SOLVING PARABOLIC PDES VIA WEAK APPROXIMATIONS OF SDES

BRIAN J. LEE

ABSTRACT. We provide an introduction to Monte Carlo methods for approximating solutions to a wide class of parabolic Cauchy problems. The central tool is the Feynman–Kac formula, which represents solutions of these PDEs as a functional integral over Wiener paths. To approximate these quantities numerically, we introduce weak approximation schemes for stochastic differential equations and show how they give rise to Monte Carlo estimators based on a sample mean. We additionally discuss techniques for estimators with reduced variance using an appropriate change of measure.

1. INTRODUCTION

As a motivating example, we begin by considering the d-dimensional *heat equation with decay*:

$$\partial_t u(t,x) = \kappa \Delta u(t,x) - \gamma(x) u(t,x), \qquad u(0,x) = f(x). \tag{1.1}$$

where $\kappa > 0$, $u \in C^{1,2}([0,T] \times \mathbb{R}^d; \mathbb{R})$, and $\gamma \in C(\mathbb{R}^d)$ is lower bounded. Here, κ is often referred to as the *diffusivity* and γ is often referred to as the *decay rate* (note γ here is spatially-dependent). In the case of $\gamma \equiv 0$, we recover the standard heat equation

$$\partial_t u = \kappa \Delta u \tag{1.2}$$

which admits the explicit heat-kernel representation [LeG16]:

$$u(t,x) = \int_{\mathbb{R}^d} f(y) \frac{1}{(4\pi\kappa t)^{d/2}} \exp\left(-\frac{|x-y|^2}{4\kappa t} \right) dy. \tag{1.3}$$

In the general case, one can instead write solutions to (1.1) as a Wiener integral

$$u(t,x) = \mathbb{E}\left[\exp\left(-\int_0^t \gamma(x + \sqrt{2\kappa} W_s)\, ds \right) f(x + \sqrt{2\kappa} W_t) \right], \tag{1.4}$$

due to the Feynman–Kac formula. In either case, however, the solutions (1.3) and (1.4) typically can not be written down in closed-form, and one has to resort to numerical methods to approximate solutions to (1.1) instead.

One straightforward deterministic approach is to discretize (1.1) directly by truncating $\mathbb{R}^d$ to a large bounded domain, say $D_R = [-R, R]^d$, and proceeding with a finite difference scheme. Taking a mesh size $\Delta x = 2R/M$, we obtain a grid with M points in each coordinate direction for a total of $O(M^d)$ total grid points. If U_i^n denotes an approximation to $u(n\Delta t, x_i)$, the standard central-difference approximation of the Laplacian is

$$\Delta_{\Delta x} U_i^n = \sum_{\ell=1}^d \frac{U_{i+e_\ell}^n - 2U_i^n + U_{i-e_\ell}^n}{(\Delta x)^2},$$

where e_ℓ is the ℓ-th coordinate vector. The explicit finite-difference scheme for (1.1) is then given by

$$U_i^{n+1} = U_i^n + \Delta t \left(\kappa \sum_{\ell=1}^d \frac{U_{i+e_\ell}^n - 2U_i^n + U_{i-e_\ell}^n}{(\Delta x)^2} - \gamma(x_i) U_i^n \right). \tag{1.5}$$

Finite difference schemes remainsimple and effective in low dimensions and thus remain the preferred method for numerical approximation in these regimes. For higher dimensional PDEs, however, these

approaches quickly become computationally infeasible due to the grid size's exponential dependence on the dimension d.

In this paper, we introduce an alternative probabilistic approach based on Monte Carlo estimation of Feynman–Kac solutions of PDEs such as (1.4). In particular, we consider the class of parabolic Cauchy problems of the form

$$\partial_t u(t,x) = \mathcal{L}u(t,x) - q(x)u(t,x), \qquad u(0,x) = f(x), \tag{1.6}$$

where the elliptical operator

$$\mathcal{L}\phi(x) = \sum_{i=1}^{d} a_i(x)\partial_i\phi(x) + \frac{1}{2}\sum_{i,j=1}^{d} c_{ij}(x)\partial_{ij}\phi(x), \qquad c(x) = \sigma(x)\sigma(x)^\top. \tag{1.7}$$

is given by the infinitesimal generator of a (time-homogenous) d-dimensional Itô diffusion:

$$dX_t = a(X_t)dt + \sigma(X_t)dW_t \qquad X_0 = x. \tag{1.8}$$

where $a : \mathbb{R}^d \to \mathbb{R}^d$, $\sigma : \mathbb{R}^d \to \mathbb{R}^{d\times m}$, and W is a m-dimensional Wiener process. Throughout this paper, we assume that both a and σ are globally Lipschitz with at most linear growth, so that the diffusion SDE a unique nonexplosive strong solution. The core idea of the Monte Carlo method then, is to create numerical approximations for the process $(X_t)_{t\geqslant 0}$ to estimate the corresponding functional integral given by Feynman–Kac. Because we consider only *functions* of the Itô process in the Feynman–Kac expectations, it suffices to have good approximations of the probability distribution of X_t rather than a close approximation of the sample paths themselves. This leads naturally to the idea of *weak approximation* which we use to analyze our numerical solvers.

We organize the paper as follows. In Section 2, we prove the Feynman–Kac formula for parabolic PDEs of the form (1.6) and then suggest simple Monte Carlo estimators for (1.6) based on simple sample averages of the corresponding Wiener integral. In Section 3, we introduce the notion of *weak approximation* of SDEs and use them to analyze the systematic error of various estimators. In Section 4, we consider the variance of our Monte Carlo estimators and develop a variance reduction method based on an appropriate change of measure.

2. The Path Integral Formulation for Parabolic PDEs

The Kolmogorov forward and backward equations provide a key connection between stochastic differential equations and certain types of parabolic PDEs. The Feynman–Kac formula, a direct corollary of the backward equation, additionally enables one to write down solutions to such a PDE as a functional integral over Wiener paths.

2.1. The Feynman–Kac Formula.

We begin with a concise proof of Feynman–Kac in the case of time-homogenous diffusions (which notably admit a semigroup) and show how it can be used to model solutions of (1.1). We streamline the presentation of [Oks03], making extensive use of the Theory of Markov Processes, for which we refer the reader to Chapter 6 of [LeG16] or Chapter 8 of my personal notes.

Theorem 1 (Feynman–Kac). *Let $X = (X_t)_{t\geqslant 0}$ be an $\mathbb{R}^d$-valued Itô diffusion with generator $\mathcal{L}$ given by (1.7) and SDE given by (1.8). If $f \in C_0^2(\mathbb{R}^d)$ and $q \in C(\mathbb{R}^d)$ is lower bounded, then*

$$u(t,x) = \mathbb{E}_x\left[e^{-\int_0^t q(X_s)\,ds} f(X_t)\right] \tag{2.1}$$

is the unique (classical) solution of the Cauchy problem

$$\partial_t u(t,x) = \mathcal{L}u(t,x) - q(x)u(t,x), \qquad u(0,\cdot) = f, \tag{2.2}$$

among functions $u \in C^{1,2}((0,\infty)\times\mathbb{R}^d)$ that are bounded on $[0,T]\times\mathbb{R}^d$ for every $T < \infty$.

Proof. Let $(Q_t)_{t\geqslant 0}$ be the transition semigroup of X, and define

$$A_t := \int_0^t q(X_s)\,ds, \qquad Q_t^q g(x) := \mathbb{E}_x[e^{-A_t} g(X_t)].$$

By the Markov property and the additivity of $t \mapsto A_t$, we have that $(Q_t^q)_{t\geqslant 0}$ satisfies the Chapman-Kolmogorov equations and that it is a semigroup. Moreover, for $g \in C_0^2(\mathbb{R}^d)$,

$$\frac{Q_h^q g(x) - g(x)}{h} = \frac{Q_h g(x) - g(x)}{h} + \mathbb{E}_x\left[\frac{e^{-A_h}-1}{h} g(X_h)\right].$$

Taking $h \downarrow 0$, the first term tends to $Lg(x)$ while $A_h/h \to q(x)$. Thus, by continuity of g,

$$\frac{e^{-A_h}-1}{h} g(X_h) \to -q(x)g(x).$$

so by dominated convergence, we get that Q_t^q has generator $\mathcal{L} - q$. Now, defining $u(t,\cdot) = Q_t^q f$, we have by the Kolmogorov Backward Equation that

$$\partial_t u(t,x) = (L-q)Q_t^q f(x) = Lu(t,x) - qu(t,x).$$

with $u(0,x) = Q_0^q f = f$, so (2.1) indeed solves the Cauchy problem (2.2). Note that $u(t,\cdot)$ is bounded on $[0,T]\times\mathbb{R}^d$ since $\sup_{0\leqslant t\leqslant T}|u(t,x)| \leqslant e^{cT}\|f\|_\infty < \infty$. For uniqueness, consider $v \in C^{1,2}((0,\infty)\times\mathbb{R}^d)$ also bounded on $[0,T]\times\mathbb{R}^d$ for $T<\infty$ that is a solution to (2.1). Applying Itô's formula and the product rule, we have for $M_s := e^{-A_s} v(T-s, X_s)$ that

$$dM_s = e^{-A_s}[-\partial_t v + Lv - qv](T-s, X_s)\,ds + dN_s,$$

where N_s is a local martingale. By assumption, we have that $\partial_t v = Lv - qv$ so the drift vanishes and by lower-boundedness of q, we get that M_s is bounded since

$$|M_s| \leqslant e^{cT}\|v\|_{\infty,[0,T]\times\mathbb{R}^d} < \infty$$

so M_s is in fact a true martingale. By the martingale property, we thus have

$$v(t,x) = \mathbb{E}_x[e^{-A_t} v(0,X_t)] = \mathbb{E}_x[e^{-A_t} f(X_t)] = u(t,x).$$

Thus, $v = u$. □

In particular, for $\kappa > 0$ and $\gamma \in C(\mathbb{R}^d)$ lower bounded, one may consider the Itô diffusion $X_t := x + \sqrt{2\kappa}W_t$ which has generator

$$\mathcal{L} = \frac{1}{2}(2\kappa\Delta) = \kappa\Delta,$$

from which one recovers (1.4):

Corollary 2 (Feynman–Kac for the Heat Equation)**.** *Under the assumptions of Theorem 1, the Cauchy problem* (1.1) *has a unique (classical) solution given by* (1.4)*.*

Remark 3. Theorem 1 is often used with a terminal value of u rather than an initial value. In particular, one can show that

$$v(t,x) = \mathbb{E}\left[\exp\left(-\int_t^T q(X_s)ds\right) g(X_T) \,\middle|\, X_t = x\right],$$

is the unique solution to (2.2) with boundary condition $v(T,x) = g(x)$ for a *terminal* time $T < \infty$ by taking the time reversal $v(t,x) = u(T-t,x)$.

2.2. **The Basic Monte Carlo Estimator.** We now introduce a simple Monte–Carlo estimator for the Cauchy problem (2.2) based on a sample estimate of (2.1). These estimators notably requires a significantly smaller number of computations compared to a finite-difference scheme such as (1.5). These come at the drawback of accuracy issues, including discretization error, convergence speed, and sample variance which we discuss in this subsection.

Let X be the diffusion from (1.8). Define the augmented process $\tilde{X} = (\bar{X}, A) \in \mathbb{R}^{d+1}$ by

$$d\bar{X}_t = a(\bar{X}_t)\,dt + \sigma(\bar{X}_t)\,dW_t, \qquad \bar{X}_0 = x, \tag{2.3}$$

$$A_t := -\int_0^t q(\bar{X}_s)\,ds. \tag{2.4}$$

Then the Feynman–Kac functional can be written as

$$u(t,x) = \mathbb{E}_x\left[e^{A_t} f(\bar{X}_t)\right] = \mathbb{E}_x[\psi(\tilde{X}_t)], \tag{2.5}$$

where we define

$$\psi(y) := f((y^1, \cdots, y^m))\exp(y^{m+1}). \tag{2.6}$$

In particular, we remark that (2.5) gives a representation for $u(t,\cdot)$ as a functional of a multi–dimensional Itô process. If one could either sample the paths of $\tilde{X}$ directly, or sample a process with exact, matching statistics for ψ of the form (2.6), one could obtain an unbiased estimator by taking a simple sample average:

$$\widehat{u}_M(t,x) := \frac{1}{M}\sum_{r=1}^{M} \psi(Y_t^{(r)}) \tag{2.7}$$

where $Y^{(r)}$ are independent samples of the process. In practice, such an exact sampling is impossible, and we instead rely on numerical methods to approximate the process instead. Typically, one discretizes the stochastic equations (**??**) with some step size $\delta > 0$ and approximates samples $\hat{Y}_t^{\delta,(r)}$, resulting in a biased estimator $\hat{u}_{M,\delta}$ due to the discretization error. In the next section, we introduce some numerical schemes and develop the requisite theory to analyze the resulting discretization errors.

We can additionally decompose the error of the resulting estimate as a sum of a systematic error due to the discretization, also referred to as the *bias*, and a statistical error due to the usage of the sample mean, also referred to as the *variance*:

$$\hat{e} = \hat{u}_{M,\delta} - u = \underbrace{(\hat{u}_{M,\delta} - \hat{u}_M)}_{e_{\text{sys}}} + \underbrace{(\hat{u}_M - u)}_{e_{\text{stat}}}. \tag{2.8}$$

For $\psi(\tilde{X}_t) \in L^2$, the statistical error converges asymptotically to a mean zero Gaussian:

$$e_{\text{stat}} = \hat{u}_R - u \sim \mathcal{N}\left(0, \frac{1}{R}\text{Var}(\psi(\tilde{X}_t))\right)$$

by the central limit theorem. In general however, $\text{Var}\psi(\tilde{(}X_t))$ may be extremely large, so one may require an extremely large number of samples to obtain sufficiently small confidence intervals. A natural question is then the construction of an estimator with reduced variance, which we return to in Section 4.

3. Weak Approximations Schemes for SDEs

In this section, we rigorously define the notion of a numerical, time-discrete approximation and then introduce the appropriate theory to analyze discretization error. In particular, we focus on the notion of weak approximation due to our focus on sample estimation. We then prove some explicit convergence results for the Euler scheme and then briefly describe how one can obtain higher order methods through a stochastic generalization of the Taylor expansion.

3.1. **Time-Discrete Approximations.** Let $T > 0$. Consider the d-dimensional Itô SDE on $[0, T]$:

$$dX_t = a(t, X_t)dt + b(t, X_t)dW_t, \qquad X_0 = x, \tag{3.1}$$

where W is a m-dimensional Wiener process.

Definition 4 (Time-discretizations of an SDE)**.** Consider a filtered probability space with filtration $(\mathcal{F}_t)_{t\geqslant 0}$ and let $\delta > 0$ be given. We define a *time discretization* with maximum step-size δ to be an increasing set of random variables

$$0 = \tau_0 < \tau_1 < \cdots < \tau_N = T$$

such that $\sup_k(\tau_{k+1} - \tau_k) \leqslant \delta$ and each τ_{k+1} is $\mathcal{F}_{\tau_k}$-measurable for each $0 \leqslant k < N$.

Intuitively, our definition implies that while our discretization times τ_k can potentially be random, it must be previsible and depend only on the information at the previous time τ_{k-1}.

Definition 5 (Time-Discrete Approximations of an SDE)**.** Let $\pi := (\tau_j : 0 \leqslant j < N)$ be a time-discretization as given by the previous definition. An $\mathbb{R}^d$-valued càdlàg process Y^δ is called a *time-discrete approximation* of X on π, if

(i) $Y^\delta_{\tau_k}$ is $\mathcal{F}_{\tau_k}$-measurable,

(ii) there exists $\ell \in \mathbb{Z}_{0+}$, $\mathcal{F}_{\tau_{k+1}}$-measurable random variables $Z_{k+1,j}$ for $1 \leqslant j \leqslant \ell$ and Borel $\Phi^\delta_{k+1} : (\mathbb{R}^d)^{k+1} \times \mathbb{R}^{k+2} \times \mathbb{R}^\ell \to \mathbb{R}^d$ such that

$$Y^\delta_{\tau_{k+1}} = \Phi^\delta_{k+1}\left(Y^\delta_{\tau_0}, \ldots, Y^\delta_{\tau_k}, \tau_0, \ldots, \tau_{k+1}, Z_{k+1,1}, \ldots, Z_{k+1,\ell}\right).$$

Definition 6 (Time-discrete scheme)**.** For each $\delta > 0$, let $\mathcal{P}_\delta$ denote the collection of all admissible time-discretizations of $[0, T]$ with maximum step-size at most δ. A *time-discrete scheme* for the SDE (3.1) is a map

$$(\delta, \pi) \longmapsto Y^{\delta,\pi}, \qquad \pi \in \mathcal{P}_\delta,$$

which assigns to each admissible discretization $\pi \in \mathcal{P}_\delta$ a time-discrete approximation $Y^{\delta,\pi}$ of X on π in the sense of the previous definition. If a particular family of discretizations $\pi^\delta \in \mathcal{P}_\delta$ is fixed, we write simply $Y^\delta := Y^{\delta,\pi^\delta}$.

Intuitively, the two conditions of Definition 5 imply that the definitions of $Y^\delta_{\tau_k}$ are recursive: the first condition implies that the computation of $Y^\delta_{\tau_k}$ should not involve more information than at the time τ_k, while the second condition implies that $Y_{\tau_{k+1}}$ depends only on the values at earlier discretization times, on the step size, and on a finite number of random variables which generate the noise mainly within the current time step. This definition, due to [KP99] notably also covers various kinds of interpolation methods, such as right continuous piecewise constant and linear interpolations. From now on, we assume a fixed interpolation scheme and define the scheme only on the grid. We also omit writing an explicit grid if it is obvious from the definition and use the words scheme and approximation interchangeably.

The simplest possible time-discrete scheme is given by a straightforward extension of the forward Euler Method for ODEs:

Algorithm 1 (Euler–Maruyama)**.** *Let X be a solution to (3.1) on $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geqslant 0}, \mathbb{P})$ and let W be the corresponding Wiener process also. Let $\pi^\delta = (\tau_k : 0 \leqslant k \leqslant N)$ be a time-discretization of $[0, T]$ on the same space. Denote $\Delta\tau_k := \tau_{k+1} - \tau_k$ and $\Delta W^\delta_{k+1} := W_{\tau_{k+1}} - W_{\tau_k}$. The Euler–Maruyama scheme associated with (3.1) on π^δ is defined recursively on the grid by $Y^\delta_{\tau_0} = x$, and for $0 \leqslant k < N$,*

$$Y^\delta_{\tau_{k+1}} = Y^\delta_{\tau_k} + a(\tau_k, Y^\delta_{\tau_k})\Delta\tau_k + b(\tau_k, Y^\delta_{\tau_k})\Delta W^\delta_{k+1}. \tag{3.2}$$

3.2. **Strong and Weak Convergence.** In order to compare and assess the convergence rates of different time-discrete schemes, we now introduce the notions of strong and weak order. For this section, let $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ be a filtered probability space and let X be a solution to (3.1).

Definition 7 (Strong convergence)**.** A time-discrete scheme Y^δ on $(\Omega, \mathcal{F}, (\mathcal{F}_t^\delta)_{t\geqslant 0}, \mathbb{P})$ (where $(\mathcal{F}_t^\delta)_{t\geqslant 0}$ is a potentially different filtration) with step-size $\delta > 0$ is said to converge to X with *strong order* $\gamma > 0$ on $[0, T]$ if there exists $\delta_0 > 0$, such that for all $\delta \in (0, \delta_0)$,

$$\left(\mathbb{E}\left[\sup_{0\leqslant t\leqslant T} |X_t - Y_t^\delta|^2\right]\right)^{1/2} = O(\delta^\gamma) \tag{3.3}$$

Under some mild regularity assumptions, one can show that the Euler-Maruyama scheme of Algorithm 1 has strong order $\gamma = 0.5$. See Theorem 10.2.2 of [KP99].

Definition 8 (Weak Convergence)**.** A time-discrete scheme Y^δ on a (potentially different) filtered probability space $(\Omega^\delta, \mathcal{F}^\delta, (\mathcal{F}_t^\delta), \mathbb{P}^\delta)$ with step-size $\delta > 0$ is said to converge to X with *weak order* $\beta > 0$ $[0, T]$, if, for every $\psi \in C_P^{2(\beta+1)}(\mathbb{R}^d; \mathbb{R})$, there exists $\delta_0 > 0$ such that, for all $\delta \in (0, \delta_0)$,

$$\sup_{0\leqslant t\leqslant T} \left|\mathbb{E}[\psi(X_t)] - \mathbb{E}^\delta[\psi(Y_t^\delta)]\right| = O(\delta^\beta).$$

where $C_P^{2(\beta+1)}$ refers to the set of $2(\beta + 1)$-differentiable functions with polynomial growth.

Intuitively, the strong order represents the convergence rate of the path, while the weak order represents the convergence rate of the distribution. This is evidenced by the fact that strong convergence requires that our time-discrete scheme be defined on the same probability space, unlike the weak convergence scheme which only requires proximity of *statistics* of the approximation and the true process. In fact, because we only require that the measure induced by Y^δ to be close, for the Euler scheme of Algorithm 1 that we can replace the Gaussian increments ΔW_{k+1}^δ with other random variables with similar moment properties. This leads to a simplified version of the Euler scheme:

Algorithm 2 (Simplified Weak Euler Scheme)**.** *Let $\delta > 0$ be given and let $\Delta\hat{W}_k^\delta : \mathbb{R}^m \to \mathbb{R}^d$ be independent $\mathcal{F}_{\tau_{k+1}}$-measurable random variables for $1 \leqslant k \leqslant N$ such that*

$$|\,\mathbb{E}(\Delta\hat{W}_k^\delta)| + |\,\mathbb{E}((\Delta\hat{W}_k^\delta)^3)| + |\,\mathbb{E}((\Delta\hat{W}_k^\delta)^2) - \delta| = O(\delta^2) \tag{3.4}$$

Then, the Simplified Weak Euler Scheme is given by taking (3.2) with ΔW_{k+1}^δ replaced by $\Delta\hat{W}_{k+1}^\delta$.

A straightforward calculation shows a two-point distribution satisfies the constraint of Algorithm 2. Often, this two-point distribution is preferred to Algorithm 1 for weak approximations because it is cheaper to sample and has bounded increments:

Lemma 9. *We have independent $\Delta\hat{W}_j^\delta$ s.t. $\mathbb{P}(\Delta\hat{W}_j^\delta = \pm\sqrt{\delta}) = 1/2$ satisfies the condition (3.4).*

We now present a result due to [MP91], which shows that if the drift a and covariance $B = bb^\top$ are slightly more than twice continuously differentiable, then the weak Euler scheme converges with weak order $\beta = 1.0$. We write $\mathcal{H}_T^{(r)}$ for the parabolic Hölder class of order r on $[0, T] \times \mathbb{R}^d$; in particular, when $r > 2$, the functions have two spatial derivatives and one time derivative with the corresponding Hölder remainder estimates.

Theorem 10 (Mikulevičius–Platen)**.** *Let $X = (X_t)_{0\leqslant t\leqslant T}$ be a d-dimensional Itô diffusion process given by (3.1) and define the covariance matrix:*

$$B(t, x) := b(t, x)b(t, x)^\top.$$

Suppose that all moments of X_T exist and that B is uniformly elliptic, i.e. there exists $\lambda > 0$ s.t.

$$\xi^\top B(t,x)\xi \geqslant \lambda|\xi|^2 \tag{3.5}$$

for all $t \in [0,T]$, $x, \xi \in \mathbb{R}^d$. Assume moreover, that for some $l \in (0,3)\backslash\{1,2\}$, we have $a, B \in \mathcal{H}_T^{(l)}$, and that $g \in \mathcal{H}^{(l+2)}$. Let Y^δ be a (weak) Euler approximation given by the Algorithm 2 with discretization $\pi^\delta = (\tau_k : 0 \leqslant k \leqslant N)$ of $[0,T]$ and step size $\delta > 0$. Then, we have

$$\left|\mathbb{E}[g(X_T)] - \mathbb{E}[g(Y_T^\delta)]\right| = O(\delta^{\chi(l)}) \tag{3.6}$$

where

$$\chi(l) = \begin{cases} 1/2, & l \in (0,1), \\ 1/(3-l), & l \in (1,2), \\ 1, & l \in (2,3). \end{cases} \tag{3.7}$$

Proof Sketch. The main idea of the proof is to first rewrite the Kolmogorov Backward Equation to an error caused by freezing the coefficients in (3.2) and then using Hölder regularity to bound the total error. Let

$$L^0 := \partial_t + \sum_{i=1}^d a^i(t,x)\partial_i + \frac{1}{2}\sum_{i,j=1}^d B^{ij}(t,x)\partial_{ij}$$

be the backward Kolmogorov operator associated to (3.1). From the uniform ellipticity of B and the parabolic Hölder regularity assumptions, the terminal value problem

$$L^0 v = 0, \qquad v(T,x) = g(x),$$

has a unique solution $v \in \mathcal{H}_T^{(l+2)}$ (see Thm 5.2, pg. 361 of [LSU68]). Moreover, for an appropriate weighted interior norm $\|\cdot\|_{\mathcal{H}_T^{(l+2)}}$, we have boundedness of the norm: $\|v\|_{\mathcal{H}_T^{(l+2)}} \leqslant C\|g\|_{\mathcal{H}^{(l+2)}}$. By Itô's formula applied to the exact diffusion X, we have

$$\mathbb{E}[v(0,X_0)] = \mathbb{E}[v(T,X_T)] = \mathbb{E}[g(X_T)].$$

Let $Y = Y^\delta$ denote the continuous Euler interpolation (one can extend to the case of the general increments by matching the required moments; we do not reproduce the separate moment-matching argument here). Thus, for $s \in [\tau_n, \tau_{n+1})$, writing $\bar{s} := \tau_n$, we have

$$dY_s = a(\bar{s}, Y_{\bar{s}})\,ds + b(\bar{s}, Y_{\bar{s}})\,dW_s.$$

Applying Itô's formula to $v(s, Y_s)$, and using $X_0 = Y_0$, gives

$$\begin{aligned}\mathbb{E}[g(Y_T)] - \mathbb{E}[g(X_T)] &= \mathbb{E}[v(T,Y_T)] - \mathbb{E}[v(0,Y_0)] \\ &= \mathbb{E}\int_0^T \left[\partial_t v(s,Y_s) + \sum_{i=1}^d a^i(\bar{s},Y_{\bar{s}})\partial_i v(s,Y_s) + \frac{1}{2}\sum_{i,j=1}^d B^{ij}(\bar{s},Y_{\bar{s}})\partial_{ij}v(s,Y_s)\right]ds.\end{aligned}$$

Since $L^0 v = 0$, this becomes

$$\mathbb{E}[g(Y_T)] - \mathbb{E}[g(X_T)] = \mathbb{E}\int_0^T \left[\sum_{i=1}^d \big(a^i(\bar{s},Y_{\bar{s}}) - a^i(s,Y_s)\big)\partial_i v(s,Y_s) \quad + \frac{1}{2}\sum_{i,j=1}^d \big(B^{ij}(\bar{s},Y_{\bar{s}}) - B^{ij}(s,Y_s)\big)\partial_{ij}v(s,Y_s)\right]ds.$$

turning the weak error into an error over the coefficients.

It remains to bound the two terms in the last display. The key estimate is Lemma 14.1.6 of [KP99], which says that for $f \in \mathcal{H}_T^{(l)}$,

$$\left|\mathbb{E}\left[f(s,Y_s) - f(\bar{s},Y_{\bar{s}}) \mid \mathcal{F}_{\bar{s}}\right]\right| \leqslant C\|f\|_{\mathcal{H}_T^{(l)}}\delta^{\chi(l)}.$$

This estimate is proved by mollifying f in space and time, applying Itô's formula to the mollified function, and then optimizing the smoothing parameters. The optimization is precisely what produces the exponent $\chi(l)$ in (3.7). To use the lemma, decompose, for example,

$$\big(a^i(\bar{s}, Y_{\bar{s}}) - a^i(s, Y_s)\big)\partial_i v(s, Y_s) = (a^i\partial_i v)(\bar{s}, Y_{\bar{s}}) - (a^i\partial_i v)(s, Y_s) - a^i(\bar{s}, Y_{\bar{s}})\big[\partial_i v(\bar{s}, Y_{\bar{s}}) - \partial_i v(s, Y_s)\big],$$

and similarly for the terms involving $B^{ij}\partial_{ij}v$. Since $a, B \in \mathcal{H}_T^{(l)}$ and $v \in \mathcal{H}_T^{(l+2)}$, the functions $a^i\partial_i v$, $\partial_i v$, B^{ij}, $\partial_{ij}v$, and $\partial_{ij}v$ all belong to $\mathcal{H}_T^{(l)}$. Hence Lemma 14.1.6 applies to each of these terms. Integrating over $s \in [0, T]$, and using the regularity bound on v, we obtain

$$\big|\mathbb{E}[g(X_T)] - \mathbb{E}[g(Y_T^\delta)]\big| \leqslant C\|g\|_{\mathcal{H}^{(l+2)}}\delta^{\chi(l)} = O(\delta^{\chi(l)}).$$

This proves the desired weak error estimate. □

3.3. **Higher Order Schemes.** We now (briefly) describe how one might obtain higher order weak approximations: the main idea is truncation of the Wagner–Platen expansion, a stochastic generalization of the integral form of Taylor's formula. The main idea is essentially a repeated application of Itô's Lemma: let $f \in C^n(\mathbb{R})$ and X_t an Itô Diffusion on $[t_0, T]$ satisfying (3.1). By Itô's Lemma, we have:

$$f(X_t) = f(X_{t_0}) + \int_{t_0}^t \left(a(X_s)\frac{\partial}{\partial x}f(X_s) + \frac{1}{2}b^2(X_s)\frac{\partial^2}{\partial x^2}f(X_s)\right)ds + \int_{t_0}^t b(X_s)\frac{\partial}{\partial x}f(X_s)dW_s \tag{3.8}$$

$$= f(X_{t_0}) + \int_{t_0}^t L^0 f(X_s)ds + \int_{t_0}^t L^1 f(X_s)dW_s \tag{3.9}$$

where we introduce the operators

$$L^0 := a\frac{\partial}{\partial x} + \frac{1}{2}b^2\frac{\partial^2}{\partial x^2}, \qquad L^1 = b\frac{\partial}{\partial x}$$

Now, for $f \equiv x$, we get obviously

$$X_t = X_{t_0} + \int_{t_0}^t a(X_s)ds + \int_{t_0}^t b(X_s)dW_s.$$

Then, repeated application of (3.9) to $f = a$ and $f = b$, we obtain

$$\begin{aligned} X_t &= X_{t_0} + \int_{t_0}^t \left(a(X_{t_0}) + \int_{t_0}^s L^0 a(X_z)\,dz + \int_{t_0}^s L^1 a(X_z)\,dW_z\right)ds + \int_{t_0}^t \left(b(X_{t_0}) + \int_{t_0}^s L^0 b(X_z)\,dz + \int_{t_0}^s L^1 b(X_z)\,dW_z\right)dW_s \\ &= X_{t_0} + a(X_{t_0})\int_{t_0}^t ds + b(X_{t_0})\int_{t_0}^t dW_s + R, \end{aligned} \tag{3.10}$$

where

$$R = \int_{t_0}^t\int_{t_0}^s L^0 a(X_z)\,dz\,ds + \int_{t_0}^t\int_{t_0}^s L^1 a(X_z)\,dW_z\,ds + \int_{t_0}^t\int_{t_0}^s L^0 b(X_z)\,dz\,dW_s + \int_{t_0}^t\int_{t_0}^s L^1 b(X_z)\,dW_z\,dW_s.$$

This gives us the simplest, lowest order stochastic Taylor expansion. One can then straightforwardly extend this for higher orders by applying (3.9) to $f = L^1 b$ (or $L^1 a$, $L^0 a$, $L^0 b$). We state the general result (which requires some more careful bookkeeping) in the Appendix.

From here, one can obtain higher order approximations by adding more terms in the Wagner–Platen expansion. For example, by keeping all terms involving double stochastic integrals, one obtains the *order* 2.0 *weak Taylor scheme* (see Algorithm 3 in the Appendix). Further schemes can then be developed by using the Taylor expansion to guess an approximate form and then check one of several sufficient moment conditions to determine if the scheme is of a desired order (see for example Theorem 14.5.2 of [KP99]).

4. Variance Reduction Methods

In the previous section, we analyzed the discretization error associated with the Monte Carlo estimator in (2.7). We now turn our attention to the second term in the bias–variance decomposition (2.8) and consider the problem of constructing an estimator with reduced variance.

The intuition is as follows: we may have that $X_t \in B_t$ for some measurable region B_t with high probability, but the value of the functional $\psi(X_t)$ may be small in that region. The key idea is to make an appropriate change of measure using Girsanov's theorem to get an estimator satisfying

$$\mathbb{E}[g(\tilde{X}_t)\Theta] = \mathbb{E}[g(X_T)]$$

for an appropriate Radon–Nikodym derivative Θ and modified Itô process $\tilde{X}$. We shall see that for an appropriate choice of $\tilde{X}$, we can control (and greatly reduce) the sample variance.

4.1. Girsanov Change of Measure.

Let $\theta = (\theta_t)_{0\leqslant t\leqslant T}$ be an $\mathbb{R}^m$-valued adapted process satisfying Novikov's criterion (see [LeG16] Thm. 5.23):

$$\mathbb{E}\left[\exp\left(\frac{1}{2}\langle L\rangle_\infty\right)\right] = \mathbb{E}\left[\exp\left(\frac{1}{2}\int_0^T |\theta_s|^2\,ds\right)\right] < \infty.$$

We then recall that the stochastic exponential

$$\Theta_t := \exp\left(\int_0^t \theta_s \cdot dW_s - \frac{1}{2}\int_0^t |\theta_s|^2\,ds\right) \tag{4.1}$$

is a u.i. martingale for $0 \leqslant t \leqslant T$. By Girsanov's Theorem, we also have that under the tilted measure $\mathbb{P}^\theta$ defined by $d\mathbb{P}^\theta = \Theta_T d\mathbb{P}$, that the drifted Wiener process W_t^θ given by

$$dW_t^\theta := dW_t - \theta_t dt \tag{4.2}$$

is a Wiener process under $\mathbb{P}^\theta$ (see Section 5.6 of [LeG16]).

It is useful to denote the inverse likelihood ratio by

$$Z_t^\theta := \left.\frac{d\mathbb{P}}{d\mathbb{P}^\theta}\right|_{\mathcal{F}_t} = \exp\left(-\int_0^t \theta_s \cdot dW_s^\theta - \frac{1}{2}\int_0^t |\theta_s|^2\,ds\right). \tag{4.3}$$

Under $\mathbb{P}^\theta$, the controlled diffusion is

$$dX_t^\theta = \big(a(X_t^\theta) + \sigma(X_t^\theta)\theta_t\big)\,dt + \sigma(X_t^\theta)\,dW_t^\theta, \qquad X_0^\theta = x. \tag{4.4}$$

Therefore, for a given test function $\psi \in C_P^{2(\beta+1)}(\mathbb{R}^d;\mathbb{R})$ that (where we denote by $\mathbb{E}_\theta$ the expectation with respect to $\mathbb{P}^\theta$):

$$\mathbb{E}[\psi(X_t)] = \mathbb{E}_\theta[\psi(X_t^\theta)Z_t^\theta]. \tag{4.5}$$

In particular, we can rewrite our Feynman–Kac solutions as follows:

Corollary 11 (Controlled Feynman–Kac)**.** *Under the assumptions of Theorem 1, the solution to (2.2) can be written as*

$$u(t,x) = \mathbb{E}_\theta\left[e^{-\int_0^t q(X_s^\theta)\,ds} f(X_t^\theta)Z_t^\theta\right]. \tag{4.6}$$

The rewriting (4.6) of $u(t,\cdot)$ has the practical benefit of allowing one to control the weighting term Z_T^θ to concentrate on regions of importance. This method is thus often referred to as *Drift-Based Importance Sampling* or *Controlled Feynman–Kac* in the context of PDE solvers. A corresponding Monte–Carlo estimator $\hat{u}_M^\theta$ can then be written as follows:

$$\widehat{u}_M^\theta(T,x) := \frac{1}{M}\sum_{r=1}^{M} e^{-\int_0^T q(X_s^{(r)})\,ds} f(X_T^{(r)}) Z_T^{\theta,(r)}, \qquad \mathbb{E}_\theta[\widehat{u}_M^\theta(T,x)] = u(T,x), \tag{4.7}$$

where $X^{(r)}$ are independent samples of (4.4) under $\mathbb{P}^\theta$, and $Z_T^{\theta,(r)}$ are the corresponding likelihood weights. This estimator can be simulated numerically in an analogous way to as described in Section 2.

4.2. **The Zero-Variance Drift.** How does one construct Θ_T such that the corresponding estimator (4.7) has low variance? The following result due to [Mil95], is interesting from a theoretical point of view: let X be an Itô diffusion satisfying (1.8) and suppose $u(t,x) > 0$ everywhere. We show that, at least theoretically, one can construct a drift such that the reweighted sampler draws paths in such a way that the reweighting term Z_T^θ *exactly* cancels the sampling bias.

Proposition 12 (Zero-Variance Guiding Drift)**.** *Define the backward process $h(t,x) := u(T-t,x)$ and suppose that h is positive and smooth. Suppose the drift*

$$\theta_t^* = \sigma(X_t)^\top \nabla \log h(t, X_t). \tag{4.8}$$

satisfies Novikov's criterion. Then, the corresponding drifted SDE

$$dX_t = (a(X_t) + \sigma(X_t)\theta_t^*)\,dt + \sigma(X_t) dW_t^{\theta^*}, \qquad X_0 = x, \tag{4.9}$$

has a Girsanov estimator that is constant:

$$e^{-\int_0^T q(X_s)\,ds} f(X_T) Z_T^{\theta^*} = h(0,x) = u(T,x) \qquad \mathbb{P}^{\theta^*}\text{-a.s.} \tag{4.10}$$

Proof. By Remark 3 we note that h is the unique solution to (2.2) with boundary condition $v(T,x) = f(x)$. Let θ^* be as defined in (4.8). Under the tilted measure $\mathbb{P}^{\theta^*}$, Itô's lemma gives

$$\begin{aligned} d(\log h(t,X_t)) &= \left[\frac{\partial_t h + \mathcal{L}h}{h}(t,X_t) + \alpha_t \cdot \theta_t^* - \frac{1}{2}|\alpha_t|^2\right] dt + \alpha_t dW_t^{\theta^*} \\ &= \left[q(X_t) + \frac{1}{2}|\theta_t^*|^2\right] dt + \theta_t^* dW_t^{\theta^*}, \end{aligned}$$

where we used $\partial_t h + \mathcal{L}h = qh$. Integrating from 0 to T and using $h(T,X_T) = f(X_T)$, we get

$$\log f(X_T) - \log h(0,x) = \int_0^T q(X_s)\,ds + \int_0^T \theta_s^* \cdot dW_s^{\theta^*} + \frac{1}{2}\int_0^T |\theta_s^*|^2\,ds.$$

Rearranging and exponentiating both sides gives

$$\begin{aligned} h(0,x) &= e^{-\int_0^T q(X_s)\,ds} f(X_T) \exp\left(-\int_0^T \theta_s^* \cdot dW_s^{\theta^*} - \frac{1}{2}\int_0^T |\theta_s^*|^2\,ds\right) \\ &= e^{-\int_0^T q(X_s)\,ds} f(X_T) Z_T^{\theta^*}. \end{aligned}$$

This proves (4.10). Hence the estimator is equal to the deterministic constant $h(0,x) = u(T,x)$ almost surely under $\mathbb{P}^{\theta^*}$, and therefore its variance is zero. □

Remark 13. In practice, the construction of the optimal drift θ^* in Proposition 12 is tautological as it requires knowledge of the exact u that we are trying to estimate. Nevertheless, the above result shows that it is possible to obtain a substantial reduction in variance by choosing an appropriate reweighting of the estimator. A practical way of implementing such a method, would be to use a preexisting (rough) estimate $\bar{u}$ of u to calculate the appropriate importance sampling weights.

5. Conclusions and Bibliographical Notes

In this paper, we described a Monte–Carlo approach to solving a wide class of parabolic Cauchy problems by utilizing the Feynman–Kac formula. In particular, by utilizing Feynman–Kac, we were able to construct an estimator based on a sample mean of the functionals of approximated diffusions. We then analyzed the discretization errors of such estimators through the lens of weak convergence. Finally, we discussed variance reduction through a measure transformation method based on Girsanov's theorem.

There are many more important topics in the theory of numerical SDEs that we have sacrificed for the sake of brevity in this paper. In Section 2, for example, we considered only a time-homogenous form of Feynman–Kac for simplicity of presentation. For more general formulations of the statement for broader boundary conditions or Feller processes, we refer the reader to [Lal; KS91]. In Section 3, we considered only explicit time-discrete approximations for diffusive Itô SDEs and ignored stability concerns. For further discussion on the strong scheme, stability considerations, and an overview of implicit methods, extrapolation methods, and predictor-corrector methods, we refer the reader to Chapters 14-15 of [KP99], while for generalizations to SDEs with Poisson jumps, we refer the reader to [PB10]. In Section 4, we considered only variance reduction methods based on a change of measure. Further variance reduction methods, for example based on the more broad theory of Monte–Carlo integration, and the construction of unbiased estimators is discussed further in Chapter 16 of [KP99].

Acknowledgements

This paper was written for MIT's Stochastic Calculus (18.676) in spring 2026, taught by Prof. Nike Sun. I would like to thank Prof. Sun for helpful discussions during office hours and Arif Dayi, Michael Doboli, Pavan Yeddanapudi, and Jeonghyun Ahn for feedback on my final presentation.

References

[KP99] Peter E. Kloeden and Eckhard Platen. *Numerical Solution of Stochastic Differential Equations*. Corrected third printing. Applications of Mathematics 23. Berlin, Heidelberg: Springer-Verlag, 1999. ISBN: 978-3-642-08107-1. DOI: 10.1007/978-3-662-12616-5.

[KS91] Ioannis Karatzas and Steven Shreve. *Brownian Motion and Stochastic Calculus*. Vol. 113. Graduate Texts in Mathematics. Springer, 1991. ISBN: 978-0-387-97655-6. DOI: 10.1007/978-1-4612-0949-2.

[Lal] Steven P. Lalley. *Lecture 12: Stochastic Differential Equations, Diffusion Processes, and the Feynman-Kac Formula*. Lecture notes. University of Chicago, Statistics 391. URL: https://www.stat.uchicago.edu/~lalley/Courses/391/Lecture12.pdf (visited on 05/04/2026).

[LeG16] Jean-François LeGall. *Brownian Motion, Martingales, and Stochastic Calculus*. Vol. 274. Graduate Texts in Mathematics. Springer Verlag, 2016. ISBN: 978-3-319-31088-6. DOI: 10.1007/978-3-319-31089-3.

[LSU68] O. A. Ladyženskaja, V. A. Solonnikov, and N. N. Ural'ceva. *Linear and Quasi-linear Equations of Parabolic Type*. Vol. 23. Translations of Mathematical Monographs. Translated from the Russian by S. Smith. Providence, RI: American Mathematical Society, 1968, p. 648.

[Mil95] G. N. Milstein. *Numerical Integration of Stochastic Differential Equations*. 1st ed. Mathematics and Its Applications. Dordrecht: Springer Dordrecht, 1995. 172 pp. ISBN: 978-94-015-8455-5. DOI: 10.1007/978-94-015-8455-5.

[MP91] Remigius Mikulevicius and Eckhard Plate. "Rate of Convergence of the Euler Approximation for Diffusion Processes". In: *Mathematische Nachrichten* 151.1 (1991), pp. 233–239. DOI: https://doi.org/10.1002/mana.19911510114. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/mana.19911510114. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/mana.19911510114.

[Oks03] Bernt Oksendal. *Stochastic Differential Equations: An Introduction with Applications*. 6th ed. Universitext. Berlin, Heidelberg: Springer, 2003. ISBN: 978-3-540-04758-2. DOI: 10.1007/978-3-642-14394-6.

[PB10] Eckhard Platen and Nicola Bruti-Liberati. *Numerical Solution of Stochastic Differential Equations with Jumps in Finance*. Vol. 64. Stochastic Modelling and Applied Probability. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010. ISBN: 978-3-642-13694-8. DOI: 10.1007/978-3-642-13694-8.

Appendix: The Wagner-Platen Expansion

5.1. **Iterated Itô Integrals.** We begin by introducing some notation to simplify bookkeeping:

Definition 14 (Multi-indices)**.** Let $\mathcal{M}$ denote the set of all finite tuples

$$\alpha = (j_1, \dots, j_\ell), \qquad j_i \in \{0, 1, \dots, m\},$$

together with an "empty" tuple v. We refer to such tuples l as "multi-indices". We write $l(\alpha) = \ell$ for the length of α, and $n(\alpha)$ for the number of zero entries in α. By convention $l(v) = n(v) = 0$. If $\alpha = (j_1, \dots, j_\ell) \neq v$, we write

$$-\alpha = (j_2, \dots, j_\ell), \qquad \alpha- = (j_1, \dots, j_{\ell-1}),$$

with the convention that deleting the only entry gives v.

We now introduce a notion of an "iterated" or "multiple" Itô integral:

Definition 15 (Multiple Itô integrals)**.** For an adapted process H, stopping times ρ, τ such that $0 \leqslant \rho \leqslant \tau \leqslant T$ a.s., and a multi-index $\alpha = (j_1, \dots, j_\ell)$, define $I_\alpha[H]_{\rho,\tau}$ recursively by

$$I_v[H]_{\rho,\tau} = H_\tau,$$

and, for $\ell \geqslant 1$,

$$I_\alpha[H]_{\rho,\tau} = \begin{cases} \displaystyle\int_\rho^\tau I_{\alpha-}[H]_{\rho,s}\, ds, & j_\ell = 0, \\ \displaystyle\int_\rho^\tau I_{\alpha-}[H]_{\rho,s}\, dW_s^{j_\ell}, & j_\ell \in \{1, \dots, m\}. \end{cases}$$

When $H \equiv 1$, we simply write I_α for $I_\alpha[1]_{\rho,\tau}$.

Intuitively, an index 0 in the multi-index represents integration with respect to time, while indices $1, \dots, m$ represent integration with respect to the corresponding Wiener coordinates. Thus, for instance,

$$I_{(0)} = \tau - \rho, \qquad I_{(j)} = W_\tau^j - W_\rho^j,$$

and

$$I_{(i,j)} = \int_\rho^\tau \int_\rho^{s_2} dW_{s_1}^i\, dW_{s_2}^j,$$

with the convention $dW_s^0 = ds$.

5.2. **Itô Coefficient Functions.** We next introduce the differential operators which appear as stochastic Taylor coefficients. Let $B(t,x) = b(t,x)b(t,x)^\top$. Define

$$L^0 = \partial_t + \sum_{i=1}^{d} a_i(t,x)\partial_i + \frac{1}{2}\sum_{i,k=1}^{d} B_{ik}(t,x)\partial_{ik},$$
$$L^j = \sum_{i=1}^{d} b_i^j(t,x)\partial_i. \qquad j \in \{1,\cdots,m\}$$

These operators act componentwise on vector-valued functions.

Definition 16 (Itô coefficient functions)**.** Consider $\alpha \in \mathcal{M}$ and $f \in C^q(\mathbb{R}_+ \times \mathbb{R}^d; \mathbb{R}^r)$ for $q = l(\alpha) + n(\alpha)$. We define the *Itô coefficient function* f_α recursively by

$$f_\alpha = \begin{cases} f, & \alpha = v \\ f_\alpha = L^{j_1} f_{-\alpha} & \alpha = (j_1,\ldots,j_\ell) \neq v. \end{cases}$$

5.3. **The General Wagner–Platen Expansion.** We now state the main theorem of this section:

Theorem 17 (Wagner–Platen Expansion)**.** *Let $A \subset \mathcal{M}$ be a finite hierarchical set, meaning that $v \in A$ and $-\alpha \in A$ whenever $\alpha \in A\backslash\{v\}$. Define the remainder index set*

$$\mathcal{B}(A) := \{\alpha \in \mathcal{M}\backslash A : \ -\alpha \in A\}.$$

If f, a, and b are sufficiently smooth so that all terms below are well-defined, then for stopping times $\rho \leqslant \tau$,

$$f(\tau, X_\tau) = \sum_{\alpha\in A} I_\alpha[f_\alpha(\rho, X_\rho)]_{\rho,\tau} + \sum_{\alpha\in\mathcal{B}(A)} I_\alpha[f_\alpha(\cdot, X_\cdot)]_{\rho,\tau}.$$

For numerical purposes, one chooses a finite hierarchical set A and drops the remainder. The Wagner–Platen expansion corresponds to taking

$$\begin{aligned} A_{\mathrm{WP}} =& \{v, (0), (0,0)\} \\ &\cup \{(j), (j,0), (0,j) : 1 \leqslant j \leqslant m\} \\ &\cup \{(j_1,j_2) : 1 \leqslant j_1, j_2 \leqslant m\} \\ &\cup \{(j_1,j_2,j_3) : 1 \leqslant j_1, j_2, j_3 \leqslant m\}. \end{aligned}$$

Taking $f(t,x) = x$, setting $\rho = t$, $\tau = t+h$, and writing $I_\alpha = I_\alpha[1]_{t,t+h}$, the expansion becomes

$$\begin{aligned} X_{t+h} = X_t + a\, I_{(0)} + \sum_{j=1}^{m} b^j\, I_{(j)} + \sum_{j_1,j_2=1}^{m} L^{j_1} b^{j_2}\, I_{(j_1,j_2)} + \sum_{j=1}^{m} \left(L^j a\, I_{(j,0)} + L^0 b^j\, I_{(0,j)}\right) \\ + \sum_{j_1,j_2,j_3=1}^{m} L^{j_1} L^{j_2} b^{j_3}\, I_{(j_1,j_2,j_3)} + L^0 a\, I_{(0,0)} + R_{\mathrm{WP}}. \end{aligned} \tag{5.1}$$

Here all coefficient functions are evaluated at (t, X_t), and

$$R_{\mathrm{WP}} = \sum_{\alpha\in\mathcal{B}(A_{\mathrm{WP}})} I_\alpha[x_\alpha(\cdot, X_\cdot)]_{t,t+h},$$

where $x(t,x) = x$. Under the usual smoothness and polynomial growth assumptions, the one-step remainder satisfies

$$\|R_{\mathrm{WP}}\|_{L^2} = O(h^2).$$

Consequently, the scheme obtained by dropping R_{WP} is the Wagner–Platen scheme, a strong Taylor scheme of order $3/2$. Equivalently, on a grid $0 = t_0 < \cdots < t_N = T$ with $h_k = t_{k+1} - t_k$, the Wagner–Platen scheme is

$$Y_{k+1} = Y_k + a\, I_{(0)} + \sum_{j=1}^{m} b^j\, I_{(j)} + \sum_{j_1,j_2=1}^{m} L^{j_1} b^{j_2}\, I_{(j_1,j_2)} + \sum_{j=1}^{m} \left(L^j a\, I_{(j,0)} + L^0 b^j\, I_{(0,j)} \right) + \sum_{j_1,j_2,j_3=1}^{m} L^{j_1} L^{j_2} b^{j_3}\, I_{(j_1,j_2,j_3)} + L^0 a\, I_{(0,0)}. \tag{5.2}$$

In (5.2), the multiple Itô integrals are taken over $[t_k, t_{k+1}]$, and all coefficient functions are evaluated at (t_k, Y_k).

5.4. **Higher Order Weak Schemes.** We conclude the appendix by giving an example of a higher order weak-scheme taken by adding the double integral terms in the Wagner–Platen expansion. To avoid excessive notation, we describe only the autonomous scalar case $d = m = 1$ with deterministic times $t_n = n\delta$. We denote

$$\Delta Z_n := I_{(1,0)} = \int_{t_n}^{t_{n+1}} \int_{t_n}^{s} dW_r\, ds = \int_{t_n}^{t_{n+1}} (W_s - W_{t_n})\, ds.$$

$$I_{(1,1)} = \frac{1}{2}\left((\Delta W_n)^2 - \delta \right), \qquad I_{(0,1)} = \delta \Delta W_n - \Delta Z_n, \qquad I_{(0,0)} = \frac{\delta^2}{2}.$$

Substituting these identities into the scalar Itô–Taylor expansion gives the following scheme.

Algorithm 3 (Order 2.0 Weak Taylor Scheme). *Let $\delta > 0$ be given. Consider previsible, independent $\Delta W_n, \Delta Z_n$ such that*

$$\Delta W_n = W_{t_{n+1}} - W_{t_n}, \qquad \Delta Z_n = \int_{t_n}^{t_{n+1}} (W_s - W_{t_n})\, ds.$$

Equivalently, one may generate this Gaussian pair by taking independent standard normal random variables ξ_n, η_n and setting

$$\Delta W_n = \sqrt{\delta}\, \xi_n, \qquad \Delta Z_n = \frac{\delta}{2} \Delta W_n + \sqrt{\frac{\delta^3}{12}}\, \eta_n.$$

The order 2.0 weak Taylor scheme is then

$$Y_{n+1} = Y_n + a\delta + b\Delta W_n + \frac{1}{2} b b' \left((\Delta W_n)^2 - \delta \right) + a' b\, \Delta Z_n + \frac{1}{2}\left(a a' + \frac{1}{2} a'' b^2 \right) \delta^2 + \left(a b' + \frac{1}{2} b'' b^2 \right) (\delta \Delta W_n - \Delta Z_n), \tag{5.3}$$

where all coefficients and derivatives are evaluated at Y_n.

As in Algoirthm 2, we remark that the increments variables ΔW_n and ΔZ_n may be replaced by simpler variables provided that the required joint moments are matched to sufficiently high order.