

Numerical Methods for Stochastic Differential Equations

Abstract

We study the Euler-Maruyama and Milstein numerical methods to solve an autonomous scalar Itô stochastic differential equation in the context of machine learning minimization. We present a brief introduction to Brownian motion and stochastic integration as well as an analysis of the numerical stability of those methods accompanied by numerical simulations motivated by standard ODE methods.

1 Introduction

Stochastic differential equations (SDEs) are, loosely speaking, differential equations in which one or more terms are stochastic processes which produce a solution that is also a stochastic process. Precisely because SDEs combine both random and deterministic effects into a model that is capable of describing the time evolution of degrees of freedom, they play a central role in science and engineering. In particular, SDEs serve as models for molecular dynamics, quantum chemistry, Bayesian statistics, weather forecasting, machine learning, and econometrics.

Now, as discussed in [5], SDEs can also play an important role in solving minimization problems, especially in machine learning. These minimization problems are usually of the form

$$\min_{x \in \mathbb{R}^d} f(x), \quad f(x) := \frac{1}{p} \sum_{i=1}^p f_i(x), \quad (1)$$

where $f_i: \mathbb{R}^d \rightarrow \mathbb{R}$ for $i = 1, \dots, p$. In common machine learning terms, x is the vector of trainable parameters; f represents the total loss function with each f_i being the loss due to the i -th training sample; and p is the training sample size, which can be extremely large [5]. This means that using standard gradient descent – which requires p gradient evaluations per step – is prohibitively expensive when $p \gg 1$. A proposed alternative to avoid this issue is the so called stochastic gradient descent, which replaces the full gradient ∇f by a “sampled version” of it that serves as its (unbiased) estimator. More precisely, the simplest stochastic gradient descent is of the form

$$x_{j+1} = x_j - \eta \nabla f_{U_j}(x_j), \quad j \in \mathbb{N}, \quad (2)$$

where $\eta \in \mathbb{R}$ is the learning rate, and $\{U_j\}_{j \in \mathbb{N}}$ is a collection of i.i.d. uniform random variables taking values in $\{1, \dots, p\}$. This method in Equation (2) has the advantage that, after sampling only a few indices from the full gradient ∇f , its computational complexity is independent of p per iterate [5]. For this reason, stochastic gradient descent has become a widely used algorithm in large scale problems.

However, even though there are many convergence results for stochastic gradient descent methods, the majority of these results are upper-bounds for strongly convex objective functions, which often do not characterize the behavior of systems in real-world settings [5]. So, translating the theoretical understanding of these methods into actual algorithms can be a very hard task. An alternative approach proposed by [5] is to rewrite the stochastic gradient descent of Equation (2) as the SDE

$$dX_t = F(X_t) dt + G(X_t) dB_t, \quad (3)$$

where F, G are a real-valued functions, and $(B_t)_{t \geq 0}$ is a standard Brownian motion. As shown in [5], under mild assumptions on f , the solution of the SDE above converges to the solution of the minimization problem in the weak topology. The advantage of considering this SDE instead of the stochastic gradient descent is that we can employ the exuberant flora of numerical methods for SDEs to study a larger range of real-world problems.

Now, with this connection between minimization problems and SDEs, certain natural questions appear about the stability, accuracy, and bias of numerical methods for SDEs. For instance, when modeling scenarios where long-time trajectories are relevant, understanding and predicting the stability of methods are extremely important;

e.g., so that we can increase the time steps in our scheme and trade accuracy for physical insights that would not be available otherwise (see [5] for more examples). In this sense, the goal of this article is to understand the long-time moment stability of two well-known methods for approximating solutions to SDEs: the Euler-Maruyama method and the Milstein method. The analysis in this article will be focused on the family of SDEs used by [5] to rewrite minimization problems (see Section 2 of [5]). Finally, we implement these methods, and perform numerical experiments with different time step sizes to verify the analytic findings.

Outline. In Section 2, we review a few background concepts from probability and stochastic calculus that are useful throughout the paper. In Section 3, we introduce the Euler-Maruyama and Milstein methods; we implement them to obtain the Ornstein-Uhlenbeck process as the solution to an SDE; and, lastly, we prove a few theoretical results on the long-time stability and accuracy of these methods. Finally, in Section 4, we perform numerical experiments to verify the theoretical findings of the previous section.

2 Preliminaries: Brownian Motion and Stochastic Calculus

In this section, we review a few concepts that will be relevant throughout this article when studying stochastic processes (a much more detailed presentation of these topics may be found in [4] and [7]). Stochastic processes describe dynamical systems that evolve probabilistically with time. More precisely, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $(E, \mathcal{G})$ be a measurable space. A *stochastic process* is a collection of random variables $(X_t: t \geq 0)$ such that, for each $t \geq 0$, X_t is a random variable from $(\Omega, \mathcal{F}, \mathbb{P})$ to $(E, \mathcal{G})$. The set E is called the *state space* of X_t , and Ω is called the *sample space*. Note that a stochastic process X_t is a function of both $t \geq 0$ and $\omega \in \Omega$. For a fixed sample point $\omega \in \Omega$, we call the function $t \mapsto X_t(\omega)$ a *sample path* of X .

For $p \geq 1$, we denote by $L^p(\Omega, \mathcal{F}, \mathbb{P})$, or simply L^p when the underlying probability space is clear, the space of real random variables X such that $|X|^p$ is integrable (with the usual identification of random variables that are equal a.s.), and we equip it with the usual norm. For $n \in \mathbb{N}$, define as $\mu_n := \mathbb{E}[X^n]$ the n -th *moment* of a random variable X , where $\mathbb{E}[\cdot]$ denotes the expectation of a random variable. A real random variable X is said to be *Gaussian* with mean m and standard deviation σ^2 , denoted $X \sim N(m, \sigma^2)$, if its law has density

$$p_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-m)^2/2\sigma^2}. \quad (4)$$

Now, one of the most important stochastic processes is Brownian motion (also known as Wiener process). We say that a real random process $(B_t)_{t \geq 0}$ is a (*standard*) *Brownian motion* if the following two conditions hold:

- (i) $B_0 = 0$ a.s., and, for every choice of $0 = t_0 < t_1 < \dots < t_p$, the increments $B_{t_i} - B_{t_{i-1}}$, $1 \leq i \leq p$, are independent and distributed as $N(0, t_i - t_{i-1})$;
- (ii) All sample paths of B are continuous.

Define a *filtration* $(\mathcal{F}_t)_{t \geq 0}$ as a collection of σ -algebras on $(\Omega, \mathcal{F}, \mathbb{P})$ such that $\mathcal{F}_s \subset \mathcal{F}_t$ for every $s \leq t$. Now, we say that a real-valued random process $(X_t)_{t \geq 0}$ such that $X_t \in L^1$ for every $t \geq 0$ is an *adapted martingale* with respect to a filtration $(\mathcal{F}_t)_{t \geq 0}$ if X_t is $\mathcal{F}_t$ -measurable for all $t \geq 0$, and we have that $\mathbb{E}[X_t | \mathcal{F}_s] = X_s$ for every $0 \leq s < t$. As an example, Brownian motion $(B_t)_{t \geq 0}$ is a martingale adapted to the canonical filtration $\mathcal{F}_t = \sigma(B_s: s \leq t)$.

Let $(X_t)_{t \geq 0}$ be a martingale. We define the *(Itô) stochastic integral* as

$$\int_0^t X_s dX_s := \lim_{n \rightarrow \infty} \sum_{i=0}^{p_n} X_{t_i^n} (X_{t_{i+1}^n} - X_{t_i^n}), \quad (5)$$

where the limit converges in probability, and the partitions $0 = t_0^n < t_1^n < \dots < t_{p_n}^n = t$ of $[0, t]$ get more and more refined as $n \nearrow \infty$. It is possible to define more general stochastic integrals, but the integral in Equation (5)

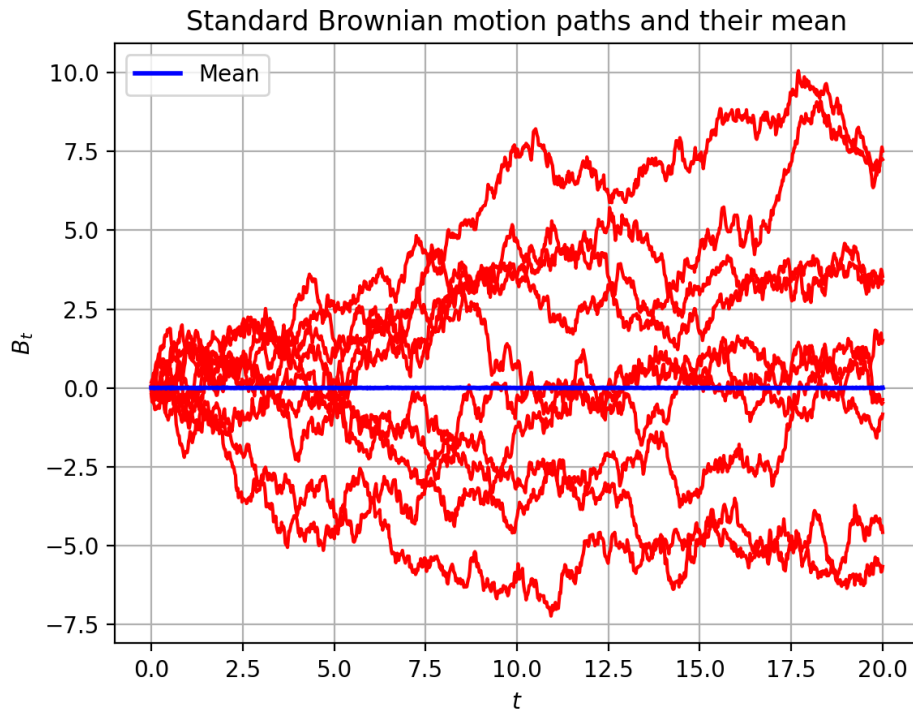


Figure 1: Plot of 10 individual paths of a standard Brownian motion and their mean at each time.

is enough for the purposes of this article. For convergence results and in-depth analysis of stochastic integrals, we direct the reader to [4] and Chapter 3 of [7].

Finally, the *autonomous scalar Itô SDE* is given by

$$dX_t = f(X_t) dt + g(X_t) dB_t, \quad t \in [0, T], \quad (6)$$

where f, g are real-valued functions. Note that, if $g \equiv 0$ (i.e., no noise coefficient), this becomes a familiar *deterministic* ODE

$$\frac{dX_t}{dt} = f(X_t), \quad t \in [0, T]. \quad (7)$$

It is worth noting that the form in which Equation (6) is written is just a notational convenience since the “infinitesimals” there have no mathematical meaning. We can write Equation (6) more precisely as

$$X_t - X_0 = \int_0^t f(X_s) ds + \int_0^t g(X_s) dB_s, \quad t \in [0, T]. \quad (8)$$

3 Euler-Maruyama and Milstein Methods

In this section we introduce the Euler-Maruyama and Milstein methods to iteratively approximate solutions to SDEs following Chapter 5 of [7] and [3]. Later in this section, we analyze the numerical stability of those methods by comparing the first and second moments of the theoretical solutions with the moments of the approximate solutions. For the purposes of this article, we consider the scalar autonomous Itô SDE given by

$$dX_t = f(X_t) dt + g(X_t) dB_t, \quad t \in [0, T], \quad (9)$$

with a deterministic initial condition $X_0 \in [0, 1]$, where $(B_t)_{t \geq 0}$ is a standard Brownian motion, and the functions f, g are affine and do not depend on time. In order to approximate solutions, we discretize our time interval $[0, T]$ into n equispaced nodes so that the length of each time subinterval is $h = T/n$. By doing this, we generate a time grid $t_j = jh$, for $j = 0, \dots, n$. To simplify notation, let x_j denote the *approximate* solution at time t_j for $j = 0, \dots, n$. Moreover, let

$$f_j := f(x_j) \quad \text{and} \quad f'_j := f'(x_j), \quad j = 0, \dots, n, \quad (10)$$

and similarly for g_j and g'_j . Lastly, in order to simulate the Brownian steps, define a sequence $(\Delta B_j)_{j=0}^n$ of i.i.d. random variables $\Delta B_j \sim N(0, h)$ for $j = 0, \dots, n$.

The Euler-Maruyama method (abbreviated here as EM) is given by

$$x_{j+1} = x_j + f_j h + g_j \Delta B_j, \quad j = 0, \dots, n. \quad (11)$$

The Milstein method (abbreviated here as Mi) is given by

$$x_{j+1} = x_j + f_j h + g_j \Delta B_j + \frac{1}{2} g'_j g_j [(\Delta B_j)^2 - h], \quad j = 0, \dots, n. \quad (12)$$

Note that EM is very similar to the well-known Euler method for iteratively approximating solutions to ODEs; in particular, EM takes Euler's method and adds a stochastic correction. Moreover, Mi takes the formula for EM and adds another correction term that depends on $(\Delta B_j)^2$, in close resemblance to Itô's formula. As shown in Chapter 5 of [7] and Section 2 of [3], the convergence of the EM method is of order 1 (with respect to the weak topology) and the convergence of the Mi method is of order 1 (also in the weak topology).

3.1 Example: Ornstein-Uhlenbeck Process

Now, as an example, we implement the methods above to numerically approximate the solution to a well-known SDE. Consider the following initial value problem

$$dX_t = -\frac{X_t - \mu}{\tau} dt + \sigma \sqrt{\frac{2}{\tau}} dB_t, \quad X_0 = 0, \quad \text{for } t \in [0, T], \quad (13)$$

where $\mu, \tau, \sigma \in \mathbb{R}_+$ are nonnegative constants, and $(B_t)_{t \geq 0}$ is a standard Brownian motion. The solution to the SDE above is called the *Ornstein-Uhlenbeck process*, and it is commonly used in physics to model the velocity of a Brownian particle (of positive mass) under a dissipative force [Chapter 3 of 7].

See Figure 2 for Ornstein-Uhlenbeck processes numerically computed for different choices of constants. For these simulations, we considered the time interval $[0, 1]$, and we discretized it with $n = 1000$ equispaced nodes; i.e., the time steps used were of length $h = 0.001$. In this case, the coefficient functions f, g are

$$f(x) = -\frac{x - \mu}{\tau}, \quad g(x) = \sigma \sqrt{\frac{2}{\tau}}. \quad (14)$$

So, the iterative scheme for *both* EM and Mi is given by

$$x_{j+1} = x_j + \left(-\frac{x_j - \mu}{\tau} \right) h + \sigma \sqrt{\frac{2}{\tau}} \Delta B_j, \quad j = 0, \dots, n, \quad (15)$$

since $g' = 0$. Now, even though both EM and Mi produce the same iterative scheme in this case, we can clearly see the non-deterministic character of the solution to this differential equation. As indicated by Figure 2, the larger the magnitude of the dB_t term, the “noisier” the solution process. This is expected since the Brownian motion term is responsible for introducing “randomness” to the equation.

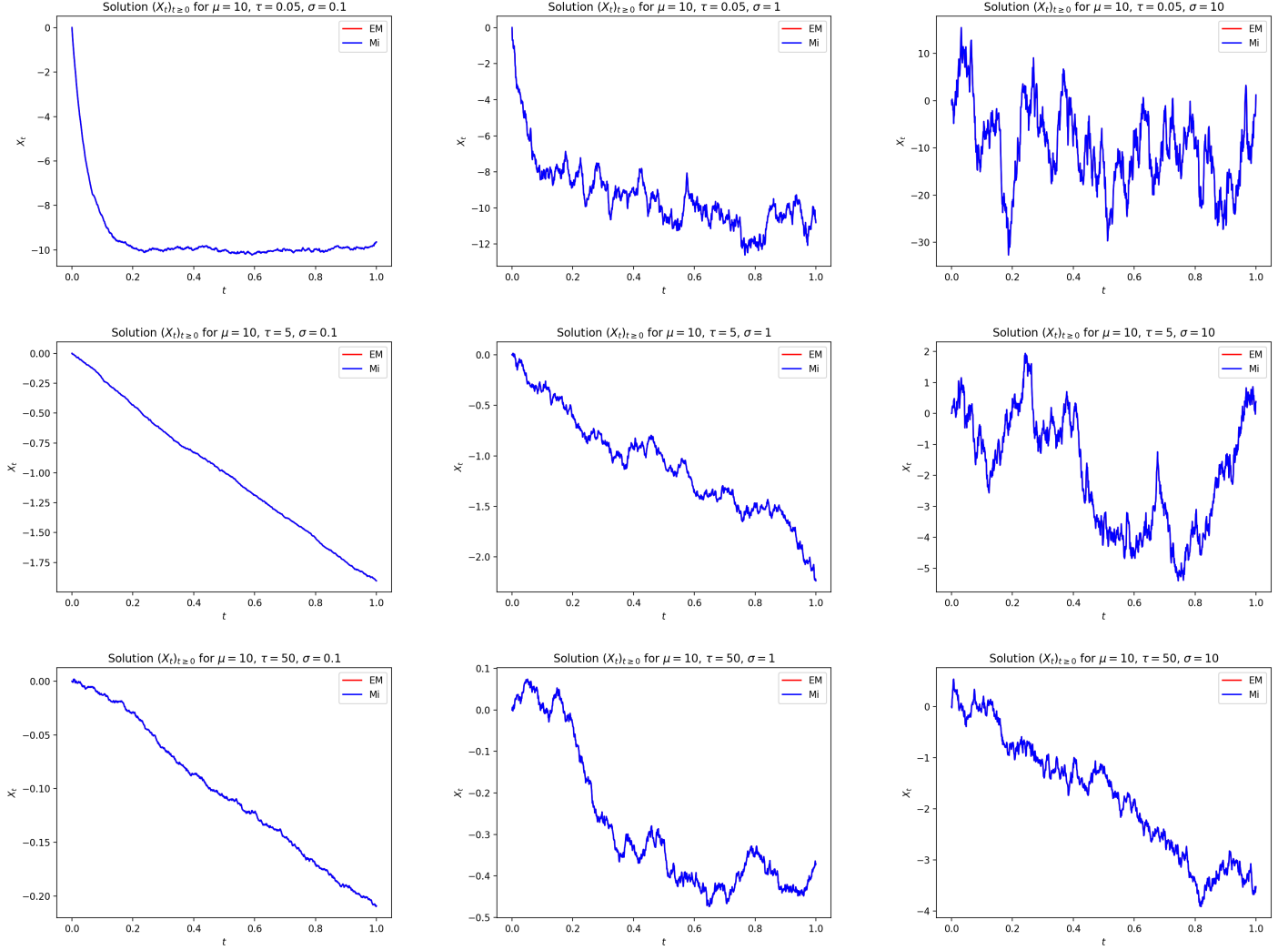


Figure 2: Approximate solutions for Equation (13) with different parameters τ and σ . The same mean $\mu = 10$ and initial condition $X_0 = 0$ were used for the simulations presented in this figure. The plots clearly show that the solution seems “more deterministic” when the dt term dominates, and “noisier” when the dB_t term dominates.

3.2 Stability

Now, we turn our attention to the asymptotic moment stability of the methods EM and Mi described above following the discussions in [2], [6], and Chapter 5 of [7]. In order to study the asymptotic stability of EM and Mi with different time steps h , we define the m -th moment of the approximate solution x_j at the j -th time step as

$$\mu_j^{(m)} := \mathbb{E}[(x_j)^m], \quad j = 0, \dots, n. \quad (16)$$

As in [6], we say that a numerical scheme with time step h of an SDE is asymptotically stable for the m -th moment if we have that

$$\limsup_{j \rightarrow \infty} \left| \mu_j^{(m)} - \mathbb{E}[(X_{t_j})^m] \right| < \infty, \quad (17)$$

where X_{t_j} is the continuous-time solution to the SDE at time t_j . Moreover, following [7], we say that a numerical scheme is asymptotically $O(h^j)$ accurate for the m -th moment if, for sufficiently small h , we have that

$$\limsup_{j \rightarrow \infty} \left| \mu_j^{(m)} - \mathbb{E}[(X_{t_j})^m] \right| = O(h^j). \quad (18)$$

Note that, by the above definitions, in order for a numerical scheme to be asymptotically accurate, it must be stable for the same moment [Chapter 5, 7]. For the purposes of this article, we will be mostly interested in first and second asymptotic moment stability and accuracy.

Finally, following the case reductions in [3], [5], and [Chapter 5, 7], we study the asymptotic moment stability and accuracy of EM and Mi for the SDE

$$dX_t = -X_t dt + (1 - \eta X_t) dB_t, \quad (19)$$

where $|\eta| < \sqrt{2}$ and $(B_t)_{t \geq 0}$ is a standard Brownian motion; that is, Equation (6) with $f(x) = -x$ and $g(x) = 1 - \eta x$. As shown in Section 3.3 of [3], the first two moments of the continuous-time solution process of Equation (19) are

$$\mathbb{E}[X_t] = X_0 e^{-t}, \quad (20)$$

and

$$\mathbb{E}[(X_t)^2] = \begin{cases} X_0^2 e^{-t(2-\eta^2)} + \frac{1}{2-\eta^2} (1 - e^{-t(2-\eta^2)}) + \frac{2\eta X_0}{1-\eta^2} (e^{-t} - e^{-t(2-\eta^2)}), & \text{if } \eta \neq \pm 1, \\ X_0^2 e^{-t} - e^{-t} + 1 \pm 2X_0 t e^{-t}, & \text{if } \eta = \pm 1 \end{cases}. \quad (21)$$

As proven also in Section 3.3 of [3], provided that $\eta^2 < 2$, we have that $\mathbb{E}[(X_t)^2] \rightarrow 1/(2-\eta^2)$ as $t \rightarrow \infty$.

3.2.1 Asymptotic Moment Stability of Euler-Maruyama (EM)

Consider the SDE in Equation (19). Taking expectations on the definition of EM (Equation (11)), and using the fact that $\Delta B_j \sim N(0, h)$ is independent of x_j for all $j = 0, \dots, n$, we obtain

$$\begin{aligned} \mu_{j+1}^{(1)} &= \mathbb{E}[x_{j+1}] = \mathbb{E}[x_j - x_j h + (1 - \eta x_j) \Delta B_j] \\ &= \mathbb{E}[(1 - h)x_j + (1 - \eta x_j) \Delta B_j] \\ &= (1 - h) \mathbb{E}[x_j] \\ &= (1 - h) \mu_j^{(1)}. \end{aligned} \quad (22)$$

Thus, $\mu_j^{(1)} = x_0(1 - h)^j$. Note that, similarly to Euler's method for approximating deterministic ODEs, in order to have asymptotic stability, we must require the time step h to be such that $|1 - h| < 1$; i.e., $h \in (0, 2)$. If this is the case, then $\mu_j^{(1)} \rightarrow 0$ as $j \rightarrow \infty$, so EM here is first moment asymptotically stable and unbiased [3, Chapter 5 of 7].

Similarly, we have that the second moment of EM's approximate solution to Equation (19) is

$$\begin{aligned} \mu_{j+1}^{(2)} &= \mathbb{E}[(x_{j+1})^2] = \mathbb{E}[(1 - h)^2 x_j^2] + \mathbb{E}[(1 - \eta x_j)^2 (\Delta B_j)^2] \\ &= \mu_j^{(2)} [(1 - h)^2 + h\eta^2] + 2\eta h \mu_j^{(1)} + h. \end{aligned} \quad (23)$$

Note that we now have a recursive formula for the second moment. So, using the standard linear analysis method introduced by [1, 2], EM is second moment asymptotically stable here whenever

$$1 > \left| \frac{\partial}{\partial \mu_j^{(2)}} \left(\mu_j^{(2)} [(1 - h)^2 + h\eta^2] + 2\eta h \mu_j^{(1)} + h \right) \right| = |(1 - h)^2 + \eta^2 h|; \quad (24)$$

i.e., whenever

$$0 < h < 2 - \eta^2. \quad (25)$$

As expected, second moment stability is more restrictive than first moment stability.

Lastly, under the constraints on h obtained above, this EM scheme is asymptotically stable for the first and second moments. Thus, $\mu_j^{(2)}$ converges to, say, $\mu_\infty^{(2)}$ as $j \rightarrow \infty$, and we have that

$$\mu_\infty^{(2)} = \mu_\infty^{(2)} [(1-h)^2 + h\eta^2] + h \quad (26)$$

since $\mu_j^{(1)} \rightarrow 0$ as $j \rightarrow \infty$ under the constraints above. So, by Equation (21),

$$\mu_\infty^{(2)} = \frac{1}{2-h-\eta^2} = \mathbb{E}[X_{t_j}^2] + O(h); \quad (27)$$

i.e., this EM scheme is asymptotically first order accurate for the second moment [3].

3.2.2 Asymptotic Moment Stability of Milstein (Mi)

Once again, consider the SDE in Equation (19). Now, the Mi method defined in Equation (12) gives

$$x_{j+1} = (1-h)x_j + (1-\eta x_j)\Delta B_j - \frac{1}{2}\eta(1-\eta x_j)((\Delta B_j)^2 - h). \quad (28)$$

So, since x_j is independent of ΔB_j for all $j \in \mathbb{N}$, after taking expectations on both sides of the equation above, we obtain that $\mu_{j+1}^{(1)} = (1-h)\mu_j^{(1)}$, just as for EM. That is, $\mu_j^{(1)} = x_0(1-h)^j$, as before. With this, we again find that, provided that $h \in (0, 2)$, this Mi scheme is asymptotically first moment stable (and unbiased since $\mu_j^{(1)} \rightarrow 0$ as $j \rightarrow \infty$) [6, 3, Chapter 5 of 7].

Now, we turn to the second moment of this Mi. After squaring and taking expectations on Equation (28), we find that

$$\mu_{j+1}^{(2)} = \mathbb{E}[x_{j+1}^2] = (1-h)^2\mu_j^{(2)} + h \left(1 + 2\eta\mu_j^{(1)} + \eta^2\mu_j^{(2)}\right) + \frac{1}{2}\eta^2h^2 \left(\eta^2\mu_j^{(2)} + 2\eta\mu_j^{(1)} + 1\right). \quad (29)$$

Thus, following [2, 6] and assuming that $h \in (0, 2)$ so that $\mu_j^{(1)} \rightarrow 0$, Mi is asymptotically stable for the second moment if

$$\begin{aligned} 1 &> \left| \frac{\partial}{\partial \mu_j^{(2)}} \left((1-h)^2\mu_j^{(2)} + h \left(1 + 2\eta\mu_j^{(1)} + \eta^2\mu_j^{(2)}\right) + \frac{1}{2}\eta^2h^2 \left(\eta^2\mu_j^{(2)} + 2\eta\mu_j^{(1)} + 1\right) \right) \right| \\ &= \left| 1 + h^2 \left(1 + \frac{\eta^4}{2}\right) - h(2 - \eta^2) \right|; \end{aligned} \quad (30)$$

i.e., if

$$0 < h < \frac{2 - \eta^2}{1 + \frac{\eta^4}{2}}. \quad (31)$$

Comparing the above condition on h for Mi to be second moment asymptotically stable with the one for EM, we see that the condition for Mi is more restrictive than for EM (cf. Equation (25)).

Finally, note that, when Mi is asymptotically second moment stable here, in the limit,

$$\mu_\infty^{(2)} = \frac{\frac{h\eta^2}{2} + 1}{2 - \eta^2 - h \left(1 + \frac{\eta^4}{2}\right)} = \frac{1}{2 - \eta^2} + O(h). \quad (32)$$

Thus, this Mi is asymptotically first order accurate for the second moment [6].

4 Numerical Experiments

We performed a series of numerical simulations (see Appendix A for the code used) to verify the analytic results presented in the previous section about the stability and accuracy of EM and Mi in the setting of Equation (19). All experiments were performed in a laptop with 16 GB of RAM and an Intel 8th Gen. Core i7-8550U CPU. We considered the time interval $[0, 20]$ for all the numerical experiments that follow. See Figure 3 for a comparison between the approximate first moments of EM and Mi with the theoretical first moment in Equation (20). For these numerical simulations, we used values of $\eta \in \{0.3, 1.4\}$ and time steps $h \in \{0.1, 0.01, 0.001\}$. From the plots in Figure 3, we can clearly see that EM and Mi are asymptotically first moment stable and accurate here for both values of η , as their “trajectories” are uniformly very close to the theoretical values – even for the largest time step $h = 0.1$. Finally, since the trajectories for these first moments appear to decay to zero as time passes, we see that EM and Mi are asymptotically first moment unbiased, in accordance with the analytic results from Section 3.2.

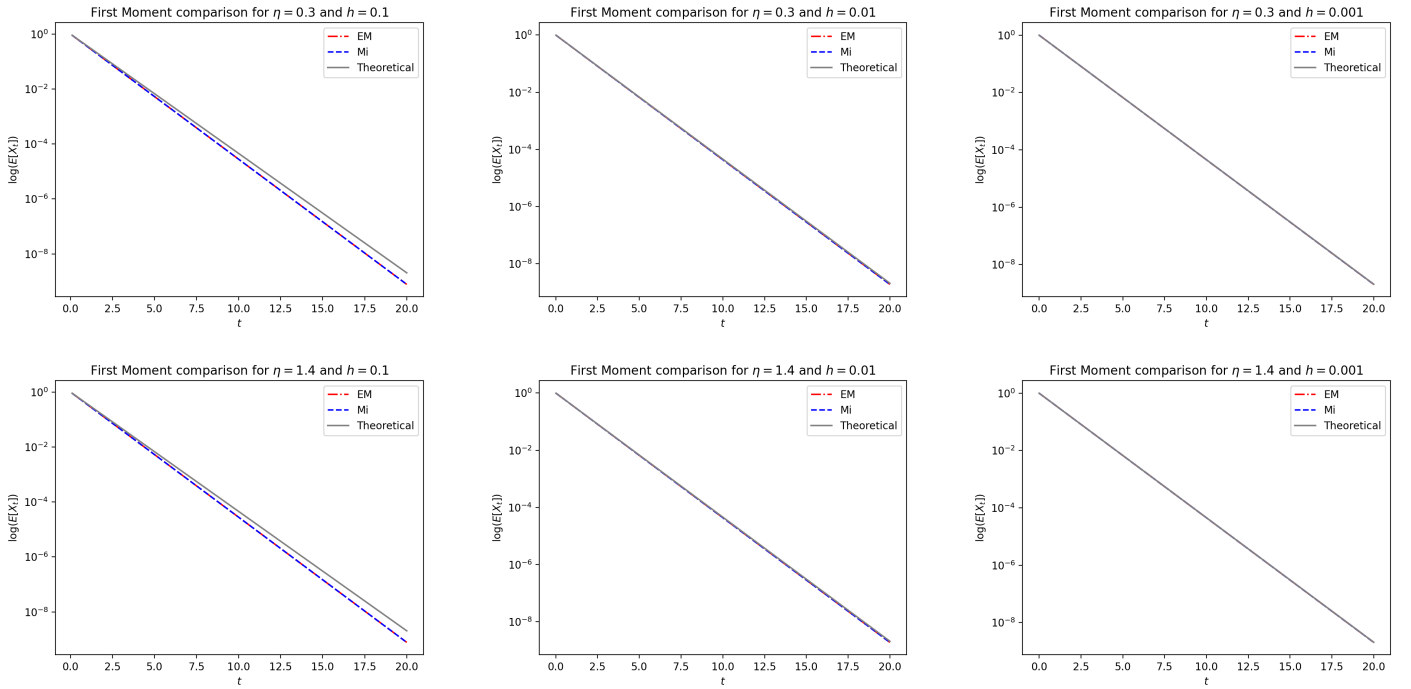


Figure 3: Comparison between the analytic first moment and the first moments of approximate solutions obtained via EM and Mi. The y -axis is shown in logarithmic scale. The first row shows first moments for $\eta = 0.3$ and the second row for $\eta = 1.4$, each row containing three different time steps h .

These plots indicate that EM and Mi are asymptotically first moment stable and accurate [3].

Now, see Figure 4 for a comparison between the theoretical and approximate second moments. Again, we used values of $\eta \in \{0.3, 1.4\}$ and time steps $h \in \{0.1, 0.01, 0.001\}$. Here, only with sufficiently small time steps h we see the second moments converging to the theoretical result from Equation (21). This is expected since we have stricter conditions on h for second moment asymptotic stability (cf. Equations (25) and (31)). Moreover, for $\eta = 0.3$, we clearly see the cancellation error expected to appear for early times in our EM and Mi approximations. Despite this loss of accuracy for small times, for $\eta = 0.3$, we see the predicted asymptotic convergence for finer time grids. For $\eta = 1.4$, the second moments of EM and Mi appear to be more sensitive to h , and they diverge for the largest time step $h = 0.1$. Still for $\eta = 1.4$, we only see convergence to the true second moment when $h = 0.001$. Finally, Mi is observed to be the most biased here, as its trajectories are always the furthest away from the theoretical curves.

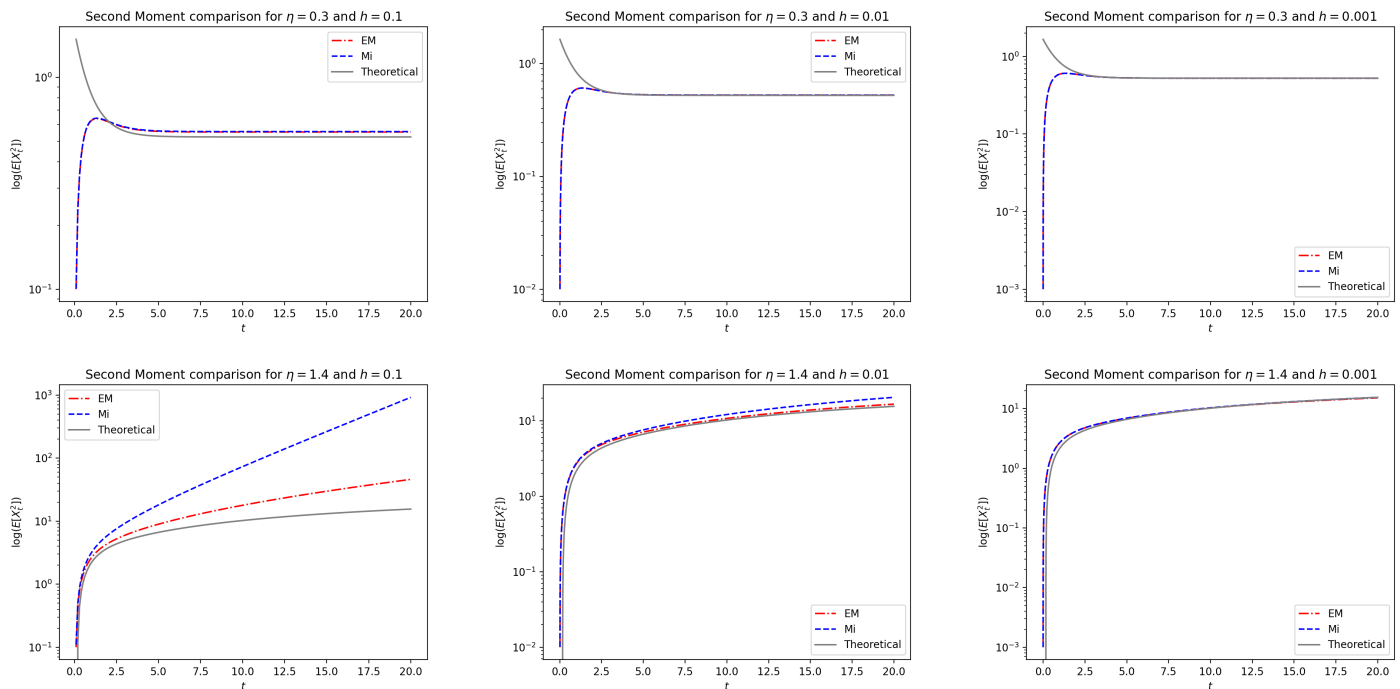


Figure 4: Comparison between the analytic second moment and the second moments of approximate solutions obtained via EM and Mi. The y -axis is shown in logarithmic scale. The first row shows second moments for $\eta = 0.3$ and the second row for $\eta = 1.4$, each row containing three different time steps h . The discrepancy in early times between second moments in the first row is due to cancellation error in the discrete second moment formula. This leads to loss of accuracy in the solutions on that region. Still, these plots indicate that EM and Mi are *asymptotically* second moment stable for small time steps h . Here, Mi appears to be somewhat more biased than EM [3].

5 Conclusions

Motivated by SDEs that can be used to solve minimization problems in machine learning, we showed that the Euler-Maruyama and Milstein methods are first and second moment asymptotically stable provided that our time steps were sufficiently small. In addition, we showed that both of these methods are second moment asymptotically accurate, and that they are first moment asymptotically accurate and unbiased – again, for small enough time steps. From numerical experiments, we observed that second moment asymptotic stability and accuracy in both methods are much more sensitive on time step sizes compared to first moment. Moreover, we also noted that we ought to be careful with potential cancellation errors introducing loss of accuracy in the second moments for small times. Lastly, natural next steps would be to test these iterative schemes for SDEs with nonlinear coefficient functions, and for SDEs with noise term coming from other appropriate stochastic processes apart of standard Brownian motion.

A Python Code

Find below the Python code used to generate the Brownian motion paths, implement the EM and Mi methods, and perform the asymptotic moment simulations shown in this article.

```
#!/usr/bin/env python3
#
import numpy as np
import matplotlib
import matplotlib.pyplot as plt
```

```

import scipy as sc
#
#
#
eps=1.0*10**(-14)
pi=np.pi
e=np.e
zero=0.0
done=1.0
#
#
#
def wiener(xi,n,dt,out=None):
    '''
    This subroutine generates a standard Wiener process with n steps.

    Input parameters:

    xi - starting position of Wiener process n - number of steps
    dt - time step

    Output parameters:

    out - array of positions at each time for the Wiener process
    '''
    xi=np.asarray(xi)

    gv=sc.stats.norm.rvs(scale=np.sqrt(dt),size=xi.shape+(n,))

    m=np.mean(gv,axis=1)

    if out is None:
        out=np.empty(gv.shape)

    np.cumsum(gv,axis=-1,out=out)
    out+=np.expand_dims(xi,axis=-1)

    return out,m
#
#
#
#####
#
# This is the beginning of the testing code proper for the Euler-Maryuama
# and Milstein methods for solving the autonomous scalar Ito SDE with
# affine coefficient functions.
#
#####
#
#
#

```

```

n=1000
t=1
dt=t/n

m=10

xi=np.zeros(n)
w=np.empty((m,n))

w,av=wiener(xi[:],n,dt,out=None)
tt=np.linspace(zero,n*dt,n)

plt.figure(dpi=200)
for i in range(m):
    plt.plot(tt,w[i],color='r')
plt.plot(tt,av,linewidth=2,color='b',label='Mean')
plt.xlabel(r'$t$')
plt.ylabel(r'$B_t$')
plt.title(r'Standard Brownian motion paths and their mean')
plt.grid(True)
plt.legend()
plt.savefig('plot21.png')
#
# Solving Ornstein-Uhlenbeck process
#
mu=10
tau=50
sigma=10

a=1/tau
b=mu/tau
c=zero
d=sigma*np.sqrt(2/tau)

rr=np.random.randn(n)

xi=zero

em=np.zeros(n)
em[0]=xi

mi=np.zeros(n)
mi[0]=xi

for i in range(n-1):
    em[i+1]=em[i]+(-a*em[i]-b)*dt+(c*em[i]+d)*np.sqrt(dt)*rr[i]
    mi[i+1]=mi[i]+(-a*mi[i]-b)*dt+(c*mi[i]+d)*np.sqrt(dt)*rr[i]
        +0.5*c*(c*mi[i]+d)*((rr[i])**2-dt)

plt.figure(dpi=200)
plt.plot(tt,em,color='r',label='EM')

```

```

plt.plot(tt,mi,color='b',label='Mi')
plt.xlabel(r'$t$')
plt.ylabel(r'$X_t$')
plt.title(r'Solution  $(X_t)_{t \geq 0}$  for  $\mu=10$ ,  $\tau=50$ ,  $\sigma=10$ ')
plt.legend()
plt.savefig('plot30.png')
#
# Moment stability of EM and Mi
#
n=20000
t=20
dt=t/n

eta=.3

em=np.zeros(n)
mi=np.zeros(n)

xi=done

em[0]=xi
mi[0]=xi

tt=np.linspace(0,n*dt,n)
rr=np.random.randn(n)

for i in range(n-1):
    em[i+1]=em[i]-em[i]*dt+(1-eta*em[i])*np.sqrt(dt)*rr[i]
    mi[i+1]=mi[i]-mi[i]*dt+(1-eta*mi[i])*np.sqrt(dt)*rr[i]
        +0.5*(-eta)*(1-eta*mi[i])*(rr[i]**2-dt)

av1=np.zeros(n)
av1=xi*np.exp(-tt)

me1=np.zeros(n)
mm1=np.zeros(n)

for i in range(1,n):
    me1[i]=xi*(1-dt)**i
    mm1[i]=xi*(1-dt)**i

plt.figure(dpi=200)
plt.semilogy(tt[1:],me1[1:],'-.',color='r',label='EM')
plt.semilogy(tt[1:],mm1[1:], '--',color='b',label='Mi')
plt.semilogy(tt[1:],av1[1:],color='gray',label='Theoretical')
plt.xlabel(r'$t$')
plt.ylabel(r'$\log(E[X_t])$')
plt.title(r'First Moment comparison for  $\eta=0.3$  and  $h=0.001$ ')
plt.legend()
plt.savefig('plot41.png')

```

```

me2=np.zeros(n)
mm2=np.zeros(n)

av2=np.zeros(n)
av2=(xi**2)*np.exp(-(2-eta**2)*tt)+(1-np.exp(-(2-eta**2)*tt))/(2-eta**2)
      +2*eta*xi*(np.exp(-tt)-np.exp(-(2-eta**2)*t))/(1-eta**2)

me2[0]=zero
mm2[0]=zero

for i in range(n-1):
    me2[i+1]=((1-dt)**2+dt*eta**2)*me2[i]+2*dt*eta*me1[i]+dt
    mm2[i+1]=(1-dt)**2*mm2[i]+dt*(1+2*eta*mm1[i]+mm2[i]*eta**2)
              +0.5*dt**2*eta**2*(1+2*eta*mm1[i]+mm2[i]*eta**2)

plt.figure(dpi=200)
plt.semilogy(tt[1:],me2[1:], '-.',color='r',label='EM')
plt.semilogy(tt[1:],mm2[1:], '--',color='b',label='Mi')
plt.semilogy(tt[1:],av2[1:],color='gray',label='Theoretical')
plt.xlabel(r'$t$')
plt.ylabel(r'$\log(E[X_t^2])$')
plt.title(r'Second Moment comparison for $\eta=0.3$ and $h=0.001$')
plt.legend()
plt.savefig('plot42.png')

```

References

- [1] M. Abramowitz and I. A. Stegun. *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, National Bureau of Standards Applied Mathematics Series, vol. 55, Washington D.C., 1964
- [2] N. Higham. *Accuracy and stability of numerical algorithms*, Other Titles in Applied Mathematics, SIAM, Philadelphia, 2002, DOI 10.1137/1.9780898718027
- [3] T. Hudson, X. H. Li, S. Murphy. *Numerical stability revisited: a family of benchmark problems for the analysis of explicit stochastic differential equations integrators*. Preprint in arXiv. <https://arxiv.org/abs/2503.19203v1>
- [4] J. F. Le Gall. *Brownian motion, martingales, and stochastic calculus*, Graduate Texts in Mathematics, vol. 274, Springer, New York, 2016, DOI 10.1007/978-3-319-31089-3
- [5] Q. Li, C. Tai, and E. Weinman (2017). Stochastic modified equations and adaptive stochastic gradient algorithms. In *International Conference on Machine Learning*, pages 2101-2110. PMLR.
- [6] G. N. Milstein. *Approximate integration of stochastic differential equations*, SIAM Theory of Probability & Its Applications **19** (1975), no. 3, 557-562, DOI 10.1137/1119062
- [7] G. A. Pavliotis. *Stochastic processes and applications: Diffusion processes, the Fokker-Planck and Langevin equations*, Texts in Applied Mathematics, vol. 60, Springer, New York, 2014, DOI 10.1007/978-1-4939-1323-7