t)f(X_t))Y_t]$$

Differentiation under the expected value was permitted, because function are smooth and bounded and time integral is regular. Our equation is exactly initial PDE if we define $u(x,t) := \mathbb{E}_x[Y_t f(X_t)]$ where small x in the lower index is to take into account boundary condition.

6.2 Solving the Schrödinger equation with the Feynman-Kac formula

One could ask, why did we even bother with this whole huge stochastic machinery? For the sheer beauty of mathematics? That's an acceptable answer, but there is so much more to it than simply that. Feynman-Kac formula is powerful tool for approximating solutions to partial differential equations numerically, and among applicable examples is the Schrödinger equation. We will work out a simple 1D example with harmonic oscillator potential, so that we know the analytic solution, and thus can compare it with our approximation.
Hamiltonian for our model is given by:

$$H = -\frac{\hbar^2}{2m}\frac{d^2}{dx^2} + \frac{1}{2}m\omega^2 x^2$$

The Schrödinger equation is given by:

$$i\hbar\frac{\partial}{\partial t}\Psi = H\Psi$$

But to use our brand new tools we would like to perform a transformation called Wick rotation, it is simply just substitution $t = -i\tau$ and look on the function $u(x,\tau) = \Psi(x,-it)$, so our equation becomes:

$$\hbar\frac{\partial}{\partial \tau}u = Hu$$

The analytical solution for the ground state is of the form:

$$u(x,\tau) = exp(-\frac{E_0\tau}{\hbar})\exp(-\frac{1}{2}\alpha x^2)$$

Now the sweet part is coming. From the previous section, we know that:

$$u(x,\tau) = \mathbb{E}_x[\exp(-\frac{1}{\hbar}\int_0^t V(B_s)ds)\cdot u(B_t,0)]$$

To approximate this formula, we need to generate random Brownian trajectories and for each one calculate the following integral:

$$w = \exp(-\frac{1}{\hbar}\int_0^t \frac{1}{2}m\omega^2 B_s^2 ds)$$

Here are some generated trajectories with weight distribution for the whole simulation:

Figure 3: Generated Brown trajectories and weights distribution

Trajectories that went further are in stronger potential, therefore, have smaller weights. With our weights, we can calculate approximated solution:

$$u(x,\tau)_{app} = \frac{1}{N}\sum_{i=1}^{N} w_i \cdot u(B_t^{(i)}, 0)$$

For 2000 trajectories and time evolution from $t = 0$ to $t = 1$ our approximated function is well suited to analytical solution.

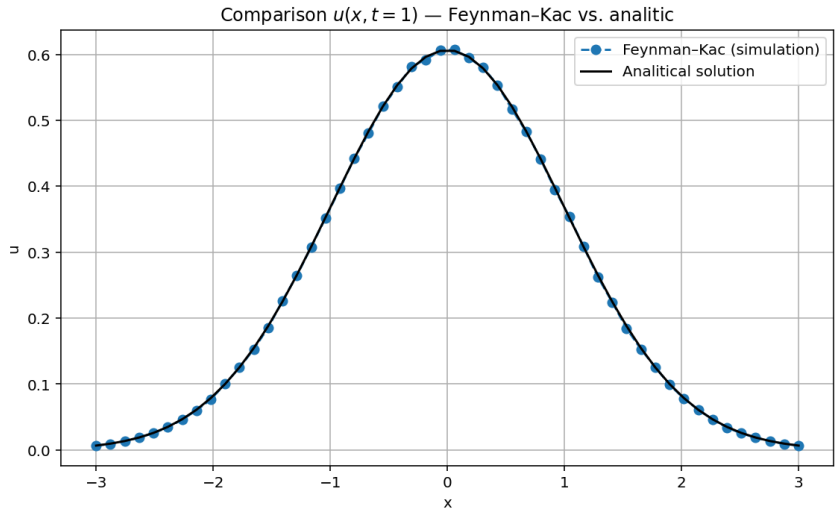

Figure 4: Comparison for analitical and approximated solution

References

[1] Reed, Michael, and Barry Simon, *Methods of Modern Mathematical Physics: II: Fourier Analysis, Self-Adjointness*. Academic Press, 1975.

[2] J. Jakubowski and R. Sztencel. *Wstep do teorii prawdopodobieństwa*. ed. IV, Script, 2010.

[3] G. A. Pavliotis,.*Stochastic Processes and Applications*. Springer, 2014.

[4] L. Lorenzi and M. Bertoldi. *Analytical Methods for Markov Semigroups*. CRC Press, New York, 2006.