Universiteit Utrecht

Bachelor thesis

Efficient simulation methods for stochastic differential equations

Author:
Roderick Groeneweg

Supervisor:
Dr. Deb Panja

January 18, 2017

Abstract

The purpose of this study, was to improve the efficiency of simulations of stochastic differential equations. We first explain what such equations and their properties are. Then we derive some methods to simulate them. The first couple of these methods are generalizations of numerical integration methods, which are not new. After these methods we derive new methods with a different approach. To test these methods, we generate the stationary probability distribution of the process, and compare it to the theory. We do see improvements in terms of stability and accuracy in some cases, but further research into their efficiency is needed to be conclusive.

Contents

1 Introduction

The main object of study of this thesis, is a quantity $X(t)$ that is placed in a potential $V(X(t))$, but also exhibits random fluctuations. To be more precise, the quantity $X(t)$ is governed by a continuous stochastic differential equation (SDE). What these equations are will be defined in Chapter 2. These equations arise in fields as mathematical finance, probability theory, ecology, and most notably microphysics. Causes of the random fluctuations in an otherwise deterministic system can be such as human action, underlying complex systems and in the last case collisions with myriad smaller particles.

An intuitive, but mathematically wrong way to write down a SDE is

$$\frac{dX(t)}{dt} = F(X(t)) + L(t) \tag{1}$$

Where $F(x) = -dV(x)/dx$ is the force due to the potential, and $L(t)$ is a force that causes the random fluctuations in $X(t)$. The problem with this approach is that $X(t)$ is unlikely to be differentiable.

We will not go into the nuances of existence and uniqueness problems of SDEs and simply assume that for our purposes a unique solution exists. The solution to a SDE is a stochastic process. It is however very rarely the case that this solution can be written down explicitly, and thus it can only be approximated numerically. The goal of this thesis is to discuss algorithms to simulate the process $X(t)$. Since the standard algorithms for this problem only accurately simulate the properties of the model for very small steps in time, we are especially interested in finding an algorithm that can take larger timesteps accurately.

Acknowledgements

First and foremost I would like to thank dr. Deb Panja for his guidance and for the time he invested in my thesis. He exposed me to an interesting problem, and motivated me to solve problems.
I would also like to thank my friends and family for their support.

2 The Model

In the introduction, we briefly discussed what kind of process we want to model. Now we shall give a more complete picture of what the properties of this process must be. Without losing generality we shall call the quantity $X(t)$ the location of a particle.

2.1 Brownian motion

For simplicity we first assume that the potential is constant, and hence does not exert any force on the particle. The movement from the particle are thus random fluctuations. We call this movement Brownian motion. A good mathematical definition for Brownian motion comes from the symmetric random walk (i.e. every timestep a 'particle' has equal probability going a unit step right as going a unit step left). One takes the limit to infinity of the number of steps (n) per unit time, while scaling the process with $\sqrt{n}$ so that the number of steps of the process does not influence the expected displacement per unit time. Intuitively this gives the idea of the particle experiencing many small collisions. Thankss to a central limit theorem, one can prove that the process one ends up with is normally distributed with mean at the start location and variance equal to the time that has past. A rigorous derivation and proof can be red in reference [4] Chapter 3, up to and including theorem 3.2.1.(Central limit).

If the steps of the process are not of unit length, then one can scale the location with a factor σ. In thermodynamics, by Einstein's relation (see [3] section 3.1.1), we know that $\sigma \sim k_B T$ where k_B is Boltzmann's constant, and T is the temperature. The proportionality factor depends on other physical quantities of the system (i.e. (kinetic) energy and (non-conservative) forces). We will only look at SDEs with constant diffusion coefficient $\sigma^2/2$.

If the steps are of unit length, or $\sigma = 1$, we call the process 'standard Brownian motion' or the 'Wiener process', which can be defined from its properties in the following way.

Definition 1 (Wiener Process)**.** *The Wiener process (or standard Brownian motion process), is the stochastic process* $\{B(t)|t \in [0,\infty)\}$ *that has the following properties.*

i) $B(t)$ *is continuous in* t *and* $B(0) = 0$ *almost surely*

ii) for all $0 = t_0 < t_1 < ... < t_m$ *the time increments*

$$B(t_1) - B(t_0), B(t_2) - B(t_1), ..., B(t_m) - B(t_{m-1})$$

are independent

iii) and each time increment is normally distributed with $B(t+\tau) - B(t) \sim N(0, \tau)$ *for each* t *and positive* τ*.*

It can be shown that the Wiener process is (almost surely) not differentiable to the time parameter. A rigorous proof involves a bit of measure theory which we are not going through here. A heuristic way to explain why the Wiener process is not differentiable, is that the symmetric walk changes direction in a non differentiable way. The Wiener process is the limit of infinitely many of these direction changes (up to a set of measure 0), and hence it is not surprising that there is no derivative of the paths of Brownian motion.

2.2 The Itô integral

For constant potential V, we thus have that $X(t) = \sigma B(t)$ for some constant σ depending on the specifics. But since Brownian motion is not differentiable, we cannot write $dX(t)/dt = L(t)$ as suggested in the introduction. The way to get around this problem is by defining an integral with respect to Brownian motion, and write the differential equation in integral form as

$$\int_{s=0}^{t} dX(s) = \int_{s=0}^{t} \sigma dB(s)$$

or in differential notation, which is just a quick way of writing down the integral expression for general interval of integration

$$dX(t) = \sigma dB(t) \tag{2}$$

This integral is called the Itô integral, which is very similar to the Riemann-Stieltjes integral. It is a limit in probability of the Riemann-sums, for the partition $0 < t_1 < ... < t_n = T$ going to infinitely many fragments.

$$\int_0^T g(t) dB(t) := \lim_{n \to \infty} \sum_{i=0}^{n} g(t_i)(B(t_{i+1}) - B(t_i))$$

In chapter 4 of reference [4] some important properties of Itô integrals are proved. Most notably theorems 4.3.1. and 4.4.9. We will sum them up in the following theorem.

Theorem 1 (properties of Itô integrals)**.** *Let $g(t)$ be a square integrable function, that is either nonrandom, or a stochastic process. Then the Itô integral $I(t) = \int_0^t g(t)dB(t)$ has the following properties.*

i) (Continuity) $I(t)$ is continuous in t.

ii) (Linearity) Addition and multiplication with a constant work the same as in ordenary (Riemann/Lebesgue) integrals.

iii) (Martingale) $I(t)$ is a martingale, i.e. $\mathbb{E}(I(t)|I(s)) = I(s)$ for $s \leq t$.

iv) (Itô isometry) $\mathbb{E}I^2(t) = \mathbb{E}\int_0^t g^2(u)du$.

v) (deterministic integrand) For $g(t)$ a nonrandom function of time, we have that $I(t) \sim \mathcal{N}(0, \int_0^t g^2(u)du)$, i.e. $I(t)$ is normally distributed.

In the case that the potential is not constant, there is a 'force', or *drift*, $F(x) = -dV(x)/dx$ acting on the particle. Adding this term to equation 2, we have now rigorously defined the law that governs the quantity $X(t)$

$$dX(t) = F(X(t))dt + \sigma dB(t) \tag{3}$$

integrated over any continuous interval.

Physicist often do not mind and still perform calculus with the derivative of Brownian motion $B'(t) = \eta(t)$, which has expected value $\mathbb{E}\eta(t) = 0$. The other important property of this 'derivative' is that the autocorrelation is given by $\mathbb{E}(\eta(t)\eta(s)) = \delta(t-s)$ where δ is the Dirac-δ. This approach works because $\eta(t)dt$ can be interpreted as $dB(t)$ and then the properties of $\eta(t)$ follow from the properties of $dB(t)$. By definition 1 we have

$$\begin{aligned} \mathbb{E}dB(t) &= 0 \\ \mathbb{E}(dB(s)dB(t)) &= 0 \text{ for } s \neq t \text{ by independence of increments, and} \\ \mathbb{E}(dB(t)dB(t)) &= dt = \delta(t-s)dsdt \end{aligned}$$

where the last property follows because Brownian motion has quadratic variance $[B, B](t) = t$, a result that is defined and proved in reference [4] section 3.4.2.

Note that we described a 'force' F that is proportional to the change in $X(t)$. A more specific name for this term in a stochastic differential equations is the *drift coefficient*, while the term $\sigma^2/2$ is called the *diffusion coefficient*.

2.3 The Itô-Doeblin formula

Before we can really derive some results about $X(t)$, we need some sort of chain rule for functions of Brownian motion. One would expect that for a differentiable function f we would have

$$df(B(t)) = f'(B(t))dB(t)$$

but unfortunately, because Brownian motion has quadratic variation $[B, B](t) = t$, we must take into account another term

$$df(B(t)) = f'(B(t))dB(t) + \frac{1}{2}f''(B(t))dt$$

For a rigorous definition of quadratic variation and a proof that Brownian motion has quadratic variation, see reference [4] section 3.4.2. Informally it means that, while normally expressions with second order differential fall off to zero, in stochastic calculus $dB(t)dB(t)) = dt$, while all other higher order differentials fall off to zero[1].

Theorem 2 (Itô-Doeblin formula)**.** *For $f(x,t)$ a three times differentiable function in both arguments, and the SDE $dX(t) = F(X(t))dt + \sigma dB(t)$, the differential $df(X(t),t)$ is given by*

$$df = (f_t + F(X(t))f_x + \frac{\sigma^2}{2}f_{xx})dt + \sigma f_x dB(t)$$

where the subscripts denote partial differentiation, and the arguments of f are left out for convenience.

Proof. Let $f(t,x)$ be three differentiable to both it's arguments, then we have that

$$df = f_t dt + f_x dx + \frac{1}{2}f_{xx}dx^2 + \mathcal{O}(dt^2, dxdt, dx^3) \tag{4}$$

[1]The identity $dB(t)dB(t)) = dt$ may seem contradictory, because the right side is deterministic and the left side is not. But when one integrates over these differentials, one get the exact result (almost surely) by the rule of large numbers

Now let $x = X(t)$ and $dX(t) = F(X(t))dt + \sigma dB(t)$, then (4) becomes

$$\begin{aligned} df &= f_t dt + (f_x) \cdot (F(X(t))dt + \sigma dB(t)) + \frac{1}{2}(f_{xx}) \cdot (F(X(t))dt + \sigma dB(t))^2 + \mathcal{O}(dt^2, dX(t)dt, dX(t)^3) \\ &= f_t dt + (f_x) \cdot F(X(t))dt + (f_x) \cdot \sigma dB(t) + \frac{1}{2}(f_{xx}) \cdot \sigma^2 dt + \mathcal{O}(dt^2, dX(t)dt, dX(t)^3, dB(t)dt) \end{aligned}$$

All temrs in $\mathcal{O}(dt^2, dX(t)dt, dX(t)^3, dB(t)dt)$ fall off to zero, and what we are left with is the Itô-Doeblin formula. □

3 Ornstein-Uhlenbeck processes

An Ornstein-Uhlenbeck process is a stochestic process $X(t)$ that satisfies the SDE

$$X(t) = \gamma(\mu - X(t))dt + \sigma dB(t), \qquad \text{for } \gamma > 0 \tag{5}$$

Because the drift $F(x) = \gamma(\mu - x)$ is linear, the process is simple enough so that we can write down an explicit solution. Equation 5 can be solved by Lagrange's method of variation of constants. We first solve the inhomogeneous differential equation $dX(t) = -\gamma X(t)dt$ which gives $X_{hom}(t) = e^{-\gamma t}$. The real solution can now be obtained by the substition

$$X(t) = X_{hom}(t)C(t)$$

for a function $C(t)$ that is defined by $C(x,t) = xe^{\gamma t}$ for $x = X(t)$. Normally we would take the derivative of function C, but since we are performing stochastic calculus, we have to apply the Itô-Doeblin formula, and find

$$\begin{aligned} dC(X(t),t) &= (\partial C(X(t),t)/\partial t)dt + (\partial C(X(t),t)/\partial x))dX(t) + \frac{1}{2}(\partial^2 C(X(t),t)/\partial x^2)dX(t)dX(t) \\ &= \gamma x e^{\gamma t}dt + e^{\gamma t}dX(t) + 0 \cdot dX(t)dX(t) \\ &= \gamma x e^{\gamma t}dt - \gamma x e^{\gamma t}dt + \gamma \mu e^{\gamma t}dt + e^{\gamma t}\sigma dB(t) \\ &= \gamma \mu e^{\gamma t}dt + e^{\gamma t}\sigma dB(t) \end{aligned}$$

And hence

$$C(t) = C(0) + \int_0^t \gamma \mu e^{\gamma s}ds + \int_0^t e^{\gamma s}\sigma dB(s)$$

Filling in the definition of $C(t)$, we yield

$$X(t) = X(0)e^{-\gamma t} + \mu(1 - e^{-\gamma t}) + e^{-\gamma t}\int_0^t e^{\gamma s}\sigma dB(s)$$

Which is the explicit solution of equation 5. Not that because $\partial^2 C/\partial x^2 = 0$, the Itô-Doeblin lemma has the same result as the ordinary chain rule.

Using theorem $1v$ we get that $X(t) \sim \mathcal{N}(X(0)e^{-\gamma t} + \mu(1 - e^{-\gamma t}), \frac{\sigma^2}{2\gamma}(1 - e^{-2\gamma t}))$

Note that when we take $-\gamma > 0$, although the process is not called an Ornstein-Uhlenbeck process, we do obtain the same explicit solution.

4 Some theorems we have used

In this section, we derive some known results that we have used to tackle the proposed problem. We derive the Fokker-Planck equation which is a partial differential equation in the probability density function, that gives an equivalent description of the process $X(t)$. We use this equation to get an expression for the stationary state of $X(t)$, and we elaborate on time reversibility and the detailed balance.

4.1 The Fokker-Planck equation

Another way to describe the process $X(t)$, is by a partial differential equation (PDE) in its probability density function of $X(t)$. There are more then one such PDEs, but the one that is the most use to us is the Fokker-Planck equation (or more generally the Kolmogorov forward equation [2]). We will now define this equation and prove that it is correct. The proof is based on multiple applications of the Itô-Doeblin formula (theorem 2).

Theorem 3 (Fokker-Planck equation)**.** *For the general STE*

$$dX(t) = F(t, X(t))dt + \sigma(t, X(t))dB(t)$$

the probability density function $p = \mathbb{P}(X(t) = y)$ *satisfies the Fokker-Planck equation*

$$\frac{\partial}{\partial t}p = -\frac{\partial}{\partial y}(F(t,y)p) + \frac{1}{2}\frac{\partial^2}{\partial y^2}(\sigma(t,y)^2 p)$$

Proof. For $b > 0$ let $h_b : \mathbb{R} \to \mathbb{R}$ be a twice differentiable function with support[3] $\text{supp}h_b = (0, b)$. We use this function as a help function to prove our theorem. By the Itô-Doeblin formula:

$$\begin{aligned} dh_b(X(u)) &= h_b'(X(u))dX(u) + \frac{1}{2}h_b''(X(u))dX(u)dX(u) \\ &= h_b'(X(u))\left(F(u, X(u))du + \sigma(u, X(u))dB(u)\right) + \frac{1}{2}h_b''(X(u))\sigma(u, X(u))^2 du \\ &= \left[h_b'(X(u))F(u, X(u)) + \frac{1}{2}h_b''(X(u))\sigma(u, X(u))^2\right]du + h_b'(X(u))\sigma(u, X(u))dB(u) \end{aligned}$$

[2] As the name implies there is also a Kolmogorov backward equation.

[3] The support of a function g is the subset of its domain where g is unequal to zero. $\text{supp}g = \{x \in \text{Dom}g | g(x) \neq 0\}$

Integrating from 0 to t gives

$$h_b(X(t))-h_b(0) = \int_0^t \left[h_b'(X(u))F(u,X(u)) + \frac{1}{2}h_b''(X(u))\sigma(u,X(u))^2 \right] du + M$$

where M is the integral w.r.t. Brownian motion, and thus by theorem $1iii$ has expected value 0. Since we are now going to take the expected value of both sides, this term disappears. Moreover since $h_b = 0$ outside $(0,b)$, we can integrate over $(0,b)$ in stead of $(-\infty,\infty)$.

$$\begin{aligned}\mathbb{E}\left[h_b(X(t)) - h_b(0)\right] &= \int_0^b h_b(y)\mathbb{P}\{X(t)=y\}dy - h_b(0) \\ &= \mathbb{E}\int_0^t \left[h_b'(X(u))F(u,X(u)) + \frac{1}{2}h_b''(X(u))\sigma(u,X(u))^2 \right] du\end{aligned}$$

thus, after changing the order of integration

$$\mathbb{E}\left[h_b(X(t)) - h_b(0)\right] = \int_0^t \int_0^b \left[h_b'(y)F(u,y) + \frac{1}{2}h_b''(y)\sigma(u,y)^2 \right] \mathbb{P}\{X(t)=y\}dydu \tag{6}$$

Using integration by parts on both parts of integrals from 0 to b, where the boundary terms disappear since $h_b(0) = h_b(b) = 0$, we get

$$\begin{aligned}\int_0^b h_b'(y)F(u,y)\mathbb{P}\{X(t)=y\}dy &= -\int_0^b h_b(y)\frac{\partial}{\partial y}(F(u,y)\mathbb{P}\{X(t)=y\})dy \\ \int_0^b \frac{1}{2}h_b''(y)\sigma(u,y)^2\mathbb{P}\{X(t)=y\}dy &= -\int_0^b \frac{1}{2}h_b'(y)\frac{\partial}{\partial y}(\sigma(u,y)^2\mathbb{P}\{X(t)=y\})dy \\ &= \int_0^b \frac{1}{2}h_b(y)\frac{\partial^2}{\partial y^2}(\sigma(u,y)^2\mathbb{P}\{X(t)=y\})dy\end{aligned}$$

Substitution these results back into (6) we get

$$\begin{aligned}&\int_0^b h_b(y)\mathbb{P}\{X(t)=y\}dy = h_b(0) + \\ &\quad \int_0^t \int_0^b \left[-\frac{\partial}{\partial y}(F(u,y)\mathbb{P}\{X(t)=y\}) + \frac{1}{2}\frac{\partial^2}{\partial y^2}(\sigma(u,y)^2\mathbb{P}\{X(t)=y\}) \right] h_b(y)dydu\end{aligned}$$

Taking the derivative w.r.t t gives

$$\int_0^b h_b(y)\frac{\partial}{\partial t}pdy = \int_0^b \left[-\frac{\partial}{\partial y}(F(t,y)p) + \frac{1}{2}\frac{\partial^2}{\partial y^2}(\sigma(t,y)^2 p) \right] h_b(y)dy$$

And since this has to be true for every h_b with given properties, the only way this equation can hold is if

$$\frac{\partial}{\partial t}p = -\frac{\partial}{\partial y}(F(t,y)p) + \frac{1}{2}\frac{\partial^2}{\partial y^2}(\sigma(t,y)^2 p)$$

Which concludes the proof. □

In the case of a linear $F(y)$, the process $X(t)$ is an Ornstein-Uhlenbeck process. In this case a Fourier transform can be used to simplify the partial differential equation, so that an explicit solution can be written down. Of course, we already know that this solution is going to be a normal curve.

Solving the Fokker-Planck equation is equivalent to solving our SDE. During this study, we did not manage to get new insights from this approach, but for more information one can consult reference [3].

Definition 2 (Probability current)**.** *The Fokker-Planck equation can also be written in the form*

$$\frac{\partial}{\partial t}p = -\frac{\partial}{\partial y}S(x,y,t)$$

where $S(x,y,t) = f(t,y)p + (1/2)(\partial/\partial y)(\sigma(t,y)^2 p)$ *is called the probability current.*

4.2 The stationary distribution

Observe that the probability current (definition 2) can be written

$$S(x,y,t) = \frac{\sigma^2}{2}e^{-\frac{2}{\sigma^2}V(y)}\frac{\partial}{\partial y}\left(e^{\frac{2}{\sigma^2}V(y)}p(x \to y,t)\right) \tag{7}$$

Where $V'(y) = -F(y)$, and $p(x \to y,t) = \mathbb{P}(X(t) = y | X(0) = x)$. If p is a stationary solution, i.e. $\partial p/\partial t = 0$, then it follows from the Fokker-Planck equation that $S(x,y,t) = S(y) = S$ is constant [4]. A particular solution to the differential equation 7 for constant S, is

$$p_p(y) = S \cdot \frac{2}{\sigma^2}e^{-\frac{2}{\sigma^2}V(y)}\int e^{\frac{2}{\sigma^2}V(y)}dy$$

[4]It can be shown that a stationary distribution exists, as long as all states commute with each other, and the mean return time of any state is finite. This result comes from the theory on continuous-time Markov chains.

The homogeneous solution (i.e.solution for $S = 0$) can be found by direct integration:

$$p_h(y) = Ne^{-2V(y)/\sigma^2}$$

where N is the normalisation. The general solution is the sum of the homogeneous and particular solution. Since this is a probability density function, if we integrate it over the whole space it yields 1, which is the same for the homogeneous solution. Therefore, the particular solution should integrate to 0, but since the integral and exponential are both positive, we must have that $S = 0$. Hence, the general solution is equal to the homogeneous solution of 7. We have now proved the following theorem.

Theorem 4. *For the SDE*

$$dX(t) = F(X(t))dt + \sigma dB(t)$$

and potential $V(x)$, the probability density function $\pi(y)$ given by

$$\pi(y) = \frac{e^{-2V(y)/\sigma^2}}{\int_{-\infty}^{\infty} e^{-2V(\xi)/\sigma^2} d\xi}$$

is the stationary distribution of $X(t)$, provided that the integral converge.

4.3 Time-reversibility

Definition 3. *A stochastic process $X(t)$ is said to be time-reversible, if the following joint probabilities are equal.*

$$\mathbb{P}(X(t) = y, X(0) = x) = \mathbb{P}(X(t) = x, X(0) = y)$$

For all x, y in the state space of X, and all t.

From definition 3 it follows that a time-reversible stochastic process $X(t)$ with stationary probability distribution π, adheres to the following identity.

$$\mathbb{P}(X(t) = y|X(0) = x)\pi(x) = \mathbb{P}(X(t) = x|X(0) = y)\pi(y)$$

The following theorem is due to Kolmogorov. A proof can be found in reference [5].

Theorem 5. *Let f be Lipschitz continuous, and let $X(t)$ be the solution to the SDE*

$$dX(t) = f(X(t))dt + \sigma dB(t)$$

Then $X(t)$ is reversible and has stationary distribution π, iff there is a function V such that $f = -\nabla V$ with $d\pi = e^{-2V(x)}dx$, and

$$\int_{-\infty}^{\infty} e^{-2V(x)}dx < \infty$$

From theorem 5 it follows that any process with a polynomial potential with positive highest order coefficient, is time-reversible. By theorem 4, the stationary probabilities follow a distribution $\pi(x) \sim e^{-2V(x)/\sigma^2}$. Combining this with the reversability we find

$$\frac{\mathbb{P}(X(t) = y|X(0) = x)}{\mathbb{P}(X(t) = x|X(0) = y)} = \frac{\pi(y)}{\pi(x)} = e^{-2(V(y)-V(x))/\sigma^2}$$

We call this identity the *detailed balance*.

5 Simulation methods based on integration

In this Chapter, we take a look into the most natural ways of simulating $X(t)$. The methods we are discussing are generalizations of the methods normally used for ordinary differential equations (ODE). All first order ODEs have the form $dy/dx = f(y, x)$, and schemes for numerical approximations of its solution are derived by approximations of dy/dx and higher order derivatives $d^2y/dx^2 = df/dx = \partial f/\partial x + \partial f/\partial y \cdot dy/dx = \partial f/\partial x + f \cdot \partial f/\partial y$, and so on. Unfortunately for us, these derivations cannot be easily generalized for a SDE, since in our case

$$dX(t) = F(X(t))dt + \sigma dB(t)$$

the process $X(t)$ is almost surely not differentiable, and the chain rule does not hold (but we do have the Itô-Doeblin formula, theorem 2). Indeed, our process is defined only in integral form

$$X(t) = X(0) + \int_0^t F(X(s))ds + \sigma \int_0^t dB(s)$$

where $X(0)$ is nonrandom, and we know that $\int_0^t dB(s) = B(t) \sim N(0, t)$. This knowledge reduces the problem of simulating down to finding a good method to approximate the Lebesgue integral $\int_0^t F(X(s))ds$. Luckily, methods for solving ODEs often have an analogous method for computing integrals. Here we will use the known order of convergence for the midpoint method, and for Simpson's rule. These methods are covered in for instance reference [1] Chapter 15.

Another useful result comes from the Itô-Doeblin formula (theorem 2), we get

$$F(X(t)) = F(X(0)) + \int_0^t \left(F'(X(s))F(X(s)) + \frac{\sigma^2}{2} F''(X(s)) \right) ds + \int_0^t F'(X(s))\sigma dB(s)$$

5.1 Simulating the path of one particle

Before discussing different methods to simulate $X(t)$ on step further in time, we shall go through some general information about how we carry out these simulation. First of all, for convenience we often specify $t = 0$ to be the starting point. For simulations of multiple steps, the method is of course

repetitively applied, where the starting point t_s is the end point of the previous timestep.

Secondly, to simulate the Wiener process $\int dB(t) \sim \mathcal{N}(0,t)$ we use a pseudo-random number generator. The specific method to generate a standard Gaussian random number is explained in reference [2] chapter 7, (sections 7.1 and 7.2).

And lastly, the SDE

$$dX(t) = -kF(X(t))dt + \sigma dB(t)$$

is equivalent to

$$dY(t) = -F(Y(t))dt + \frac{\sigma}{\sqrt{k}}dB(t)$$

by the transformation $Y(t) \equiv X(t/k)$; meaning that these random variables have the same distribution.
Let $k = \sigma^2$ and we have a process of unit diffusion. This will be the standard in our simulations, so that different methods will be compared with each other for the same diffusion coefficient.

5.2 The Euler Forward method

The simplest way to approximate the integral is

$$\int_0^{\Delta t} F(X(t))dt \approx F(X(0)) \cdot \Delta t$$

By the Itô-Doeblin result above, the error is given by

$$\int_0^{\Delta t} \left[\int_0^t \left(F'(X(s))F(X(s)) + \frac{\sigma^2}{2}F''(X(s)) \right) ds + \int_0^t F'(X(s))\sigma dB(s) \right] dt$$

Because $X(s)$ is continuous, the mean value theorem holds for the left side of the expression, and therefore there is $c \in [0,t]$ such that the error is

$$\int_0^{\Delta t} \left[\left(F'(X(c))F(X(c)) + \frac{\sigma^2}{2}F''(X(c)) \right) \cdot t + \int_0^t F'(X(s))\sigma dB(s) \right] dt$$

It can now be seen that the left integral converges with at least order $\mathcal{O}(\Delta t^2)$. Keep in mind that c is also dependent on t, but since it is always true that

$c \in [0, \Delta t]$ this dependence can be eliminated when determining the order by taking a supremum over $[0, \Delta t]$. The Itô integral on the right is less straight forward to determine. From theorem $1iii$ (martingale property) we see that the expected value is 0, and from theorem $1iv$ (Itô isometry) we deduce that the variance is $\mathbb{E}\int_0^t \sigma^2 F'(X(s))^2 ds \approx \sigma^2 t F(X(0))^2$.

We have now shown that the Euler forward method converges with order at least 2, and also the variance of the error goes to zero.

5.3 The Midpoint method

From numerical integration, we know that the midpoint method is superior to the Euler method. The error term in the midpoint method is over order $\mathcal{O}(\Delta t^3)$ for ODEs/deterministic integrals. A precise estimate of the error ($f''(\xi)\Delta t^3/24$ for ξ between 0 and Δt) cannot be given in our case, since the derivative to t has no meaning. To ad further complication, in the case of our integral of interest, the midpoint method uses the value of $X(\Delta t/2)$, which we cannot generally compute precisely. We can however simulate it with the Euler method discussed above. Here we find

$$\begin{aligned}
\int_0^{\Delta t} F(X(t)dt &= \Delta t \cdot F(X(\Delta t/2)) + \mathcal{O}(\Delta t^3) \\
&= \Delta t \cdot F\left(X(0) + \int_0^{\Delta t/2} F(X(t))dt + \sigma \int_0^{\Delta t/2} dB(t)\right) + \mathcal{O}(\Delta t^3) \\
&= \Delta t \cdot F\left(X(0) + F(X(0)) \cdot \frac{\Delta t}{2} + \mathcal{O}(\Delta t^2) + \sigma \int_0^{\Delta t/2} dB(t)\right) + \mathcal{O}(\Delta t^3) \\
&= \Delta t \cdot F\left(X(0) + F(X(0)) \cdot \frac{\Delta t}{2} + \sigma \int_0^{\Delta t/2} dB(t)\right) + \mathcal{O}(\Delta t^3)
\end{aligned}$$

A method of order at least 3.

5.4 Simpson's rule/Runge Kutta

Much like the way we generalized the midpoint method for SDEs, we can use Simpson's rule to find a method that in deterministic ODEs would be

the 4th order Runge-Kutta method.

$$\begin{aligned}\int_0^{\Delta t} F(X(t)dt &= \frac{\Delta t}{6}\left[F(X(0)) + 4F(X(\Delta t/2)) + F(X(\Delta t))\right] + \mathcal{O}(\Delta t^5) \\ &= \frac{\Delta t}{6}[F(X(0)) + 4F\left(X(0) + \frac{\Delta t^2}{2}F(X(0)) + \mathcal{O}(\Delta t^2) + \int_0^{\Delta t/2}\sigma dB(t)\right) \\ &\quad + F\left(X(0) + \Delta t F(X(0)) + \mathcal{O}(\Delta t) + \int_0^{\Delta t}\sigma dB(t)\right)] + \mathcal{O}(\Delta t^5) \\ &= \frac{\Delta t}{6}[F(X(0)) + 4F\left(X(0) + \frac{\Delta t}{2}F(X(0)) + \int_0^{\Delta t/2}\sigma dB(t)\right) \\ &\quad + F\left(X(0) + \Delta t F(X(0)) + \int_0^{\Delta t}\sigma dB(t)\right)] + \mathcal{O}(\Delta t^3)\end{aligned}$$

A method of order is still at least 3, because unlike a deterministic integrand, we need still use Euler's method to simulate parts of the process $X(t)$. Intuitively we expect that Simpson's rule gives a better approximation per step size. The set back is that it is much more computationally expensive.

5.5 Simulations

To test the proposed methods to simulate $X(t)$, we first simulate the stationary state for a quadratic potential. This is done by simulation 1 million particles and collecting their locations after they reach a stationary distribution. The results are plotted in figures 1, 2 and 3. We see significant improvement from the Euler method to the midpoint method, and slighter improvement in the Simpson's rule method.

From the theory, we know that $\int F(X(t))dt$ is well approximated by all methods, when $F(x)$ is relatively constant. For high values of x this is not the case, and we will therefore test the proposed methods for a smaller time step ($\Delta t = 0.4$), starting at $X(0) = 1$. The generated distributions are plotted in figure 4.

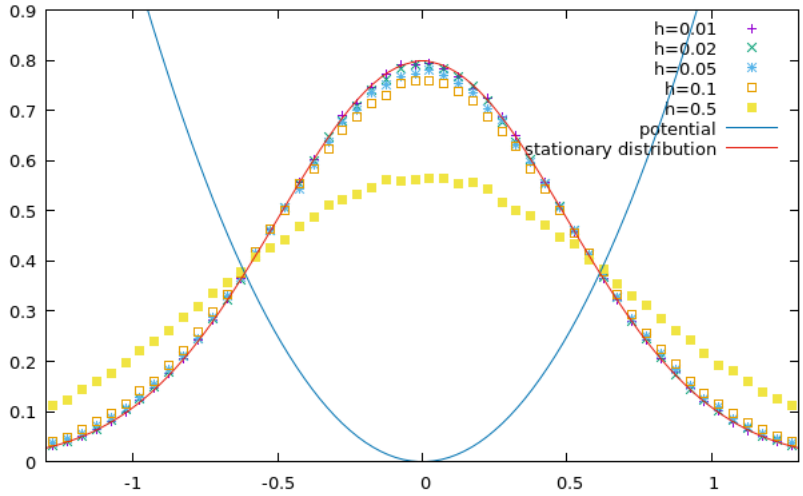

Figure 1: Using the Euler forward method for different stepsizes h, we simulated 1 million particles with $F(x) = -2x$ and $\sigma = 1$ until a stationary distribution is reached. Then a probability distribution was generated form their locations and plotted against the theoretical distribution. We see that for stepsizes $h > 0.1$ the Euler forward method is not very reliable, but for smaller stepsizes we get an accurate stationary probability distribution. The results suggest quadratic convergence. We would like to see other methods do better for larger timesteps.

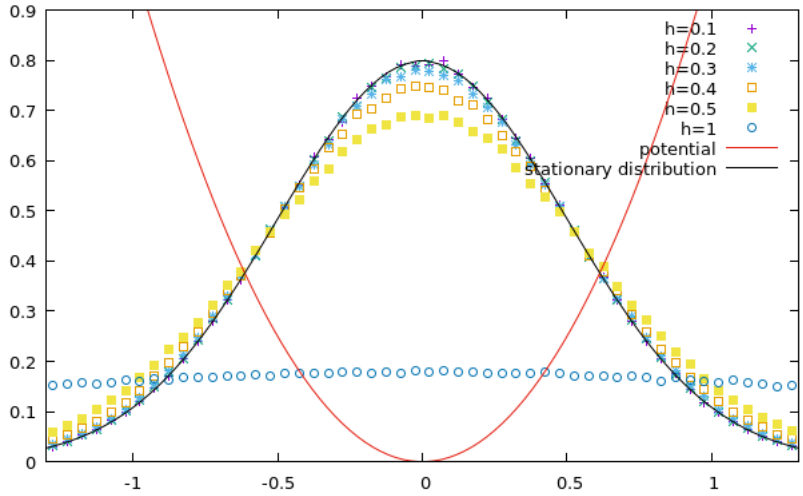

Figure 2: In this figure, we generated the stationary probability density function for $X(t)$ for the same specification as in figure 1, only this time we used the midpoint method (section 5.3). Compared to the Euler forward method, we see substantial improvement, with the stationary distribution becoming quite accurate for $h < 0.3$ which looks like an improvement with a factor 3.

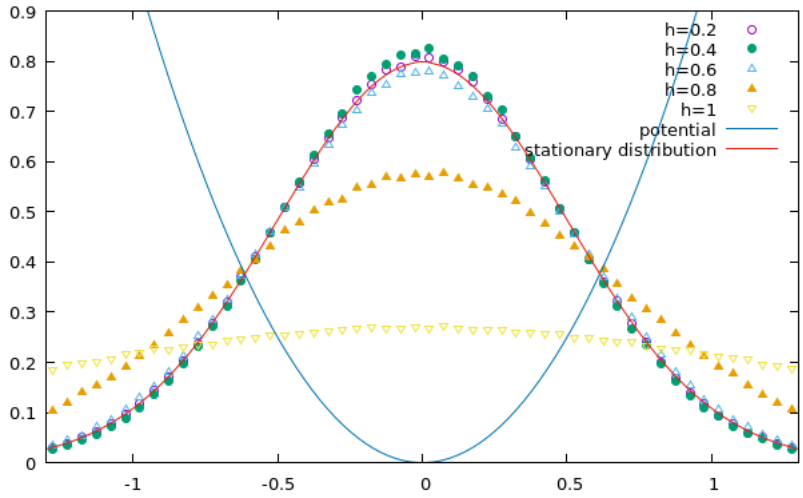

Figure 3: In this figure, we generated the stationary probability density function for $X(t)$ for the same specification as in figure 1 (and 2), only this time we used the Simpson's rule method (section 5.4). For $h = 0.6$ the Simpson's rule method does quite a bit better than the midpoint method, but for higher h it deteriorates quickly. Although this method is slower that the midpoint method, we do expect that in most cases it is a (slight) improvement.

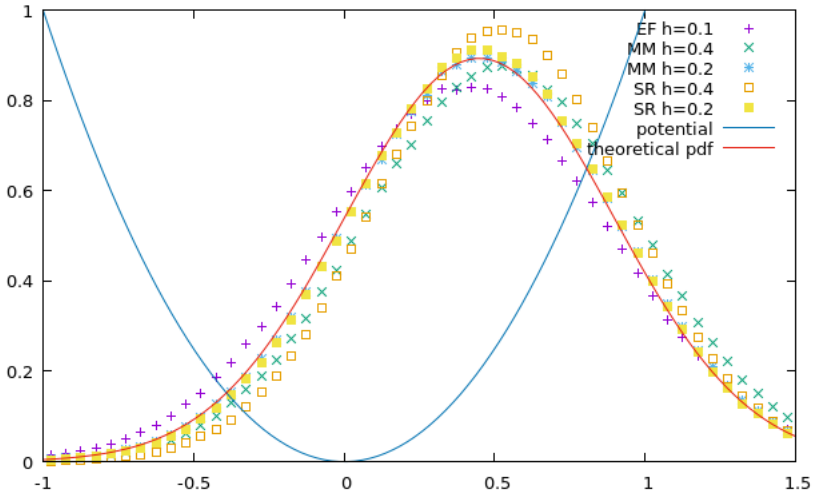

Figure 4: In this figure, we generated the probability density function for $X(0.4)$ for $X(0) = 1$, $F(x) = -2x$ and $\sigma = 1$ using different methods at stepsizes where they accurately generated the stationary distribution in earlier simulation. We compare these generated distribution with the theoretical distribution, and see that all methods are less accurate than in earlier simulations. We expect that this is due to the high value of F in this simulation, compared to the earlier ones. Especially the Euler forward method performs poorly, which can be explained by the big difference in the used (constant) drift, and the real linear drift. The midpoint method does very well, and the Simpson's rule method fail to perform better than the midpoint method. We argue that this is the case because the drift is linear, and for more complicated drift coefficients the Simpson's rule method would overtake the midpoint method in performance with relatively large stepsizes.

6 Simulation methods based on approximating the drift

If we look into the literature about methods to simulate SDEs, we come across methods described in section 5. For a quadratic potential V, or equivalently said a linear drift F, these methods have no use since we know the explicit solution $X(t)$ of such a SDE and can easily simulate it directly. In this section we elaborate on the idea to do the same thing for a more general potential V (or drift F). In this case we do not know $X(t)$ explicitly, but if we approximate V locally, we can use the explicit solution nonetheless. We might yield a simulation method that is more efficient than those proposed thus far. To our knowledge, such methods have not been used before, and so there are few theorems that can be used and still many ways to fine-tune and test the proposed algorithms. We are interested in methods to get an idea of the error of this approach, and in ways we can get the error as low as possible without compromising too much efficiency.

6.1 Taylor expanding the potential

While we tried to think of ways to get better algorithms to simulate $X(t)$, we came upon the idea to Taylor expand the potential to second order around $X(0) = x$.

$$V(y) = V(x) + \epsilon V'(x) + \epsilon^2 \frac{V''(x)}{2} + \epsilon^3 \frac{V'''(\xi)}{6}$$

where $\epsilon = y - x$ and ξ lies between x and y. Therefore

$$F(y) = -V'(y) = -\partial_\epsilon V(x + \epsilon) = -V'(x) - \epsilon V''(x) - \epsilon^2 \frac{V'''(\xi)}{2}$$

If we leave out the last term, we obtain a linear drift, for which we know that the solution of the SDE is given by the Ornstein-Uhlenbeck process. For a path $X(t)$ that stays close to $X(0)$, this Ornstein-Uhlenbeck process is going to be a good approximation. The solution of this approximation is given by

$$X_{OU}(t) \sim \mathcal{N}(\mu(X(0), t), \Sigma(X(0), t)^2)$$

with mean

$$\mu(x, t) = x e^{-V''(x)t} + \left(\frac{V'(x)}{V''(x)} - x \right) \left(e^{-V''(x)t} - 1 \right)$$

and variance

$$\Sigma(x,t)^2 = \frac{\sigma^2}{2V''(x)}\left(1 - e^{-2V''(x)t}\right)$$

We have tested this method in the double-well potential $V(x) = x^4/4 - x^2$. The generated stationary distribution for different stepsizes is plotted in figure 5.

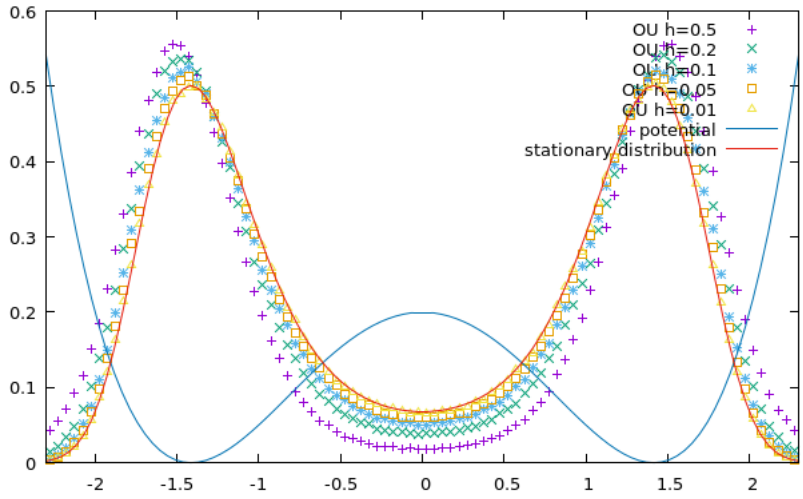

Figure 5: In this figure, we generated the stationary distribution for $X(t)$ in a double well potential $V(x) = x^4/4 - x^2$. We approximated $X(t)$ with an Ornstein-Uhlenbeck process as described in section 6.1, for different stepsizes h. As stepsizes become smaller, the smaller the neighbourhood in which $X(t)$ becomes, and thus the better the approximation of the potential is. We see that for this method to be accurate, we have to use smaller stepsizes ($h < 0.2$).

6.2 Controling the error with the detailed balance

In section 4.3 we discussed the detailed balance

$$p(x \to y, t) = e^{-2(V(y)-V(x))/\sigma^2} p(y \to x, t)$$

where $p(x \to y, t) = \mathbb{P}(X(t) = y | X(0) = x)$ for all x, y. Now, we are going to use this identity.

let $p_{OU}(x \to y, t)$ be the probability density function of $X_{OU}(t)$, the approximation of $X(t)$ that we described in section 6.1, for given $X(0) = x$. We define the relative error in the detailed balance e_{db}, by

$$p_{OU}(x \to y, t) = e^{-2(V(y)-V(x))/\sigma^2} p_{OU}(y \to x, t) + e_{db} \cdot p_{OU}(x \to y, t)$$

If $|e_{db}| \approx 1$, then our approximation fails to satisfy the detailed balance by an error of the order $p_{OU}(x \to y, t)$. We have to be careful however: in the above identity we use two terms that are approximations; $p_{OU}(x \to y, t)$ and $p_{OU}(y \to x, t)$. It could very well be the case that the error in the first cancels out to the error in the latter, so that e_{db} will be very small, in this case we could mistakenly think that $p_{OU}(x \to y, t)$ is a good approximation. It could also be the case that the error in the first is small, but the error in the latter is big, so that e_{db} is large, which is misleading as well. That being said, if the potential between the points x and y is well approximated by an expansion around $X(0) = x$, then it seems intuitive that the error $|e_{db}|$ should be small. If the expansion is a bad approximation for the potential, then the Taylor series should look quite different around $X(0) = y$, and thus $|e_{db}|$ should be large.

A way to use the detailed balance in an algorithm to simulate $X(t)$, is by rejecting the outcome for $X(t) = y$ if e_{db} is too large. But if we do this, without a proper way of correcting for the rejected values, we get probability distributions that are too high in the acceptance region, and 0 in the rejection region, which is not accurate. In stead, for a rejected step we now take two steps with half the timestep, so that the potential is approximated more precise. We also use the same random number again, to keep the integral $\int \sigma dB(t)$ normally distributed.

Using this method, the stationary distribution for the double-well potential has been generated in figure 6. In our simulations with this method, we have not observed better results compared to the Ornstein-Uhlenbeck approximation of section 6.1, while is should have been an improvement. Further research into this approach is needed to be more conclusive.

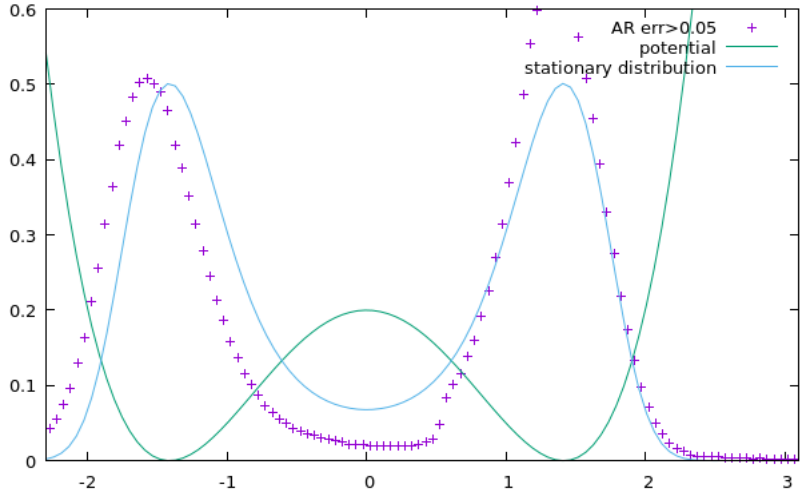

Figure 6: In this figure, we generated the stationary distribution for $X(t)$ in a double well potential $V(x) = x^4/4 - x^2$. We approximated $X(t)$ with the acceptance rejection method from section 6.2. Here we see the method for stepsize $h = 0.05$ and maximal error $e_{db} = 0.05$. Unfortunately, we have not found any values for which they yield a good approximation of the process. Moreover, we see symmetry breaking which we cannot explain. In figure 7 we symmetrized the data.

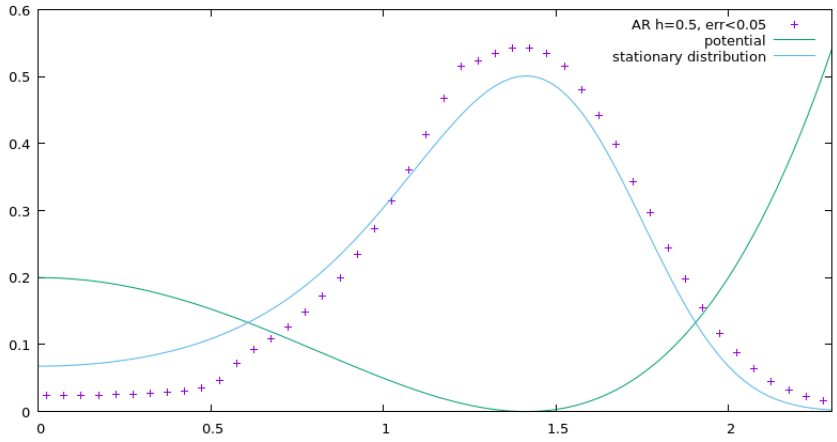

Figure 7: In this figure, we see the symmetrized data from figure 6. The Acceptance rejection method seems not to give significant progress. However we cannot be conclusive, and further research might pay off.

6.3 Using the Fokker-Planck equation

A big part of this study has been devoted to using the Fokker-Planck equation (3) to get a grip of the probability distribution compared to the distribution we obtain from approximating the process with an Ornstein-Uhlenbeck process. Unfortunately this has not provided significant steps forward.

If we substitute the approximate potential from section 6.1 into the Fokker-Planck equation, we get

$$\partial_t p(x \to y, t) = \partial_y \left((V'(x) + \epsilon V''(x) + \epsilon^2 \frac{V'''(\xi(y,t))}{2}) p(x \to y, t) \right) + \frac{\sigma^2}{2} \partial_y^2 p(x \to y, t)$$

With $\epsilon = y - x$ and $\xi(y,t)$ between x and y. The probability density function $p(x \to y, t)$ is given by $p(x \to y, t) = \int_0^t \partial_s p(x \to y, s) ds$, where the integrand is given by the Fokker-Planck equation above. The truncation error in $p(x \to y, t)$ is given by

$$e_{\text{trunc}} = \int_0^t \partial_y \left(\frac{V'''(\xi(y,s))}{2} \epsilon^2 p(x \to y, s) \right) ds$$

for $x \neq y$. This expression for the error term is unfortunately very complex, and we did not manage to get any results from it.

6.4 Another Orstein-Uhlenbeck approximation

The algorithm suggested in section 6.1 uses a Taylor expansion of the potential at $X(0)$. The farther the particle travels from the location $X(0)$, the less accurate the algorithm becomes. In this section we propose a different approximation to the potential, that still makes us of the explicit solution of our SDE when the drift F is linear. The idea is to simulate $X(t) = y$ first with an efficient, but not so accurate method (Euler forward or the midpoint method with a large stepsize), and then approximating the drift term with a line through $X(0) = x_0$ and y. This line is given by

$$L(x) = ax + b, \qquad a = \frac{F(y) - F(x_0)}{y - x_0}, \qquad b = F(x_0) - ax_0$$

With $F(x) \approx L(x)$ the drift term F is well approximated in the neighbourhoods of both x_0 and y. Our y is however not an accurate approximation of $X(t)$, and we find a better approximation by using the same value for $\int_0^t \sigma dB(s) =: B \sim \mathcal{N}(0, \sigma^2 t)$, but making use of the explicit solution

$$X(t) \sim \mathcal{N}(X(0)e^{-at}, \frac{\sigma^2}{2a}(e^{2at-1}))$$

thus we get

$$X(t) = \sqrt{\frac{\sigma^2}{2a}(e^{2at-1})}B + X(0)e^{-at}$$

as our approximation. Although this method is computationally more expensive, it might pay off. In figure 8 the stationary probability density function has been generated for different timesteps, using this method. We see an accurate generation of the stationary distribution for $h = 0.4$!

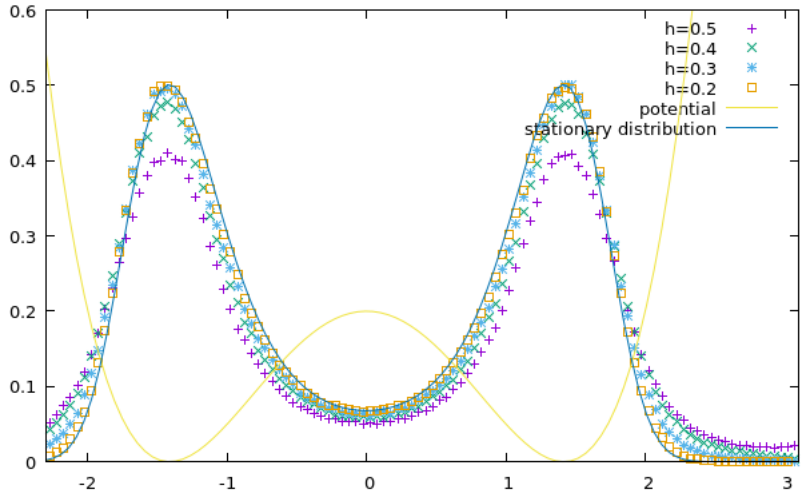

Figure 8: In this figure, we generated the stationary distribution for $X(t)$ in a double well potential with the method described in section 6.4. We used the midpoint method for the approximation of the drift by a linear function, and then use the explicit solution to generate the actual point $X(t) = y$. This method performs better than the Taylor expansion of the potential/drift.

6.5 Comparison with section 5 methods

In this section we compare the methods of section 5 with the methods of this section. We shall call the method of section 6.1 where we Taylor expand the potential and then simulate $X(t)$ as an Ornstein-Uhlenbeck process the OU-method, and we shall call the method of section 6.4, the MMOU-method, as it is a combination of the midpoint method and the OU-method.

When we used the Simpson's rule method and the midpoint method in the double well potential $V(x) = x^4/4 - x^2$, we found that they become unstable very quickly when we used larges stepsizes. For $h = 0.2$ the generated stationary distribution was very accurate, but for $h > 0.2$ both methods would sometimes blow up. The logic explanation is that the potential becomes steeper much more quickly than a quadratic potentialr, which is where $\int F(X(t))dt$ is harder to approximate. The MMOU-method shows the same blow-ups at stepsizes larger than $h > 0.8$, and interestingly the OU-method begins to blow up for $h > 4$, which is very high. A possible explanation for this, is the fact that the quadratic potential does go up again, whereas for the Euler forward and the midpoint method, we approximate the integral $\int F(X(t))dt$ with a linear one; for the Simpson's rule method this is more subtle, and therefore it is unlikely that this is a complete explanation.

In figure 9 we see the generated stationary distributions for the methods in this section for high stepsizes where they do not yet blow up. In the following table, we summarize the observed behaviour of the different methods in the described double well potential for different stepsizes, in simulations of the stationary state.

h	MM	SR	OU	MMOU
accurate	< 0.2	< 0.2	$< 0.1 \sim 0.2$	< 0.4
poor	-	-	$0.2 \sim 4$	$0.5 \sim 0.8$
explode	> 0.2	> 0.2	> 4	> 0.8

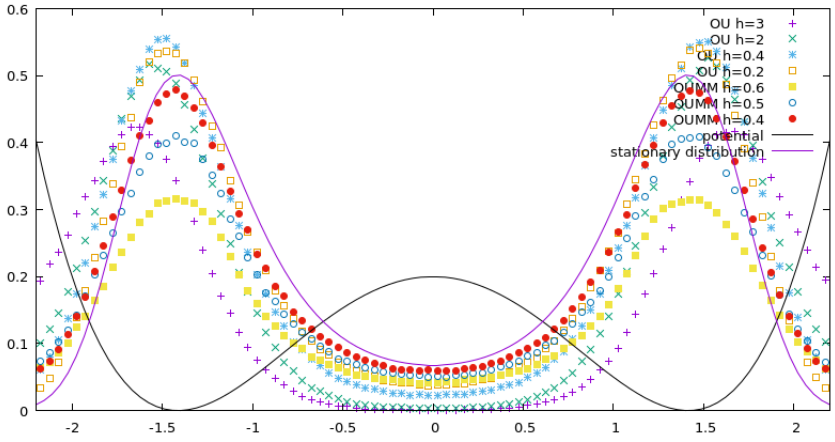

Figure 9: In this figure, we generated the stationary distribution for $X(t)$ in a double well potential with the methods described in this section, for large stepsizes where they do not yet explode. We discovered that the methods from section 5 blow up for relatively small stepsizes in the double well potential. The most interesting finding, is that the OU-method doesn blow up until stepsizes were highered to $h = 4$. Nevertheless the generated stationary distributions for high stepsizes far from accurate.

7 Conclusions

The problem of simulating SDEs

$$dX(t) = F(X(t))dt + \sigma dB(t)$$

for nonlinear drift terms $F(X(t))$, remains a very difficult problem. We will now concisely summerize our findings.

We have found that methods based on approximating the Lebesgue integral $\int F(X(t))dt$, section 5, are a good way to go, although for accurate simulations, they can be slow. The newly conceived methods of section 6 have the potential to provide a useful alternative, although more tests are needed to be conclusive. In tests where we generated the stationary distribution of the double-well potential, we found that the Ornstein-Uhlenbeck solution of a Taylor expanded potential provides an algorithm that is much more stable than other known methods, although for accurate approximations we still need to use relatively small stepsizes. Using the detailed balance and acceptance-rejection sampling, section 6.2, we did not manage to improve this method, although the approach could be improved by further research. In section 6.4, we found that combining the two approaches yields a method that is accurate for approximately twice as large stepsizes; but we have not yet answered the question as to how efficient this method is compared to other methods.

References

[1] Uri M. Asher, Chen Greif, *A First Course in Numerical Methods*, Society for Industrial and Applied Mathematics, 2011.

[2] William H. Press, Saul A. Teukolsky, William T. Vetterling, Brian P. Flannery, *Numerical Recipes in C; the Art of Scientific Computing*, 2nd edition, Cambridge University Press, 1992.

[3] Hannes Risken, *The Fokker-Planck Equations; Methods of Solution and Applications*, Springer-Verlag, Berlin Heidelberg, 1984.

[4] Steven E. Shreve, *Stochastic Calculus for Finance II:, Continuous-Time Models*, Springer Finance, New York, 2004.

[5] Jochen Voß, *Some large deviation results for diffusion processes*, Universität Kaiserlautern, PhD thesis, 2004.