M5A44 COMPUTATIONAL STOCHASTIC PROCESSES

Professor G.A. Pavliotis

Department of Mathematics Imperial College London, UK
g.pavliotis@imperial.ac.uk

WINTER TERM 2014-2015
14/01/2015
Lectures: Mondays, 11:00-13:00, Wednesdays 12:00-13:00, Huxley 658.

Office Hours: Mondays 14:00-15:00, Wednesdays 14:00-15:00 or by appointment. Huxley 621.

Course webpage:
http://www.ma.imperial.ac.uk/~pavl/comp_stoch_proc.htm

Text: Lecture notes, available from the course webpage. Also, recommended reading from various textbooks/review articles.
This is an introductory course on computational stochastic processes, aimed towards 4th year, MSc and MRes students in applied mathematics and theoretical physics.

Prior knowledge of basic stochastic processes in continuous time, scientific computing and a programming language such as matlab or C is assumed.
PART I: INTRODUCTION

- Random number generators.
- Random variables, probability distribution functions.
- Introduction to Monte Carlo techniques.
- Variance reduction techniques.

SIMULATION OF STOCHASTIC PROCESSES

- Introduction to stochastic processes.
- Brownian motion and related stochastic processes and their simulation.
- Gaussian stochastic processes.
- Karhunen-Loeve expansion and simulation of Gaussian random fields.
- Numerical solution of ODEs and PDEs with random coefficients.
3. NUMERICAL SOLUTION OF STOCHASTIC DIFFERENTIAL EQUATIONS.

- Basic properties of stochastic differential equations.
- Itô’s formula and stochastic Taylor expansions.
- Examples of numerical methods: The Euler-Marayama and Milstein schemes.
- Theoretical issues: convergence, consistency, stability. Weak versus strong convergence.
- Implicit schemes and stiff SDEs.
4. ERGODIC DIFFUSION PROCESSES
   ▶ Deterministic methods: numerical solution of the Fokker-Planck equation.
   ▶ Numerical calculation of the invariant measure/steady state simulation.
   ▶ Calculation of expectation values, the Feynman-Kac formula.

5. MARKOV CHAIN MONTE CARLO (MCMC).
   ▶ The Metropolis-Hastings and MALA algorithms.
   ▶ Diffusion limits for MCMC algorithms.
   ▶ Peskun and Tierney orderings and Markov Chain comparison.
   ▶ Optimization of MCMC algorithms.
   ▶ Bias correction and variance reduction techniques.
6. STATISTICAL INFERENCE FOR DIFFUSION PROCESSES.
   ▶ Estimation of the diffusion coefficient (volatility).
   ▶ Maximum likelihood estimation of drift coefficients in SDEs.
   ▶ Nonparametric and Bayesian techniques.

7. APPLICATIONS
   ▶ Probabilistic methods for the numerical solution of partial differential equations.
   ▶ Exit time problems.
   ▶ Molecular dynamics/computational statistical mechanics.
   ▶ Molecular motors, stochastic resonance.
Prerequisites

- Basic knowledge of ODEs and PDEs.
- Familiarity with the theory of stochastic processes.
- Stochastic differential equations.
- Basic knowledge of functional analysis and stochastic analysis.
- Numerical methods and scientific computing.
- Familiarity with a programming language: Matlab, C, Fortran....
Assessment

- Based on 1 project (25%) and a final exam (75%).
- Mastery exam: additional project/paper to study and write a report.
- The project will be (mostly) computational.
- The exam will be theoretical (e.g. analysis of algorithms).
Bibliography

- Lecture notes will be provided for all the material that we will cover in this course. They will be posted on the course webpage.

- Books that cover (parts of) the contents of this course are
Lectures

- Slides and whiteboard.
- Computer experiments in Matlab.
why Computational Stochastic Processes?

- Many models in science, engineering and economics are probabilistic in nature and we have to deal with uncertainty.
- Many deterministic problems can be solved more efficiently using probabilistic techniques.
FIG. 1. $\delta^{18}$O record from the NGRIP ice core during the last glacial period.
Fit this data to a bistable SDE

\[ dx = -V'(x; \mathbf{a}) \, dt + \sigma \, dW, \quad V(x) = \sum_{j=1}^{4} a_j x^j. \]  

(1)

Estimate the coefficients in the drift from the paleoclimate data using the unscented Kalman filter.

the resulting potential is highly asymmetric.
FIG. 8. Potential derived by least-squares fit from the probability density of the ice-core data (solid) together with probability densities of the model (dashed) and the data (dotted).
Sample from the Boltzmann distribution:

$$\pi(x) = \frac{1}{Z} e^{-\beta U(x)}, \quad (2)$$

where $\beta = (k_B T)^{-1}$ is the inverse temperature and the normalization constant $Z$ is the \textbf{partition function}

$$Z = \int_{\mathbb{R}^{3k}} e^{-\beta U(x)} \, dx. \quad (3)$$

$x$ denotes the configuration of the system $x = \{x_i : i = 1, \ldots k\}$,

$$x_i = (x_{i1}, x_{i2}, x_{i3}).$$

The potential $U(x)$ is a sum of pairwise interactions

$$U(x) = \sum_{i,j} \Phi(|x_i - x_j|), \quad \Phi(r) = 4\varepsilon \left[ \left( \frac{\sigma}{r} \right)^{12} - \left( \frac{\sigma}{r} \right)^6 \right].$$
We need to be able to calculate $Z(\beta)$, which is a very high dimensional integral.

We want to be able to calculate expectations with respect to $\pi(x)$:

$$\mathbb{E}_\pi f(x) = \int_{\mathbb{R}^3 k} f(x) \pi(x) \, dx.$$

One method for doing this is by simulating a diffusion process (MCMC):

$$dX_t = -\nabla U(X_t) \, dt + \sqrt{2\beta^{-1}} \, dW_t.$$
Some interesting questions

- How many experiments should we do?
- How do we compute expectations with respect to stationary distributions?
- Can we exploit the problem structure to speed up the computation?
- How do we efficiently compute probabilities of rare events?
- How do we estimate the sensitivity of a stochastic model to changes in a parameter?
- How do we use stochastic simulation to optimize our choice of decision parameters?
Example: Brownian motion in a periodic potential

Consider the Langevin dynamics in a smooth periodic potential

\[ \ddot{q} = -\nabla V(q) - \gamma \dot{q} + \sqrt{2\gamma \beta^{-1}} \dot{W}, \quad (4) \]

where \( W(t) \) denotes standard \( d \)-dimensional Brownian motion, \( \gamma > 0 \) the friction coefficient and \( \beta \) the inverse temperature (strength of the noise).

Write as a first order system:

\[ dq_t = p_t \, dt, \quad (5a) \]
\[ dp_t = -\nabla_q V(q_t) \, dt - \gamma p_t \, dt + \sqrt{2\gamma \beta^{-1}} \, dW_t. \quad (5b) \]
At long times the solution to $q_t$ performs an effective Brownian motion with diffusion coefficient $D$, $\mathbb{E} q_t^2 \sim 2Dt$, $t \ll 1$.

We compute the diffusion coefficient for $V(q) = \cos(q)$ using the formula

$$D = \lim_{t \to \infty} \frac{\text{Var}(q_t)}{2t},$$

as a function of $\gamma$, at a fixed temperature $\beta^{-1}$.

The SDE (5) has to be solved numerically using an appropriate scheme.

The numerical simulations suggest that

$$D \sim \frac{1}{\gamma}, \quad \text{for both } \gamma \ll 1 \text{ and } \gamma \gg 1.$$  \hspace{1cm} (6)
a. $\text{Var}(q_t) / 2t$ vs $t$

b. $D$ vs $\gamma$

Figure: Variance and effective diffusion coefficient for various values of $\gamma$. 
The One-Dimensional Random Walk

We let time be discrete, i.e. \( t = 0, 1, \ldots \). Consider the following stochastic process \( S_n \):

- \( S_0 = 0 \);
- at each time step it moves to \( \pm 1 \) with equal probability \( \frac{1}{2} \).

In other words, at each time step we flip a fair coin. If the outcome is heads, we move one unit to the right. If the outcome is tails, we move one unit to the left.

Alternatively, we can think of the random walk as a sum of independent random variables:

\[
S_n = \sum_{j=1}^{n} X_j,
\]

where \( X_j \in \{-1, 1\} \) with \( \mathbb{P}(X_j = \pm 1) = \frac{1}{2} \).
We can simulate the random walk on a computer:

- We need a *(pseudo)*random number generator to generate \( n \) independent random variables which are uniformly distributed in the interval \([0,1]\).
- If the value of the random variable is \( \geq \frac{1}{2} \) then the particle moves to the left, otherwise it moves to the right.
- We then take the sum of all these random moves.
- The sequence \( \{S_n\}_{n=1}^N \) indexed by the discrete time \( T = \{1, 2, \ldots N\} \) is the **path** of the random walk. We use a linear interpolation (i.e. connect the points \( \{n, S_n\} \) by straight lines) to generate a **continuous path**.
Figure: Three paths of the random walk of length $N = 50$. 
Figure: Three paths of the random walk of length $N = 1000$. 
Every path of the random walk is different: it depends on the outcome of a sequence of independent random experiments.

We can compute statistics by generating a large number of paths and computing averages. For example, \( \mathbb{E}(S_n) = 0 \), \( \mathbb{E}(S_n^2) = n \).

The paths of the random walk (without the linear interpolation) are not continuous: the random walk has a jump of size 1 at each time step.

This is an example of a **discrete time, discrete space** stochastic processes.

The random walk is a **time-homogeneous** (the probabilistic law of evolution is independent of time) **Markov** (the future depends only on the present and not on the past) process.

If we take a large number of steps, the random walk starts looking like a continuous time process with continuous paths.
Consider the sequence of continuous time stochastic processes

\[ Z^n_t := \frac{1}{\sqrt{n}} S_{nt} \]

In the limit as \( n \to \infty \), the sequence \( \{Z^n_t\} \) converges (in some appropriate sense) to a Brownian motion with diffusion coefficient \( D = \frac{\Delta x^2}{2\Delta t} = \frac{1}{2} \).
Figure: Sample Brownian paths.
Brownian motion $W(t)$ is a continuous time stochastic processes with continuous paths that starts at 0 ($W(0) = 0$) and has independent, normally distributed Gaussian increments.

We can simulate the Brownian motion on a computer using a random number generator that generates normally distributed, independent random variables.
We can write an equation for the evolution of the paths of a Brownian motion $X_t$ with diffusion coefficient $D$ starting at $x$:

$$dX_t = \sqrt{2D}dW_t, \quad X_0 = x.$$ 

This is an example of a **stochastic differential equation**.

The probability of finding $X_t$ at $y$ at time $t$, given that it was at $x$ at time $t = 0$, the **transition probability density** $\rho(y, t)$ satisfies the PDE

$$\frac{\partial \rho}{\partial t} = D \frac{\partial^2 \rho}{\partial y^2}, \quad \rho(y, 0) = \delta(y - x).$$

This is an example of the **Fokker-Planck equation**.

The connection between Brownian motion and the diffusion equation was made by Einstein in 1905.
ELEMENTS OF PROBABILITY THEORY
A collection of subsets of a set $\Omega$ is called a $\sigma$–algebra if it contains $\Omega$ and is closed under the operations of taking complements and countable unions of its elements.

A sub-$\sigma$–algebra is a collection of subsets of a $\sigma$–algebra which satisfies the axioms of a $\sigma$–algebra.

A measurable space is a pair $(\Omega, \mathcal{F})$ where $\Omega$ is a set and $\mathcal{F}$ is a $\sigma$–algebra of subsets of $\Omega$.

Let $(\Omega, \mathcal{F})$ and $(E, \mathcal{G})$ be two measurable spaces. A function $X : \Omega \mapsto E$ such that the event

$$\{\omega \in \Omega : X(\omega) \in A\} =: \{X \in A\}$$

belongs to $\mathcal{F}$ for arbitrary $A \in \mathcal{G}$ is called a measurable function or random variable.
Let \((\Omega, \mathcal{F})\) be a measurable space. A function \(\mu : \mathcal{F} \mapsto [0, 1]\) is called a *probability measure* if \(\mu(\emptyset) = 1\), \(\mu(\Omega) = 1\) and 
\[
\mu\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} \mu(A_k)
\]
for all sequences of pairwise disjoint sets \(\{A_k\}_{k=1}^{\infty} \in \mathcal{F}\).

The triplet \((\Omega, \mathcal{F}, \mu)\) is called a *probability space*.

Let \(X\) be a random variable (measurable function) from \((\Omega, \mathcal{F}, \mu)\) to \((E, \mathcal{G})\). If \(E\) is a metric space then we may define *expectation* with respect to the measure \(\mu\) by

\[
\mathbb{E}[X] = \int_{\Omega} X(\omega) \, d\mu(\omega).
\]

More generally, let \(f : E \mapsto \mathbb{R}\) be \(\mathcal{G}\)-measurable. Then,

\[
\mathbb{E}[f(X)] = \int_{\Omega} f(X(\omega)) \, d\mu(\omega).
\]
Let $U$ be a topological space. We will use the notation $\mathcal{B}(U)$ to denote the Borel $\sigma$–algebra of $U$: the smallest $\sigma$–algebra containing all open sets of $U$. Every random variable from a probability space $(\Omega, \mathcal{F}, \mu)$ to a measurable space $(E, \mathcal{B}(E))$ induces a probability measure on $E$:

$$
\mu_X(B) = \mathbb{P}X^{-1}(B) = \mu(\omega \in \Omega; X(\omega) \in B), \quad B \in \mathcal{B}(E).
$$

The measure $\mu_X$ is called the distribution (or sometimes the law) of $X$.

**Example**

Let $\mathcal{I}$ denote a subset of the positive integers. A vector $\rho_0 = \{\rho_0,i, \ i \in \mathcal{I}\}$ is a distribution on $\mathcal{I}$ if it has nonnegative entries and its total mass equals 1: $\sum_{i \in \mathcal{I}} \rho_0,i = 1$. 
We can use the distribution of a random variable to compute expectations and probabilities:

\[ \mathbb{E}[f(X)] = \int_S f(x) \, d\mu_X(x) \]

and

\[ \mathbb{P}[X \in G] = \int_G d\mu_X(x), \quad G \in \mathcal{B}(E). \]

When \( E = \mathbb{R}^d \) and we can write \( d\mu_X(x) = \rho(x) \, dx \), then we refer to \( \rho(x) \) as the probability density function (pdf), or density with respect to Lebesgue measure for \( X \).

When \( E = \mathbb{R}^d \) then by \( L^p(\Omega; \mathbb{R}^d) \), or sometimes \( L^p(\Omega; \mu) \) or even simply \( L^p(\mu) \), we mean the Banach space of measurable functions on \( \Omega \) with norm

\[ \|X\|_{L^p} = \left( \mathbb{E}|X|^p \right)^{1/p}. \]
Example

Consider the random variable $X : \Omega \mapsto \mathbb{R}$ with pdf

$$
\gamma_{\sigma,m}(x) := (2\pi\sigma)^{-\frac{1}{2}} \exp\left(-\frac{(x - m)^2}{2\sigma}\right).
$$

Such an $X$ is termed a **Gaussian** or **normal** random variable. The mean is

$$
\mathbb{E}X = \int_{\mathbb{R}} x \gamma_{\sigma,m}(x) \, dx = m
$$

and the variance is

$$
\mathbb{E}(X - m)^2 = \int_{\mathbb{R}} (x - m)^2 \gamma_{\sigma,m}(x) \, dx = \sigma.
$$
Example (Continued)

Let \( m \in \mathbb{R}^d \) and \( \Sigma \in \mathbb{R}^{d \times d} \) be symmetric and positive definite. The random variable \( X : \Omega \mapsto \mathbb{R}^d \) with pdf

\[
\gamma_{\Sigma,m}(x) := (\frac{1}{2\pi})^d \det \Sigma \exp \left( -\frac{1}{2} \langle \Sigma^{-1}(x - m), (x - m) \rangle \right)
\]

is termed a **multivariate Gaussian** or **normal** random variable. The mean is

\[
\mathbb{E}(X) = m
\]  

(7)

and the covariance matrix is

\[
\mathbb{E} \left( (X - m) \otimes (X - m) \right) = \Sigma.
\]

(8)
Since the mean and variance specify completely a Gaussian random variable on \( \mathbb{R} \), the Gaussian is commonly denoted by \( \mathcal{N}(m, \sigma) \). The **standard normal** random variable is \( \mathcal{N}(0, 1) \).

Since the mean and covariance matrix completely specify a Gaussian random variable on \( \mathbb{R}^d \), the Gaussian is commonly denoted by \( \mathcal{N}(m, \Sigma) \).
The Characteristic Function

- Many of the properties of (sums of) random variables can be studied using the Fourier transform of the distribution function.
- The **characteristic function** of the random variable $X$ is defined to be the Fourier transform of the distribution function

$$
\phi(t) = \int_{\mathbb{R}} e^{i t \lambda} d\mu_X(\lambda) = \mathbb{E}(e^{i t X}).
$$

(9)

- The characteristic function determines uniquely the distribution function of the random variable, in the sense that there is a one-to-one correspondence between $F(\lambda)$ and $\phi(t)$.
- The characteristic function of a $\mathcal{N}(m, \Sigma)$ is

$$
\phi(t) = e^{\langle m, t \rangle - \frac{1}{2} \langle t, \Sigma t \rangle}.
$$
Lemma

Let \( \{X_1, X_2, \ldots X_n\} \) be independent random variables with characteristic functions \( \phi_j(t), j = 1, \ldots n \) and let \( Y = \sum_{j=1}^{n} X_j \) with characteristic function \( \phi_Y(t) \). Then

\[
\phi_Y(t) = \prod_{j=1}^{n} \phi_j(t).
\]

Lemma

Let \( X \) be a random variable with characteristic function \( \phi(t) \) and assume that it has finite moments. Then

\[
E(X^k) = \frac{1}{i^k} \phi^{(k)}(0).
\]
Types of Convergence and Limit Theorems

- One of the most important aspects of the theory of random variables is the study of limit theorems for sums of random variables.
- The most well known limit theorems in probability theory are the law of large numbers and the central limit theorem.
- There are various different types of convergence for sequences or random variables.
Definition

Let \( \{Z_n\}_{n=1}^{\infty} \) be a sequence of random variables. We will say that

(a) \( Z_n \) converges to \( Z \) with probability one (almost surely) if

\[
\mathbb{P}(\lim_{n\to+\infty} Z_n = Z) = 1.
\]

(b) \( Z_n \) converges to \( Z \) in probability if for every \( \varepsilon > 0 \)

\[
\lim_{n\to+\infty} \mathbb{P}(|Z_n - Z| > \varepsilon) = 0.
\]

(c) \( Z_n \) converges to \( Z \) in \( L^p \) if

\[
\lim_{n\to+\infty} \mathbb{E}[|Z_n - Z|^p] = 0.
\]

(d) Let \( F_n(\lambda), n = 1, \cdots + \infty, F(\lambda) \) be the distribution functions of \( Z_n n = 1, \cdots + \infty \) and \( Z \), respectively. Then \( Z_n \) converges to \( Z \) in distribution if

\[
\lim_{n\to+\infty} F_n(\lambda) = F(\lambda)
\]
Let \( \{X_n\}_{n=1}^{\infty} \) be iid random variables with \( \mathbb{E}X_n = V \). Then, the **strong law of large numbers** states that average of the sum of the iid converges to \( V \) with probability one:

\[
\mathbb{P}\left( \lim_{n \to +\infty} \frac{1}{N} \sum_{n=1}^{N} X_n = V \right) = 1.
\]

The strong law of large numbers provides us with information about the behavior of a sum of random variables (or, a large number or repetitions of the same experiment) on average.

We can also study fluctuations around the average behavior. Indeed, let \( \mathbb{E}(X_n - V)^2 = \sigma^2 \). Define the centered iid random variables \( Y_n = X_n - V \). Then, the sequence of random variables

\[
\frac{1}{\sigma \sqrt{N}} \sum_{n=1}^{N} Y_n
\]

converges in distribution to a \( \mathcal{N}(0, 1) \) random variable:

\[
\lim_{n \to +\infty} \mathbb{P}\left( \frac{1}{\sigma \sqrt{N}} \sum_{n=1}^{N} Y_n \leq a \right) = \int_{-\infty}^{a} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \, dx.
\]
Assume that $E|X| < \infty$ and let $\mathcal{G}$ be a sub-$\sigma$–algebra of $\mathcal{F}$. The **conditional expectation** of $X$ with respect to $\mathcal{G}$ is defined to be the function $E[X|\mathcal{G}] : \Omega \mapsto E$ which is $\mathcal{G}$–measurable and satisfies

$$\int_{\mathcal{G}} E[X|\mathcal{G}] \, d\mu = \int_{\mathcal{G}} X \, d\mu \quad \forall \, G \in \mathcal{G}.$$

We can define $E[f(X)|\mathcal{G}]$ and the conditional probability $P[X \in F|\mathcal{G}] = E[I_F(X)|\mathcal{G}]$, where $I_F$ is the indicator function of $F$, in a similar manner.
Random Number Generators

- How can a deterministic machine produce random numbers?
- We can use a chaotic dynamical system:
  \[ x_{n+1} = f(x_n), \quad x_0 = x, \quad n = 1, 2, \ldots \]

  Provided that this dynamical system is “sufficiently chaotic”, then we can hope that for \( n \) sufficiently large the sequence of numbers \( x_n, x_{n+1}, x_{n+2}, \ldots \) is random with nice statistical properties.

- Statistical tests can be used in order to determine whether this sequence of pseudo-random numbers can be used in order to perform Monte Carlo simulations, simulate stochastic processes etc.

- Modern programming languages have well tested build in RNG. In Matlab:
  - `rand U(0,1)`.  
  - `randn N(0,1)`.  

Definition
A uniform pseudo-number generator is an algorithm which, starting from an initial value $x_0$ (the seed), produces a sequence $u_i = D^i u_0$ of values in $[0, 1]$.

Lemma
Suppose $Y \sim U(0, 1)$ and $F$ a 1d cumulative distribution function. Then $X = F^{-1}(U)$ with $F^{-1}(u) = \inf \{x; F(x) \geq u\}$ has the distribution $F$.

We can use the Box-Muller algorithm to generate Gaussian random variables, given a pseudonumber generator of uniform r.v.: Given $U_1, U_2 \sim U(0, 1)$ independent, then

$$Z_0 = \sqrt{-2 \ln U_1 \cos(2\pi U_2)},$$
$$Z_1 = \sqrt{-2 \ln U_1 \sin(2\pi U_2)}$$

are independent $N(0, 1)$ r.v.
Fast Chaotic Noise

- **Fast/slow system:**

\[
\begin{align*}
\frac{dx}{dt} &= x - x^3 + \frac{\lambda}{\varepsilon} y_2, \\
\frac{dy_1}{dt} &= \frac{10}{\varepsilon^2} (y_2 - y_1), \\
\frac{dy_2}{dt} &= \frac{1}{\varepsilon^2} (28y_1 - y_2 - y_1y_3), \\
\frac{dy_3}{dt} &= \frac{1}{\varepsilon^2} (y_1y_2 - \frac{8}{3} y_3)
\end{align*}
\]

- **Effective Dynamics:** [Melbourne, Stuart ’11]

\[
dX_t = A (X_t - X_t^3) \ dt + \sqrt{\sigma} \ dW_t
\]

- **true values:**

\[
A = 1, \quad \lambda = \frac{2}{45}, \quad \sigma = 2\lambda^2 \int_0^\infty \lim_{T \to \infty} \frac{1}{T} \int_0^T \psi^s(y) \psi^{s+t}(y) \ ds \ dt
\]
Values for $\sigma$ reported in the literature ($\varepsilon = 10^{-3/2}$)

- $0.126 \pm 0.003$ via Gaussian moment approx.
- $0.13 \pm 0.01$ via HMM

Here: $\varepsilon = 10^{-1} \rightarrow \hat{\sigma} \approx 0.121$ and $\varepsilon = 10^{-3/2} \rightarrow \hat{\sigma} \approx 0.124$

But we estimate also $\hat{A}$
Generation of Multidimensional Gaussian Random Variables

- We are given $Z \sim \mathcal{N}(0, I)$, where $I \in \mathbb{R}^{d \times d}$ denotes the identity matrix.
- We want to generate a Gaussian random vector $Y \sim \mathcal{N}(b, \Sigma)$, $\mathbb{E}Y = b$, $\mathbb{E}((Y - b)(Y - b)^T) = \Sigma$.
- $X = Y - b$ is mean zero, sufficient to consider this case. 
- We can use the Cholesky decomposition: Assuming that $\Sigma > 0$ there exists a unique lower triangular matrix such that

\[ \Sigma = \Lambda \Lambda^T. \]  

Then: $X = \Lambda Z$ is $X \sim \mathcal{N}(0, \Sigma)$.
- The computational cost is $O(d^3)$.
- In matlab:

\[ \text{chol}(\Sigma) \]
Eigenvalue decomposition of \( \text{Cov}(X) \)

- The Cholesky decomposition is not unique for covariance matrices that are not strictly positive definite.
- The eigenvalues and eigenvectors of the covariance matrix provide us complete information about the Gaussian random vector.
- Principal Component Analysis.

Let \( X \sim \mathcal{N}(0, \Sigma) \), \( \Sigma \geq 0 \). Then

\[
\Sigma = PDP^T, \quad PP^T = I, \quad D = \text{diag}(d_1, d_d).
\] (11)

- Set \( L = P\sqrt{D} \) and use the Cholesky decomposition.
- In Matlab:

\[
[P,D] = \text{eig} \left( \Sigma \right)
\]
Other techniques for generating random numbers with a given distribution:

- Acceptance-rejection algorithm.
- Markov Chain Monte Carlo.
Monte Carlo Techniques

• Example:

\[ I = \int_D f(x) \, dx = \mathbb{E}f(X), \quad x \sim (U(D))^d \]

\[ \approx \frac{1}{N} \sum_{i=1}^{N} f(x_i) =: I_N, \quad x_i \text{ iid } U(D). \]

• To prove convergence use the (strong) law of large numbers:

\[ \lim_{{N \to +\infty}} \frac{1}{N} \sum_{i=1}^{N} f(x_i) = I \quad \text{a.s.} \]
To obtain an estimate on the rate of convergence use the central limit theorem:

\[ \sqrt{N}(I_N - I) \rightarrow N(0, \sigma^2), \quad \text{in distribution,} \]

where \( \sigma^2 = \text{Var}_\pi(f(x)) \).

The error of the Monte Carlo approximation is \( O(N^{-1/2}) \), regardless of the dimensionality of \( x \).

For smooth functions, deterministic approximations (Riemann sum, trapezoidal rule..) can have better convergence rate, but they don’t scale well as the number of dimensions increases.
More generally, consider the integral

\[ I = \int_{\mathbb{R}^d} f(x) \pi(x) \, dx, \tag{12} \]

where \( \pi(x) \) is a probability distribution. We can always write integrals in this form (multiply and divide by \( \pi(x) \)).

Monte Carlo estimator:

\[ I_N = \frac{1}{N} \sum_{i=1}^{N} f(X_i), \quad X_i \sim \pi(x), \text{ iid.} \tag{13} \]

The MC estimator is unbiased:

\[ \mathbb{E}_\pi I_N = I. \]

Variance of the estimator:

\[ \sigma_I^2 := \text{var}_\pi (I_N) = \frac{1}{N} \text{var}_\pi (f(x)) = \frac{1}{N} \int_{\mathbb{R}^d} (f(x) - I)^2 \pi(x) \, dx. \]
From the strong law of large numbers:

$$\lim_{N \to +\infty} I_N = I \quad \text{a.s.}$$

From the weak LLN and Chebyshev’s inequality:

$$\mathbb{P}(|I_N - I| \leq \varepsilon) \geq 1 - \frac{\sigma^2_I}{\varepsilon^2} = 1 - \frac{\sigma^2_f}{N\varepsilon^2}. \quad (14)$$

Accuracy of the MC estimator depends on $N$ and on the variance $\sigma^2_I$.

In order for

$$\mathbb{P}(I_N \in [I - \varepsilon, I + \varepsilon]) = 1 - \alpha,$$

for a confidence level $\alpha \in (0, 1)$ we need

$$N > \frac{\sigma^2_f}{\alpha\varepsilon^2}.$$ 

MC error scales like $1/\sqrt{N}$, independently of the number of dimensions. It is expected that $\sigma^2_f$ increases as the number of dimensions increases.
Example (Huynh et al Stochastic Simulation and Applications in finance ch. 5)

```
function [Call, VarCall]=CalculateCall2(N)

% Price a call option
% N: Number of generated points to compute the expectation

x = exp(sqrt(0.1) * randn(1,N) + 5);
Call = mean(max(x - 110,0));
VarCall = var(max(x - 110,0))/N;
end
```
VARIANCE REDUCTION TECHNIQUES

To improve the accuracy of the MC estimator and to reduce the computational cost we can use variance reduction techniques.

Examples of such techniques are:
- Antithetic variables.
- Control variates.
- Importance sampling.

To implement these techniques we need to be able to estimate the mean and variance "on the fly":

\[
I_{N+1} = \frac{1}{N+1} (NI_N + f(X_{N+1}))
\]

and

\[
\text{var}(I_N) = \frac{1}{N} \left( \frac{1}{N-1} \sum_{i=1}^{N} |f(X_i)|^2 - \frac{N}{N-1} I_N^2 \right).
\]
Antithetic Variables: generate additional random variables that have the same distribution as $X_i, \ i = 1, \ldots N$ but are negatively correlated to $X_i$.

Example: suppose we want to estimate the mean $\mu_X$ of iid random variables $X_i, \ i = 1, \ldots N$. We introduce the auxiliary r.v. $X_i^a, \ i = 1, \ldots N$ with

$$E[(X_i^a - \mu_x)(X_j - \mu_x)] = -\delta_{ij}\sigma_X^2.$$ 

The variance of the sample mean is

$$\text{Var}(\hat{\mu}_a) = \frac{1}{2N^2}\sigma_X^2 + \frac{1}{4N^2} \sum_{k,j=1}^{N} E[(X_i^a - \mu_x)(X_j - \mu_x)]$$

$$= \frac{1}{4N}\sigma_X^2.$$ 

Not possible in general to find perfectly anti-correlated random variables.
Example
Let $X \sim U(0, 1)$. Then $X^a = 1 - X$ is an antithetic variable of $X$.

Example
Let $X \sim \mathcal{N}(\mu, \sigma^2)$. Then $X^a = 2\mu - X$ is an antithetic variable of $X$.

Example
Let $X \sim \mathcal{N}(\mu, \Sigma)$ a 2d Gaussian random vector. We use antithetic variables to reduce the variance and increase the accuracy of the estimation of $\mathbb{E}f(X_1, X_2)$. We will estimate the expectation of $Y_j = e^{X_j}$, $j = 1, 2$. These integrals can be calculated analytically.
function Rep=SimulationsAnti
%calculate the mean of an exponential gaussian using
%antithetic variables.
NbTraj=100;
% construct the covariance matrix
sigma1=0.2; sigma2=0.5; rho=-0.3;
MoyTheo=[10;15];
CovTheo=[sigma1^2,rho * sigma1 * sigma2;rho * sigma1 * sigma2,sigma2^2];
L=chol(CovTheo)';
%Simulation of the random variables.
Sample=randn(2,NbTraj);
SampleSimple=repmat(MoyTheo,1,NbTraj)+L * Sample;
%Introduction of the antithetic variables
XAnti=cat(2,SampleSimple,2 * repmat(MoyTheo,1,NbTraj)-SampleSimple
%Transformation of Xi into Zi
Z1Anti=exp(XAnti(1,:))
Z2Anti=exp(XAnti(2,:))
Rep=mean(Z1Anti,2);
end
CONTROL VARIATES

Let $Z$ be a random variable. We want to estimate $\mathbb{E}Z$.

We look for a r.v. $W$ that is strongly correlated with $Z$ and with known $\mathbb{E}W$.

Consider the new r.v.

$$X = Z + \alpha(W - \mathbb{E}W).$$

We have

$$\text{Var}(X) = \text{Var}(Z) + \alpha^2 \text{Var}(W) + 2\alpha \text{Cov}(Z, W).$$

We minimize the variance with respect to $\alpha$:

$$\alpha = -\frac{\text{Cov}(Z, W)}{\text{Var}(W)} \quad \text{and} \quad \text{Var}_{\text{min}}(X) = \text{Var}(Z)(1 - \rho^2).$$
CONTROL VARIATES

- The estimator (16) is unbiased and the variance is reduced when the correlation coefficient $\rho^2 = \frac{\text{Cov}(Z,W)}{\text{Var}(Z)\text{Var}(W)}$ is close to 1.
- The optimal value of $\alpha$ and $\text{Var}(X)$ can be estimated using the empirical means.
- We can also construct confidence intervals using the empirical values.
- We can add $R$ control variates:

$$X = Z + \sum_{j=1}^{R} \alpha(W_j - \mathbb{E}W_j).$$

- A similar idea can be used for vector valued r.v. $X$, or even in infinite dimensions.
Importance sampling: choose a good probability distribution from which to simulate random variables:

\[
\int_{\mathbb{R}^d} f(x) \pi(x) \, dx \quad = \quad \int_{\mathbb{R}^d} \frac{f(x) \pi(x)}{\psi(x)} \psi(x) \, dx
\]

\[
= \mathbb{E}_\psi \left( \frac{f(x) \pi(x)}{\psi(x)} \right).
\]

where \( \psi(x) \) is a probability distribution.

Consider the MC estimator

\[
I_N^\psi = \frac{1}{N} \sum_{j=1}^{N} \frac{f(X_j) \pi(X_j)}{\psi(X_j)}, \quad X_j \sim \psi(x).
\]

The estimator is unbiased. The variance is

\[
\text{var}(I_N^\psi) = \frac{1}{N} \text{var} \left( \frac{f(X) \pi(X)}{\psi(X)} \right).
\]
We can choose \( \psi(x) \) so that it has similar shape to \( f(x) \pi(x) \).

The closer \( \frac{f(X) \pi(X)}{\psi(X)} \) is to a constant the smaller the variance is.

**Example**

We use importance sampling to estimate the integral

\[
I = \int_0^1 c \exp(-a(x - b)^4) \, dx = 18.128
\]

with \( c = 100, \ a = 10000, \ b = 0.5 \). We choose \( \psi = \gamma(0.5, 0.05) \);

For \( N = 10000 \) we have \( I_N = 18.2279, \ \text{var}(I_N) = 0.1193, \ \text{err}_N = 0.0999 \), \( I_{IC} = 18.1271, \ \text{var}(I_{IC}) = 0.0041, \ \text{err}(I_{IC}) = 8.56 \times 10^{-4} \).
Figure: $c \exp(-a(x - b)^4)$ and $\gamma(0.5, 0.05)$
function [yun,varyun,errun,yIC,varyIC,errIC] = quarticIS(N)
% Use importance sampling to calculate the mean
% of exp(-V) where V is a quartic polynomial
a = 10000;b=0.5;c=100;
intexact = c*0.18128;
x = rand(N,1);
yun = mean(c*exp(-a*(x-b).^4));
varyun = var(c*exp(-a*(x-b).^4))/N;
errun = abs(intexact-yun);
mu = 0.5;sigma = 0.05;
xx = mu + sigma*randn(N,1);
yIC = mean((c*exp(-a*(xx-b).^4))./normpdf(xx,mu,sigma));
varyIC = var((c*exp(-a*(xx-b).^4))./normpdf(xx,mu,sigma))/N;
errIC = abs(intexact-yIC);
ELEMENTS OF THE THEORY OF STOCHASTIC PROCESSES
Let $T$ be an ordered set. A **stochastic process** is a collection of random variables $X = \{X_t; t \in T\}$ where, for each fixed $t \in T$, $X_t$ is a random variable from $(\Omega, \mathcal{F})$ to $(E, \mathcal{G})$.

The measurable space $\{\Omega, \mathcal{F}\}$ is called the **sample space**. The space $(E, \mathcal{G})$ is called the **state space**.

In this course we will take the set $T$ to be $[0, +\infty)$.

The state space $E$ will usually be $\mathbb{R}^d$ equipped with the $\sigma$–algebra of Borel sets.

A stochastic process $X$ may be viewed as a function of both $t \in T$ and $\omega \in \Omega$. We will sometimes write $X(t), X(t, \omega)$ or $X_t(\omega)$ instead of $X_t$. For a fixed sample point $\omega \in \Omega$, the function $X_t(\omega) : T \mapsto E$ is called a **sample path** (realization, trajectory) of the process $X$. 
The **finite dimensional distributions** (fdd) of a stochastic process are the distributions of the $E^k$-valued random variables $(X(t_1), X(t_2), \ldots, X(t_k))$ for arbitrary positive integer $k$ and arbitrary times $t_i \in T$, $i \in \{1, \ldots, k\}$:

$$F(x) = \mathbb{P}(X(t_i) \leq x_i, i = 1, \ldots, k)$$

with $x = (x_1, \ldots, x_k)$.

We will say that two processes $X_t$ and $Y_t$ are equivalent if they have same finite dimensional distributions.

From experiments or numerical simulations we can only obtain information about the (fdd) of a process.
A **Gaussian process** is a stochastic processes for which $E = \mathbb{R}^d$ and all the finite dimensional distributions are Gaussian

$$F(x) = \mathbb{P}(X(t_i) \leq x_i, i = 1, \ldots, k) = (2\pi)^{-n/2} (\det K_k)^{-1/2} \exp \left[ -\frac{1}{2} \langle K_k^{-1}(x - \mu_k), x - \mu_k \rangle \right],$$

for some vector $\mu_k$ and a symmetric positive definite matrix $K_k$.

- A Gaussian process $x(t)$ is characterized by its mean
  $$m(t) := \mathbb{E}x(t)$$
  and the covariance function
  $$C(t, s) = \mathbb{E}\left( (x(t) - m(t)) \otimes (x(s) - m(s)) \right).$$

- Thus, the first two moments of a Gaussian process are sufficient for a complete characterization of the process.
Examples of Gaussian stochastic processes

- Random Fourier series: let $\xi_i, \zeta_i \sim \mathcal{N}(0, 1), \ i = 1, \ldots, N$ and define
  \[
  X(t) = \sum_{j=1}^{N} (\xi_j \cos(2\pi jt) + \zeta_j \sin(2\pi jt)).
  \]

- Brownian motion is a Gaussian process with $m(t) = 0, \ C(t, s) = \min(t, s)$.

- Brownian bridge is a Gaussian process with $m(t) = 0, \ C(t, s) = \min(t, s) - ts$.

- The Ornstein-Uhlenbeck process is a Gaussian process with $m(t) = 0, \ C(t, s) = \lambda e^{-\alpha|t-s|}$ with $\alpha, \lambda > 0$. 
To simulate a Gaussian stochastic process:

- Fix $\Delta t$ and define $t_j = (j - 1) \Delta t$, $j = 1, \ldots, N$.
- Set $X_j := X(t_j)$ and define the Gaussian random vector $X^N = \{X_j\}_{j=1}^N$. Then $X^N \sim \mathcal{N}(\mu^N, \Gamma^N)$ with $\mu^N = (\mu(t_1), \ldots, \mu(t_N))$ and $\Gamma^N_{ij} = C(t_i, t_j)$.
- Then $X^N = \mu^N + \Lambda\mathcal{N}(0, I)$ with $\Gamma^N = \Lambda\Lambda^T$.

We can calculate the square root of the covariance matrix either using the Cholesky factorization or by calculating its eigenvalue decomposition (PCA).

$\Gamma$ might be a full matrix and the calculation of $\Lambda$ can be computationally expensive.
function [t,x,gamma,mx,varx] = gaussiansimulation(T,dt,M)
% simulate a mean zero gaussian process
% use the eigenvalue decomposition to calculate \sqrt{\Gamma}
% Input: T: length of path
% dt: timestep
% gamma: covariance
% M : number of paths simulated
% calculate first and second moment
% t = 0:dt:T;
N = length(t);
for i =1:N,
    for j=1:N,
        gamma(i,j) = covBM(t(i),t(j));
        gamma(i,j) = covOU(t(i),t(j),1,1);
        gamma(i,j) = covBB(t(i),t(j));
        gamma(i,j) = covfBM(t(i),t(j),0.2);
    end
end
[vv,dd] = eig(gamma);lambda = vv*sqrt(dd)*vv';
x = lambda*randn(N,M);
xm = mean(x'); varx = mean(x'.*x');
figure;plot(t,x(:,1:10),'Linewidth',1.0);
function rr = covBM(x,y)
% covariance function of Brownian motion
rr = min(x,y);
end

function rr = covBB(x,y)
% covariance function of Brownian motion
rr = min(x,y) - x*y;
end

function rr = covOU(x,y,dd,aa)
% covariance function of Ornstein–Uhlenbeck process
% aa: inverse correlation time
% dd: diffusion coefficient, rr(o)= dd/aa
rr = (dd/aa)*exp(-aa*abs(x-y));
end
function rr = covfBM(x,y,H)
% covariance function of fractional Brownian motion
% H: Hurst exponent
rr = (1/2) * (abs(x)^(2*H) + abs(y)^(2*H) - abs(x-y)^(2*H) ...
); 
end
Let \((\Omega, \mathcal{F}, \mathbb{P})\) be a probability space. Let \(X_t, t \in T\) (with \(T = \mathbb{R}\) or \(\mathbb{Z}\)) be a real-valued random process on this probability space with finite second moment, \(\mathbb{E}|X_t|^2 < +\infty\) (i.e. \(X_t \in L^2\)).

**Definition**

A stochastic process \(X_t \in L^2\) is called **second-order stationary** or **wide-sense stationary** if the first moment \(\mathbb{E}X_t\) is a constant and the second moment \(\mathbb{E}(X_tX_s)\) depends only on the difference \(t - s\):

\[
\mathbb{E}X_t = \mu, \quad \mathbb{E}(X_tX_s) = C(t - s).
\]
The constant $m$ is called the **expectation** of the process $X_t$. We will set $m = 0$.

The function $C(t)$ is called the **covariance** or the **autocorrelation function** of the $X_t$.

Notice that $C(t) = \mathbb{E}(X_tX_0)$, whereas $C(0) = \mathbb{E}(X_t^2)$, which is finite, by assumption.

Since we have assumed that $X_t$ is a real valued process, we have that $C(t) = C(-t)$, $t \in \mathbb{R}$.
The correlation function of a second order stationary process enables us to associate a time scale to $X_t$, the **correlation time** $\tau_{\text{cor}}$:

$$
\tau_{\text{cor}} = \frac{1}{C(0)} \int_{0}^{\infty} C(\tau) \, d\tau = \int_{0}^{\infty} \frac{\mathbb{E}(X_\tau X_0)}{\mathbb{E}(X_0^2)} \, d\tau.
$$

The slower the decay of the correlation function, the larger the correlation time is. We have to assume sufficiently fast decay of correlations so that the correlation time is finite.
Example

Consider the mean zero, second order stationary process with covariance function

\[ R(t) = \frac{D}{\alpha} e^{-\alpha |t|}. \]  

(17)

The spectral density of this process is:

\[
 f(x) = \frac{1}{2\pi} \frac{D}{\alpha} \int_{-\infty}^{+\infty} e^{-ixt} e^{-\alpha |t|} dt 
\]

\[
 = \frac{1}{2\pi} \frac{D}{\alpha} \left( \int_{-\infty}^{0} e^{-ixt} e^{\alpha t} dt + \int_{0}^{+\infty} e^{-ixt} e^{-\alpha t} dt \right) 
\]

\[
 = \frac{1}{2\pi} \frac{D}{\alpha} \left( \frac{1}{ix + \alpha} + \frac{1}{-ix + \alpha} \right) 
\]

\[
 = \frac{D}{\pi} \frac{1}{x^2 + \alpha^2}. 
\]
Example (Continued)

- This function is called the **Cauchy** or the **Lorentz** distribution.
- The Gaussian stochastic process with covariance function (17) is called the **stationary Ornstein-Uhlenbeck process**.
- The correlation time is (we have that $C(0) = D/\alpha$)

$$
\tau_{cor} = \int_{0}^{\infty} e^{-\alpha t} \, dt = \alpha^{-1}.
$$
Second order stationary processes are ergodic: time averages equal phase space (ensemble) averages. An example of an ergodic theorem for a stationary processes is the following $L^2$ (mean-square) ergodic theorem.

**Theorem**

Let $\{X_t\}_{t \geq 0}$ be a second order stationary process on a probability space $\Omega$, $\mathcal{F}$, $\mathbb{P}$ with mean $\mu$ and covariance $R(t)$, and assume that $R(t) \in L^1(0, +\infty)$. Then

$$\lim_{T \to +\infty} \mathbb{E} \left\| \frac{1}{T} \int_0^T X(s) \, ds - \mu \right\|^2 = 0. \quad (18)$$
Proof.

We have

\[ \mathbb{E} \left| \frac{1}{T} \int_0^T X(s) \, ds - \mu \right|^2 = \frac{1}{T^2} \int_0^T \int_0^T R(t - s) \, dt \, ds \\
= \frac{2}{T^2} \int_0^T \int_0^t R(t - s) \, ds \, dt \\
= \frac{2}{T^2} \int_0^T (T - v)R(u) \, du \to 0, \]

using the dominated convergence theorem and the assumption \( R(\cdot) \in L^1 \). In the above we used the fact that \( R \) is a symmetric function, together with the change of variables \( u = t - s, \quad v = t \) and an integration over \( v \).
Definition

A stochastic process is called (strictly) stationary if all finite dimensional distributions are invariant under time translation: for any integer $k$ and times $t_i \in T$, the distribution of $(X(t_1), X(t_2), \ldots, X(t_k))$ is equal to that of $(X(s + t_1), X(s + t_2), \ldots, X(s + t_k))$ for any $s$ such that $s + t_i \in T$ for all $i \in \{1, \ldots, k\}$. In other words,

$$
\mathbb{P}(X_{t_1+t} \in A_1, X_{t_2+t} \in A_2 \ldots X_{t_k+t} \in A_k) = \mathbb{P}(X_{t_1} \in A_1, X_{t_2} \in A_2 \ldots X_{t_k} \in A_k), \quad \forall t \in T.
$$

Let $X_t$ be a strictly stationary stochastic process with finite second moment (i.e. $X_t \in L^2$). The definition of strict stationarity implies that $\mathbb{E}X_t = \mu$, a constant, and $\mathbb{E}((X_t - \mu)(X_s - \mu)) = C(t - s)$. Hence, a strictly stationary process with finite second moment is also stationary in the wide sense. The converse is not true.
1. A sequence \( Y_0, Y_1, \ldots \) of independent, identically distributed random variables is a stationary process with 
\[
R(k) = \sigma^2 \delta_{k0}, \quad \sigma^2 = \mathbb{E}(Y_k)^2.
\]

2. Let \( Z \) be a single random variable with known distribution and set 
\( Z_j = Z, \ j = 0, 1, 2, \ldots \). Then the sequence \( Z_0, Z_1, Z_2, \ldots \) is a stationary sequence with 
\( R(k) = \sigma^2 \).

3. The first two moments of a Gaussian process are sufficient for a complete characterization of the process. A corollary of this is that a second order stationary Gaussian process is also a (strictly) stationary process.
The most important continuous time stochastic process is **Brownian motion**. Brownian motion is a mean zero, continuous (i.e. it has continuous sample paths: for a.e $\omega \in \Omega$ the function $X_t$ is a continuous function of time) process with independent Gaussian increments.

A process $X_t$ has **independent increments** if for every sequence $t_0 < t_1 \ldots t_n$ the random variables

$$X_{t_1} - X_{t_0}, X_{t_2} - X_{t_1}, \ldots, X_{t_n} - X_{t_{n-1}}$$

are independent.

If, furthermore, for any $t_1$, $t_2$ and Borel set $B \subset \mathbb{R}$

$$\mathbb{P}(X_{t_2+s} - X_{t_1+s} \in B)$$

is independent of $s$, then the process $X_t$ has **stationary independent increments**.
Definition

A one dimensional standard Brownian motion \( W(t) : \mathbb{R}^+ \rightarrow \mathbb{R} \) is a real valued stochastic process with the following properties:

1. \( W(0) = 0 \);
2. \( W(t) \) is continuous;
3. \( W(t) \) has independent increments.
4. For every \( t > s \geq 0 \) \( W(t) - W(s) \) has a Gaussian distribution with mean 0 and variance \( t - s \). That is, the density of the random variable \( W(t) - W(s) \) is

\[
g(x; t, s) = \left( 2\pi(t - s) \right)^{-\frac{1}{2}} \exp \left( -\frac{x^2}{2(t - s)} \right); \tag{19}
\]
A $d$–dimensional standard Brownian motion $W(t) : \mathbb{R}^+ \to \mathbb{R}^d$ is a collection of $d$ independent one dimensional Brownian motions:

$$W(t) = (W_1(t), \ldots, W_d(t)),$$

where $W_i(t), i = 1, \ldots, d$ are independent one dimensional Brownian motions. The density of the Gaussian random vector $W(t) - W(s)$ is thus

$$g(x; t, s) = \left(2\pi(t - s)\right)^{-d/2} \exp\left(-\frac{\|x\|^2}{2(t - s)}\right).$$

Brownian motion is sometimes referred to as the **Wiener process**.
To generate Brownian paths we can use the general algorithm for simulating Gaussian processes with \( \mu^N = 0 \) and 
\[
\{ \Gamma = \min(t_i, t_j) \}_{i,j=1}^N.
\]

It is easier to use the fact that Brownian motion has independent increments with \( dW \sim \mathcal{N}(0, dt) \):

- \( W_j = W_{j-1} + dW_j, j = 1, 2, \ldots N, \)
- \( \text{where } dW_j = \sqrt{\Delta t} \mathcal{N}(0, 1). \)

After having generated Brownian paths we can compute statistics using Monte Carlo

\[
\mathbb{E}f(W_t) \approx \frac{1}{N} \sum_{j=1}^{N} f(W^j_t).
\]

We can also simulate processes related to Brownian motion such as the **geometric Brownian motion**:

\[
S_t = S_0 e^{at+\sigma W_t}.
\]
Figure: Brownian sample paths
It is possible to prove rigorously the existence of the Wiener process (Brownian motion):

**Theorem**

(Wiener) There exists an almost-surely continuous process \( W_t \) with independent increments such and \( W_0 = 0 \), such that for each \( t \geq 0 \) the random variable \( W_t \) is \( \mathcal{N}(0, t) \). Furthermore, \( W_t \) is almost surely locally Hölder continuous with exponent \( \alpha \) for any \( \alpha \in (0, \frac{1}{2}) \).

Notice that Brownian paths are not differentiable.
Brownian motion is a Gaussian process. For the $d$–dimensional Brownian motion, and for $I$ the $d \times d$ dimensional identity, we have (see (7) and (8))

$$\mathbb{E}W(t) = 0 \quad \forall t \geq 0$$

and

$$\mathbb{E}\left((W(t) - W(s)) \otimes (W(t) - W(s))\right) = (t - s)I. \quad (21)$$

Moreover,

$$\mathbb{E}\left(W(t) \otimes W(s)\right) = \min(t, s)I. \quad (22)$$
From the formula for the Gaussian density \( g(x, t - s) \), eqn. (19), we immediately conclude that \( W(t) - W(s) \) and \( W(t + u) - W(s + u) \) have the same pdf. Consequently, Brownian motion has stationary increments.

Notice, however, that Brownian motion itself is not a stationary process.

Since \( W(t) = W(t) - W(0) \), the pdf of \( W(t) \) is

\[
g(x, t) = \frac{1}{\sqrt{2\pi t}} e^{-x^2/2t}.
\]

We can easily calculate all moments of the Brownian motion:

\[
\mathbb{E}(x^n(t)) = \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{+\infty} x^n e^{-x^2/2t} \, dx
\]

\[
= \begin{cases} 
1.3 \ldots (n - 1)t^{n/2}, & n \text{ even}, \\
0, & n \text{ odd}.
\end{cases}
\]
We can define the OU process through the Brownian motion via a time change.

**Lemma**

Let $W(t)$ be a standard Brownian motion and consider the process

$$V(t) = e^{-t} W(e^{2t}).$$

Then $V(t)$ is a Gaussian second order stationary process with mean 0 and covariance

$$K(s, t) = e^{-|t-s|}.$$
Definition

A (normalized) fractional Brownian motion $W_t^H$, $t \geq 0$ with Hurst parameter $H \in (0, 1)$ is a centered Gaussian process with continuous sample paths whose covariance is given by

$$
\mathbb{E}(W_t^H W_s^H) = \frac{1}{2} (s^{2H} + t^{2H} - |t - s|^{2H}).
$$

(23)

Fractional Brownian motion has the following properties.

1. When $H = \frac{1}{2}$, $W_t^{\frac{1}{2}}$ becomes the standard Brownian motion.
2. $W_0^H = 0$, $\mathbb{E}W_t^H = 0$, $\mathbb{E}(W_t^H)^2 = |t|^{2H}$, $t \geq 0$.
3. It has stationary increments, $\mathbb{E}(W_t^H - W_s^H)^2 = |t - s|^{2H}$.
4. It has the following self similarity property

$$
(W_{\alpha t}^H, t \geq 0) = (\alpha^H W_t^H, t \geq 0), \quad \alpha > 0,
$$

where the equivalence is in law.
A very useful result is that we can expand every centered \( \mathbb{E}X_t = 0 \) stochastic process with continuous covariance (and hence, every \( L^2 \)-continuous centered stochastic process) into a random Fourier series.

Let \( (\Omega, \mathcal{F}, \mathbb{P}) \) be a probability space and \( \{X_t\}_{t \in T} \) a centered process which is mean square continuous. Then, the **Karhunen-Loéve** theorem states that \( X_t \) can be expanded in the form

\[
X_t = \sum_{n=1}^{\infty} \xi_n \phi_n(t),
\]

(24)

where the \( \xi_n \) are an orthogonal sequence of random variables with \( \mathbb{E}|\xi_k|^2 = \lambda_k \), where \( \{\lambda_k \phi_n\}_{k=1}^{\infty} \) are the eigenvalues and eigenfunctions of the integral operator whose kernel is the covariance of the processes \( X_t \). The convergence is in \( L^2(\mathbb{P}) \) for every \( t \in T \).

If \( X_t \) is a Gaussian process then the \( \xi_k \) are independent Gaussian random variables.
Set $T = [0, 1].$ Let $\mathcal{K} : L^2(T) \to L^2(T)$ be defined by

$$\mathcal{K}\psi(t) = \int_0^1 K(s, t)\psi(s) \, ds.$$ 

The kernel of this integral operator is a continuous (in both $s$ and $t$), symmetric, nonnegative function. Hence, the corresponding integral operator has eigenvalues, eigenfunctions

$$\mathcal{K}\phi_n = \lambda_n, \quad \lambda_n \geq 0, \quad (\phi_n, \phi_m)_{L^2} = \delta_{nm},$$

such that the covariance operator can be expanded in a uniformly convergent series

$$B(t, s) = \sum_{n=1}^{\infty} \lambda_n \phi_n(t)\phi_n(s).$$
The random variables $\xi_n$ are defined as

$$\xi_n = \int_0^1 X(t)\phi_n(t) \, dt.$$ 

The orthogonality of the eigenfunctions of the covariance operator implies that

$$\mathbb{E}(\xi_n\xi_m) = \lambda_n \delta_{nm}.$$ 

Thus the random variables are orthogonal. When $X(t)$ is Gaussian, then $\xi_k$ are Gaussian random variables. Furthermore, since they are also orthogonal, they are independent Gaussian random variables.
Example

The Karhunen-Loéve Expansion for Brownian Motion

We set $T = [0, 1]$. The covariance function of Brownian motion is $C(t, s) = \min(t, s)$. The eigenvalue problem $C \psi_n = \lambda_n \psi_n$ becomes

$$\int_0^1 \min(t, s) \psi_n(s) \, ds = \lambda_n \psi_n(t).$$

Or,

$$\int_0^1 s \psi_n(s) \, ds + t \int_t^1 \psi_n(s) \, ds = \lambda_n \psi_n(t).$$

We differentiate this equation twice:

$$\int_t^1 \psi_n(s) \, ds = \lambda_n \psi_n'(t) \quad \text{and} \quad -\psi_n(t) = \lambda_n \psi_n''(t),$$

where primes denote differentiation with respect to $t$. 
Example

From the above equations we immediately see that the right boundary conditions are \( \psi(0) = \psi'(1) = 0 \). The eigenvalues and eigenfunctions are

\[
\psi_n(t) = \sqrt{2} \sin \left( \frac{1}{2} (2n - 1) \pi t \right), \quad \lambda_n = \left( \frac{2}{(2n - 1) \pi} \right)^2.
\]

Thus, the Karhunen-Loéve expansion of Brownian motion on \([0, 1]\) is

\[
W_t = \sqrt{2} \sum_{n=1}^{\infty} \xi_n \frac{\sin \left( \left( n - \frac{1}{2} \right) \pi t \right)}{\left( n - \frac{1}{2} \right) \pi}.
\] (25)
Example KL expansion for Brownian motion

```matlab
function [t,bb] = KLBmotion(dt)
% generate paths of Brownian motion on [0,1]
% using the KL expansion

t = 0:dt:1;
N = length(t);
bb = zeros(1,N);
for i = 1:N,
    xi = sqrt(2) * randn/((i-0.5)*pi);
    bb = bb + xi * sin((i-0.5)*pi*t);
end
figure;plot(t,bb,'Linewidth',2);
end
```
Example

The Karhunen-Loéve expansion of the Brownian bridge $B_t = W_t - tW_1$ on $[0, 1]$ is

$$B_t = \sum_{n=1}^{\infty} \xi_n \frac{\sqrt{2} \sin (n\pi t)}{n\pi}.$$  

(26)
Example KL expansion for Brownian bridge

```matlab
function [t,bb] = KLBBridge(dt)
% generate paths of Brownian bridge on [0,1]
% using the KL expansion

t = 0:dt:1;
N = length(t);
bb = zeros(1,N);
for i = 1:N,
    xi = sqrt(2) * randn/(i*pi);
    bb = bb + xi * sin(i*pi*t);
end
figure;plot(t,bb,'Linewidth',2);
end
```
STOCHASTIC DIFFERENTIAL EQUATIONS (SDEs)
In this part of the course we will study stochastic differential equation (SDEs): ODEs driven by Gaussian white noise.

Let $W(t)$ denote a standard $m$–dimensional Brownian motion, $h : Z \rightarrow \mathbb{R}^d$ a smooth vector-valued function and $\gamma : Z \rightarrow \mathbb{R}^{d \times m}$ a smooth matrix valued function (in this course we will take $Z = \mathbb{T}^d, \mathbb{R}^d$ or $\mathbb{R}^l \oplus \mathbb{T}^{d-l}$).

Consider the SDE

$$\frac{dz}{dt} = h(z) + \gamma(z) \frac{dW}{dt}, \quad z(0) = z_0. \quad (27)$$

We think of the term $\frac{dW}{dt}$ as representing Gaussian white noise: a mean-zero Gaussian process with correlation $\delta(t - s)I$.

The function $h$ in (27) is sometimes referred to as the *drift* and $\gamma$ as the *diffusion coefficient*. 
Such a process exists only as a distribution. The precise interpretation of (27) is as an integral equation for $z(t) \in C(\mathbb{R}^+, \mathcal{Z})$:

$$z(t) = z_0 + \int_0^t h(z(s)) ds + \int_0^t \gamma(z(s)) dW(s).$$  \hspace{1cm} (28)

In order to make sense of this equation we need to define the stochastic integral against $W(s)$. 

The Itô Stochastic Integral

- For the rigorous analysis of stochastic differential equations it is necessary to define stochastic integrals of the form

\[ I(t) = \int_0^t f(s) \, dW(s), \]  

(29)

- where \( W(t) \) is a standard one dimensional Brownian motion. This is not straightforward because \( W(t) \) does not have bounded variation.

- In order to define the stochastic integral we assume that \( f(t) \) is a random process, adapted to the filtration \( \mathcal{F}_t \) generated by the process \( W(t) \), and such that

\[ \mathbb{E} \left( \int_0^T f(s)^2 \, ds \right) < \infty. \]
The Itô stochastic integral \( I(t) \) is defined as the \( L^2 \)–limit of the Riemann sum approximation of (29):

\[
I(t) := \lim_{K \to \infty} \sum_{k=1}^{K-1} f(t_{k-1}) (W(t_k) - W(t_{k-1})) ,
\]

(30)

where \( t_k = k\Delta t \) and \( K\Delta t = t \).

Notice that the function \( f(t) \) is evaluated at the left end of each interval \([t_{n-1}, t_n]\) in (30).

The resulting Itô stochastic integral \( I(t) \) is a.s. continuous in \( t \).

These ideas are readily generalized to the case where \( W(s) \) is a standard \( d \) dimensional Brownian motion and \( f(s) \in \mathbb{R}^{m \times d} \) for each \( s \).
The resulting integral satisfies the Itô isometry

$$\mathbb{E}|I(t)|^2 = \int_0^t \mathbb{E}|f(s)|_F^2 ds,$$

where $| \cdot |_F$ denotes the Frobenius norm $|A|_F = \sqrt{\text{tr}(A^T A)}$.

The Itô stochastic integral is a martingale:

$$\mathbb{E}I(t) = 0$$

and

$$\mathbb{E}[I(t)|\mathcal{F}_s] = I(s) \quad \forall t \geq s,$$

where $\mathcal{F}_s$ denotes the filtration generated by $W(s)$. 
Example

- Consider the Itô stochastic integral

\[ I(t) = \int_0^t f(s) \, dW(s), \]

- where \( f, W \) are scalar–valued. This is a martingale with quadratic variation

\[ \langle I \rangle_t = \int_0^t (f(s))^2 \, ds. \]

- More generally, for \( f, W \) in arbitrary finite dimensions, the integral \( I(t) \) is a martingale with quadratic variation

\[ \langle I \rangle_t = \int_0^t (f(s) \otimes f(s)) \, ds. \]
The Stratonovich Stochastic Integral

In addition to the Itô stochastic integral, we can also define the Stratonovich stochastic integral. It is defined as the $L^2$–limit of a different Riemann sum approximation of (29), namely

$$I_{\text{strat}}(t) := \lim_{K \to \infty} \sum_{k=1}^{K-1} \frac{1}{2} \left( f(t_{k-1}) + f(t_k) \right) \left( W(t_k) - W(t_{k-1}) \right),$$

where $t_k = k \Delta t$ and $K \Delta t = t$. Notice that the function $f(t)$ is evaluated at both endpoints of each interval $[t_{n-1}, t_n]$ in (32).

The multidimensional Stratonovich integral is defined in a similar way. The resulting integral is written as

$$I_{\text{strat}}(t) = \int_0^t f(s) \circ dW(s).$$
The limit in (32) gives rise to an integral which differs from the Itô integral.

The situation is more complex than that arising in the standard theory of Riemann integration for functions of bounded variation: in that case the points in \([t_{k-1}, t_k]\) where the integrand is evaluated do not effect the definition of the integral, via a limiting process.

In the case of integration against Brownian motion, which does not have bounded variation, the limits differ.

When \(f\) and \(W\) are correlated through an SDE, then a formula exists to convert between them.
The Itô and Stratonovich stochastic integrals coincide provided that the integrand is sufficiently regular then the Itô and Stratonovich stochastic integrals coincide.

Suppose that there exist $C < \infty, \varepsilon > 0$ such that

$$\mathbb{E}|f(t, \omega) - f(s, \omega)|^2 \leq C|t - s|^{1+\varepsilon},$$

for all $s, t \in [0, T]$. Then

$$\int_0^T f(t, \omega) \, dW_t = \int_0^t f(t, \omega) \circ dW_t.$$

This is more than the regularity of the solution to an SDE.

On the other hand:

$$\int_0^T W_t \, dW_t = \frac{1}{2}W_T^2 - \frac{1}{2}T, \quad \int_0^T W_t \circ dW_t = \frac{1}{2}W_T^2.$$
Itô and Stratonovich stochastic integrals (Higham 2001)

```matlab
function [itoerr, straterr] = stint(T,N)

% Ito and Stratonovich integrals of W dW

dt = T/N;

dW = sqrt(dt) * randn(1,N);  % increments
W = cumsum(dW);  % cumulative sum

ito = sum([0,W(1:end-1)].*dW)
strat = sum((0.5*([0,W(1:end-1)]+W) + ...
            0.5*sqrt(dt)*randn(1,N)).*dW)

itoerr = abs(ito - 0.5*W(end)^2-T)
straterr = abs(strat - 0.5*W(end)^2)

end
```
We can associate a second order differential operator with and SDE, the generator of the process.

Consider the Itô SDE

\[ dX_t = \mathbf{b}(X_t) \, dt + \mathbf{\sigma}(X_t) \, dW(t), \quad (33) \]

We define

\[ \Sigma(x) = \mathbf{\sigma}(x)\mathbf{\sigma}(x)^T. \quad (34) \]

The **generator** of \( X_t \) is

\[ \mathcal{L} = \mathbf{b} \cdot \nabla + \frac{1}{2} \sum_{i,j=1}^{d} \Sigma_{ij}(x) \frac{\partial^2}{\partial x_i \partial x_j}, \quad (35) \]
Using the generator we can write Itô’s formula, which enables us to calculate the rate of change in time of functions $V(t, X_t)$.

In the absence of noise the rate of change of $V$ can be written as

$$
\frac{d}{dt} V(t, X_t) = \frac{\partial V}{\partial t}(t, x_t) + \mathcal{A} V(t, x(t)),
$$

(36)

where $X_t$ is the solution of the ODE $\dot{X}_t = b(x)$ and $\mathcal{A}$ is

$$
\mathcal{A} = b(x) \cdot \nabla.
$$

(37)

For the SDE the chain rule (36) has to be modified by the addition of a term due to noise:

$$
\frac{d}{dt} V(t, X_t) = \frac{\partial V}{\partial t}(t, X_t) + \mathcal{A} V(t, X_t) + \left\langle \nabla V(t, X_t), \sigma(X_t) \frac{dW}{dt}(t) \right\rangle.
$$

(38)
The additional term in $\mathcal{L}V$ is proportional to $\Sigma$ and it arises from the lack of smoothness of Brownian motion.

The precise interpretation of the expression for the rate of change of $V$ is in integrated form:

$$V(t, X_t) = V(x(0)) + \int_0^t \frac{\partial V}{\partial s}(s, X_s) \, ds + \int_0^t \mathcal{L}V(s, X_s) \, ds$$

$$+ \int_0^t \left\langle \nabla V(s, X_s), \sigma(X_s) \, dW(s) \right\rangle .$$ (39)

The Brownian differential scales like the square root of the differential in time: $\mathbb{E}(dW_t)^2 = dt$. We can write Itô’s formula in the form

$$dV(t, x(t)) = \frac{\partial V}{\partial t} \, dt + \sum_{i,j=1}^d \frac{\partial V}{\partial x_i} \, dx_i + \frac{1}{2} \frac{\partial^2 V}{\partial x_i \partial x_j} \, dx_i dx_j,$$ (40)

We are using the convention

$$dW_i(t)dW_j(t) = \delta_{ij} \, dt, \quad dW_i(t)\,dt = 0, \quad i, j = 1, \ldots d.$$

Thus, we can think of (40) as a generalization of Leibniz’ rule (38) where second order differentials are kept.
Taking the expectation of (40) and using the fact that the stochastic integral is a martingale we obtain (we assume that \( V \) is independent of time and that the I.C. are deterministic)

\[
\mathbb{E} V(X_t) = \mathbb{E} \int_0^t \mathcal{L} V(s, X_s) \, ds.
\]

We can use Itô’s formula to calculate the expectation value of functionals of the solution on an SDE.

For a Stratonovich SDE the rules of standard calculus apply; let \( X_t \) denote the solution of the Stratonovich SDE

\[
dX_t = b(X_t) \, dt + \sigma(X_t) \circ dW_t,
\]

where \( b : \mathbb{R}^d \mapsto \mathbb{R}^d \) and \( \sigma : \mathbb{R}^d \mapsto \mathbb{R}^{d \times d} \). Then the generator of \( X_t \) is

\[
\mathcal{L} \cdot = b \cdot \nabla \cdot + \frac{1}{2} \sigma : \nabla (\sigma^T \nabla \cdot). \tag{42}
\]

Furthermore, the Newton-Leibniz chain rule applies: let \( h \in C^2(\mathbb{R}^d) \) and define \( Y_t = h(X_t) \). Then

\[
dY_t = \nabla h(X_t) \cdot (b(X_t) \, dt + \sigma(X_t) \circ dW_t).
\]

\[
\tag{43}
\]
Example (Geometric Brownian motion)

- Consider the linear Itô SDE with multiplicative noise

\[ dX_t = \lambda X_t \, dt + \sigma X_t \, dW_t, \quad X_0 = x, \]  

(44)

- The generator of \(X_t\) is

\[ \mathcal{L} = \lambda x \frac{\partial}{\partial x} + \frac{\sigma^2 x^2}{2} \frac{\partial^2}{\partial x^2}. \]

- The solution of (44) is

\[ X(t) = x \exp \left( (\lambda - \frac{\sigma^2}{2})t + \sigma W(t) \right). \]  

(45)
Example (Geometric Brownian motion contd.)

To derive this formula, we apply Itô’s formula to the function \( f(x) = \log(x) \):

\[
\begin{align*}
\frac{d}{dt} \log(X_t) &= \mathcal{L} \left( \log(X_t) \right) dt + \sigma x \partial_x \log(X_t) \, dW_t \\
&= \left( \frac{1}{X_t} \lambda X_t + \frac{\sigma^2 X_t^2}{2} \left( -1 + \frac{1}{X_t^2} \right) \right) dt + \sigma \, dW_t \\
&= \left( \lambda - \frac{\sigma^2}{2} \right) dt + \sigma \, dW_t.
\end{align*}
\]

Consequently:

\[
\log \left( \frac{X_t}{X_0} \right) = \left( \lambda - \frac{\sigma^2}{2} \right) t + \sigma W(t),
\]

from which (45) follows.
Existence and Uniqueness of solutions for SDEs

Definition

By a solution of (27) we mean an $\mathbb{R}^d$-valued stochastic process $\{X_t\}$ on $t \in [0, T]$ with the properties:

1. $X_t$ is continuous and $\mathcal{F}_t -$adapted, where the filtration is generated by the Brownian motion $W(t)$;
2. $h(X_t) \in L^1((0, T)), \gamma(X_t) \in L^2((0, T))$;
3. equation (27) holds for every $t \in [0, T]$ with probability 1.

The solution is called unique if any two solutions $X_i(t), i = 1, 2$ satisfy

$$\mathbb{P}(X_1(t) = X_2(t), \ \forall t \in [0,T]) = 1.$$
It is well known that existence and uniqueness of solutions for ODEs (i.e. when $\gamma \equiv 0$ in (27)) holds for globally Lipschitz vector fields $h(x)$.

A very similar theorem holds when $\gamma \neq 0$.

As for ODEs the conditions can be weakened, when \textit{a priori} bounds on the solution can be found.
Assume that both $h(\cdot)$ and $\gamma(\cdot)$ are globally Lipschitz on $\mathbb{R}^d$ and that $x_0$ is a random variable independent of the Brownian motion $W(t)$ with

$$
\mathbb{E}|x_0|^2 < \infty.
$$

Then the SDE (27) has a unique solution $z(t) \in C(\mathbb{R}^+; \mathbb{R}^d)$ with

$$
\mathbb{E}\left[ \int_0^T |X_t|^2 \, dt \right] < \infty \quad \forall \ T < \infty.
$$

Furthermore, the solution of the SDE is a Markov process.
Remarks

- The Stratonovich analogue of (27) is

\[ \frac{dX}{dt} = h(X) + \gamma(X) \circ \frac{dW}{dt}, \quad X(0) = X_0. \]  

(46)

- By this we mean that \( z \in C(\mathbb{R}^+, \mathbb{R}^d) \) satisfies the integral equation

\[ X(t) = X(0) + \int_0^t h(X(s)) ds + \int_0^t \gamma(X(s)) \circ dW(s). \]  

(47)

- By using definitions (30) and (32) it can be shown that \( z \) satisfying the Stratonovich SDE (46) also satisfies the Itô SDE

\[ \frac{dX}{dt} = h(X) + \frac{1}{2} \nabla \cdot (\gamma(X)\gamma(X)^T) - \frac{1}{2} \gamma(X) \nabla \cdot (\gamma(X)^T) + \gamma(X) \frac{dW}{dt}, \]  

(48a)

\[ X(0) = X_0, \]  

(48b)

- provided that \( \gamma(z) \) is differentiable.
White noise is, in most applications, an idealization of a stationary random process with short correlation time. In this context the Stratonovich interpretation of an SDE is particularly important because it often arises as the limit obtained by using smooth approximations to white noise.

On the other hand the martingale machinery which comes with the Itô integral makes it more important as a mathematical object.

It is very useful that we can convert from the Itô to the Stratonovich interpretation of the stochastic integral.

There are other interpretations of the stochastic integral, e.g. the Klimontovich stochastic integral.
Examples of SDEs

- The Ornstein-Uhlenbeck process \((\alpha, \sigma > 0)\):
  \[
  dX_t = -\alpha X_t \, dt + \sigma \, dW_t.
  \]  
  \((49)\)

- The solution is
  \[
  X_t = e^{-\alpha t} X_0 + \sigma \int_0^t e^{-\alpha (t-s)} \, dW_s.
  \]

- Geometric Brownian motion:
  \[
  dX_t = r X_t \, dt + \sigma X_t \, dW_t.
  \]  
  \((50)\)

- The solution is
  \[
  X_t = X_0 e^{\sigma W_t + (r - \frac{1}{2} \sigma^2) t}.
  \]

- Note that the solution is different in for the Itô and the Stratonovich SDEs.

- The Cox-Ingersoll-Ross SDE \((\alpha, b > 0)\):
  \[
  dX_t = \alpha (b - X_t) \, dt + \sigma \sqrt{X_t} \, dW_t.
  \]  
  \((51)\)
Examples of SDEs (contd)

- Stochastic Verhulst equation (population dynamics)

\[
dX_t = (\lambda X_t - X_t^2) \, dt + \sigma X_t \, dW_t. \tag{52}
\]

- Lotka-Volterra SDEs:

\[
dX_i(t) = X_i(t) \left( a_i + \sum_{j=1}^{d} b_{ij} X_j(t) \right) \, dt + \sigma_i X_i(t) \, dW_i(t). \tag{53}
\]

- Protein kinetics:

\[
dX_t = (\alpha - X_t + \lambda X_t(1 - X_t)) \, dt + \sigma X_t(1 - X_t) \circ dW_t. \tag{54}
\]
Examples of SDEs (contd)

- Tracer particle (turbulent diffusion)

\[ dX_t = u(X_t, t) \, dt + \sigma \, dW_t, \quad \nabla \cdot u(x, t) = 0. \quad (55) \]

- Josephson junction (pendulum with friction and noise)

\[ \ddot{\phi}_t = - \sin(\phi_t) - \gamma \dot{\phi}_t + \sqrt{2\gamma \beta^{-1}} \dot{W}_t. \quad (56) \]

- Noisy Duffing oscillator (stochastic resonance)

\[ \dddot{X}_t = -\beta X_t - \alpha X_t^3 - \gamma \dot{X}_t + A \cos(\omega t) + \sigma \dot{W}_t \quad (57) \]

- Stochastic heat equation

\[ \partial_t u = \partial_x^2 u + \partial_t W(x, t), \quad (58) \]

on \([0, 1]\) with Dirichlet boundary conditions and \(W(x, t)\) denoting an infinite dimensional Brownian motion.
Consider the SDE (80). The generator $\mathcal{L}$ and its $L^2$-adjoint are

$$\mathcal{L} = b(x)\partial_x + \frac{1}{2}\sigma^2(x)\partial_x^2,$$

and

$$\mathcal{L}^* \cdot = -\partial_x \left( -b(x) \cdot + \frac{1}{2}\partial_x (\sigma^2(x) \cdot) \right).$$

We can obtain evolution equations for the expectation of functionals of the solution of the SDE and for the transition probability density.
Theorem

(Kolmogorov) Let \( f(x) \in C_b(\mathbb{R}) \) and let

\[
    u(x,s) := \mathbb{E}(f(X_t) | X_s = x) = \int f(y)p(y,t|x) \, dy \in C^2_b(\mathbb{R}).
\]

Assume furthermore that the functions \( b(x), \Sigma(x) = \sigma^2(x) \) are continuous in \( x \). Then \( u(x,s) \in C^{2,1}(\mathbb{R} \times \mathbb{R}^+) \) and it solves the final value problem

\[
    -\frac{\partial u}{\partial s} = b(x)\frac{\partial u}{\partial x} + \frac{1}{2} \Sigma(x,s)\frac{\partial^2 u}{\partial x^2}, \quad \lim_{s \to t} u(s,x) = f(x). \tag{59}
\]
Theorem

(Kolmogorov) Assume that $p(y, t|\cdot, )$, $b(y), \Sigma(y) \in C^2(\mathbb{R} \times \mathbb{R}^+)$. Then the transition probability density of $X_t$ satisfies the equation

$$
\frac{\partial p}{\partial t} = -\frac{\partial}{\partial y} (b(y)p) + \frac{1}{2} \frac{\partial^2}{\partial y^2} (\Sigma(y)p), \quad p(0, y|x) = \delta(x - y). 
$$

(60)
Assume that initial distribution of $X_t$ is $\rho_0(x)$. Define

$$p(y, t) := \int p(y, t|x) \rho_0(x) \, dx.$$ 

We multiply the forward Kolmogorov equation (60) by $\rho_0(x)$ and integrate with respect to $x$ to obtain the equation

$$\frac{\partial p(y, t)}{\partial t} = -\frac{\partial}{\partial y} (b(y, t)p(y, t)) + \frac{1}{2} \frac{\partial^2}{\partial y^2} (\Sigma(y)p(t, y)), \quad (61)$$

together with the initial condition

$$p(y, 0) = \rho_0(y). \quad (62)$$
Numerical Solution of SDEs

- We consider the scalar Itô SDE

\[ dX_t = b(X_t) \, dt + \sigma(X_t) \, dW_t, \quad X_0 = x, \quad (63) \]

- posed on the interval \([0, T]\). The initial condition can be either deterministic or random.

- Our goal is to obtain an approximate solution to this equation to compute quantities of the form \(\mathbb{E}f(X_t)\), where \(\mathbb{E}\) denotes the expectation with respect to the law of the process \(X_t\).

- The simplest numerical method is the Euler-Marayama method that is the analogue of the explicit Euler method for ODEs.
We partition the interval $[0, T]$ into $N$ equal subintervals of size $\Delta t$ with $N \Delta t = T$.

We set

$$ X_j := X(j \Delta t), \quad j = 0, 1, \ldots N. $$

The *Euler-Marayama method* for the SDE (63) is

$$ X_j = X_{j-1} + b(X_{j-1}) \Delta t + \sigma(X_{j-1}) \Delta W_j, \quad j = 1, \ldots N, \quad (64) $$

where $\Delta W_j = W_j - W_{j-1}$ denotes the $j$th Brownian increment.

We can write $\Delta W_j = \xi_j \sqrt{\Delta t}$ with $\xi_j \sim \mathcal{N}(0, 1)$ i.i.d.

The Euler-Marayama scheme can be written as

$$ X_j = X_{j-1} + b(X_{j-1}) \Delta t + \sigma(X_{j-1}) \sqrt{\Delta t} \xi_j, \quad j = 1, \ldots N. \quad (65) $$
Figure: Five sample paths of the Ornstein-Uhlenbeck process.
We solve the Ornstein–Uhlebeck SDE in one dimension using the EM scheme.

\[dX_t = -\alpha X_t \, dt + \sqrt{2\sigma} \, dW_t.\] (66)

We solve (66) for \(\alpha = 1\), \(\sigma = \frac{1}{2}\) for \(t \in [0, 10]\) with \(\Delta t = 0.0098\) and initial conditions \(X_0 \sim 2U(0, 1)\), where \(U(0, 1)\) denotes the uniform distribution in the interval \((0, 1)\).

In Figure 7 we present five sample paths of the OU process.

In Figure 8 we plot the first two moments of the Euler-Marayama approximation of the OU process. We compare against the theoretical solution.
a. $\mathbb{E}X_t$

b. $\mathbb{E}X_t^2$

**Figure:** First and second moments of the Ornstein-Uhlenebeck process using the Euler-Marayama method.
We use the analytical solution to investigate the error of the Euler-Marayama scheme, as a function of $\Delta t$.

We generate 1,000 paths using the EM scheme as well as the exact solution, using the same Brownian paths.

Since the noise in the OU SDE is additive, we expect that the **strong order of convergence** is 1.

This is verified by our numerical experiments, see Figure 9.
Figure: Strong order of convergence for the Euler-Marayama method for the Ornstein-Uhlenbeck process. The linear equation with slope 1 is plotted for comparison.
Consider the motion of a Brownian particle in a bistable potential:

\[ dX_t = -V'(X_t) \, dt + \sqrt{2\beta^{-1}} \, dW_t, \]  \hspace{1cm} (67)

with

\[ V(x) = \frac{x^4}{4} - \frac{x^2}{2}. \]  \hspace{1cm} (68)

The deterministic dynamical system has two stable equilibria.

For weak noise strengths the solution of the SDE spends most time oscillating around the minima of the potential.

In Figure 10 we present a sample path of the SDE with \( \beta = 10 \), obtained using the EM algorithm.
**Figure:** Sample path of the SDE (67).
Figure: Second moment of the bistable SDE (67).
Another standard numerical method for solving stochastic differential equations is the Milstein’s method

\[ X_j = X_{j-1} + b(X_{j-1}) \Delta t + \sigma(X_{j-1}) \xi_j \sqrt{\Delta t} + \frac{1}{2} \sigma(X_{j-1}) \sigma'(X_{j-1}) \Delta t \left( \xi_j^2 - 1 \right) \]  

(69)

Notice that there is an additional term, in comparison to the EM scheme.

The Milstein scheme converges strongly with order 1.
The drift term is discretized in the same way in the Euler and Milstein schemes.

On the other hand, the Milstein scheme has an additional term, which is related to the approximation of the stochastic term in (63).

We can derive the Milstein scheme as follows. First, we write an increment of the solution to the SDE in the form

$$ X_{j+1} = X_j + \int_{j\Delta t}^{(j+1)\Delta t} b(X_s) \, ds + \int_{j\Delta t}^{(j+1)\Delta t} \sigma(X_s) \, dW_s. $$  \hspace{1cm} (70)

We apply Itô’s formula to the drift and diffusion coefficients to obtain

$$ b(X_s) = b(X_j) + \int_{j\Delta t}^{s} (\mathcal{L} b)(X_\ell) \, d\ell + \int_{j\Delta t}^{s} (b' \sigma)(X_\ell) \, dW_\ell $$

and

$$ \sigma(X_s) = \sigma(X_j) + \int_{j\Delta t}^{s} (\mathcal{L} \sigma)(X_\ell) \, d\ell + \int_{j\Delta t}^{s} (\sigma' \sigma)(X_\ell) \, dW_\ell, $$

for $ s \in [j\Delta t, (j + 1)\Delta t] $, where $ \mathcal{L} $ denotes the generator of the process $ X_t $. 

Pavliotis (IC)

CompStochProc

143 / 298
We substitute these formulas into (70) to obtain, with $\xi_j \sim \mathcal{N}(0, 1)$,

$$
X_{j+1} = X_j + b(X_j) \Delta t + \sigma(X_j) \Delta W_j \\
+ \int_{j\Delta t}^{(j+1)\Delta t} \int_{j\Delta t}^s (\mathcal{L} b)(X_\ell) \, d\ell \, ds + \int_{j\Delta t}^{(j+1)\Delta t} \int_{j\Delta t}^s (b' \sigma)(X_\ell) \, dW_\ell \, ds \\
+ \int_{j\Delta t}^{(j+1)\Delta t} \int_{j\Delta t}^s (\mathcal{L} \sigma)(X_\ell) \, d\ell \, dW_s + \int_{j\Delta t}^{(j+1)\Delta t} \int_{j\Delta t}^s (\sigma' \sigma)(X_\ell) \, dW_\ell \, dW_s \\
= X_j + b(X_j) \Delta t + \sigma(X_j) \Delta W_j + (\sigma' \sigma)(X_j) \int_{j\Delta t}^{(j+1)\Delta t} \int_{j\Delta t}^s dW_\ell \, dW_s + O(\Delta t^{\alpha+\beta/2}) \\
\approx X_j + b(X_j) \Delta t + \sigma(X_j) \Delta W_j + \frac{1}{2} (\sigma' \sigma)(X_j) (\Delta W_j^2 - \Delta t) \\
= X_j + b(X_j) \Delta t + \sigma(X_j) \Delta W_j + \frac{1}{2} \Delta t (\sigma' \sigma)(X_j) (\xi_j^2 - 1),
$$

In the above, we have used the fact that $(\Delta t)^{\alpha} (\Delta W_j)^\beta = O((\Delta t)^{\alpha+\beta/2})$ and that, in one dimension,

$$
\int_{j\Delta t}^{(j+1)\Delta t} \int_{j\Delta t}^s dW_\ell \, dW_s = \frac{1}{2} (\Delta W_j^2 - \Delta t).
\tag{71}
$$
We will say that a numerical scheme has **strong order of convergence** $\alpha$ if there exists a positive constant $C$ such that

$$
\mathbb{E}|\hat{X}_j - X(j\Delta t)| \leq C\Delta t^\alpha,
$$

(72)

for any $j = 1, \ldots, N$ and for $\Delta t$ sufficiently small.

- We usually evaluate (72) at $t = T$.
- We will show that the Euler-Marayama method has strong order of convergence $\frac{1}{2}$.
- The Milstein scheme has strong order 1.
- The strong order of convergence for the Euler-Marayama method becomes 1 when the noise is additive.
- There are applications such as filtering and statistical inference where strong convergence of the numerical solution of the SDE is needed.
Quite often we are interested in the calculation of statistical quantities of solutions to SDEs, such as moments, rather than in pathwise properties.

For such a purpose, the strong convergence (72) is more than what we actually need.

We will say that a numerical method has weak order of convergence $\beta$ provided that there exists a positive constant $C$ such that for all $f \in C^\ell_p(\mathbb{R})$ (the class of $\ell$ times continuously differentiable functions which, together with their derivatives up to and including order $\ell$ have at most polynomial growth)

$$\left| \mathbb{E} f(\hat{X}_j) - \mathbb{E} f(X(j\Delta t)) \right| \leq C \Delta t^\beta,$$

for any $j = 1, \ldots, N$ and for $\Delta t$ sufficiently small.

Both the Euler-Marayama and the Milstein schemes have weak order of convergence 1.

The Milstein scheme requires more function evaluations.
For the calculation of the expectation in (72) and (73) we need to use a Monte Carlo method for calculating the expectation.

We generate $N$ sample paths of the exact solution of the SDE and of the numerical discretization that are driven by the same noise:

$$
\varepsilon = \frac{1}{N} \sum_{k=1}^{N} |\hat{X}_T^k - X_T^k|.
$$

$\varepsilon$ is a random variable for which we can prove a central limit theorem.

In addition to the discretization error we also have the Monte Carlo error.

We can perform similar numerical experiments in order to study the weak order of convergence. In this case the exact and the approximate solutions do not need to be driven by the same noise.
To estimate the strong order of convergence of a numerical scheme we plot the error as a function of step size in log-log and use least squares fitting:

$$\log(\text{err}) = \log C + \gamma \log \Delta t.$$ 

The weak convergence error consists of two steps, the **discretization error** and the **statistical error**, due to the MC approximation.

We estimate ($\hat{X}_T$ denotes the solution of the numerical scheme):

$$\varepsilon = \left| \mathbb{E}f(X_T) - \frac{1}{M} \sum_{m=1}^{M} f(\hat{X}_{N\Delta t}) \right|$$

$$\leq \left| \mathbb{E}f(X_T) - \mathbb{E}f(\hat{X}_T) \right| + \left| \mathbb{E}f(\hat{X}_T) - \frac{1}{M} \sum_{m=1}^{M} f(X_{N\Delta t}) \right|$$

$$\leq CN^{-\gamma} + CM^{-1/2}.$$ 

Fixed computational resources: optimize over $N, M$. 
The EM or Milstein schemes provide us with the solution of the SDE only at the discretization times.

The values at intermediate instants can be estimated using interpolation:

- Constant interpolation:
  \[
  \hat{X}(t) = X_{j_t}, \quad j_t = \max\{j = 0, 1, 2 \ldots, N : j\Delta t \leq t\},
  \]

- Linear interpolation:
  \[
  \hat{X}(t) = \hat{X}_{j_t} + \frac{t - \tau_{j_t}}{\tau_{j_t+1} - \tau_{j_t}}(\hat{X}_{j_t+1} - \hat{X}_{j_t}),
  \]

where \(\tau_j = j\Delta t\) and where we have used \(\hat{X}(t)\) to denote the interpolated numerical solution of the SDE.
Consider an approximate Brownian motion $W^h(t)$ constructed as a Gaussian random walk at $t_n = nh$, $h = T/N$, $T = 1$ and by linear interpolation in between. Then

$$\mathbb{E} \int_0^1 |W^h(t) - W(t)| \, dt = \frac{c_1}{N^{1/2}},$$

where $c_1 = \sqrt{\pi/32}$. Furthermore

$$\lim_{N \to +\infty} \sqrt{\frac{N}{\log N}} \mathbb{E} \sup_{0 \leq t \leq 1} |W^h(t) - W(t)| = c_2,$$

for some constant $c_2 \in (0, +\infty)$.

We can obtain similar results for diffusion processes.

It is possible to simulate exactly a diffusion process: **Exact simulation of diffusions** A. Beskos, G.O. Roberts - The Annals of Applied Probability, 2005
Consider an SDE in arbitrary dimensions:

$$dX_t = b(X_t) \, dt + \sigma(X_t) \, dW_t,$$  \hspace{1cm} (74)

where $X_t : [0, T] \mapsto \mathbb{R}^d$, $b \in C^\infty(\mathbb{R}^d)$, $\sigma \in C^\infty(\mathbb{R}^{d \times d})$, $W_t$ standard $d$-dimensional Brownian motion.

The EM scheme is (we use the notation $b^i_n = b^i(X_n)$ for the $i$th component of the vector $b$ etc.)

$$X^i_{n+1} = X^i_n + b^i_n \Delta t + \sum_{j=1}^{d} \sigma^{ij} \Delta W^j_n.$$  \hspace{1cm} (75)

It is straightforward to implement this scheme.
For the Milstein scheme we have:

\[
X_{n+1}^i = X_n^i + b_n^i \Delta t + \sum_{j=1}^{d} \sigma_{ij}^i \Delta W_n^j + \sum_{j_1,j_2=1,\ell=1}^{d} \sigma_{n}^{j\ell} \frac{\partial}{\partial x_\ell} \sigma_{1j2} I_{j_1j_2}, \tag{76}
\]

where \( I_{j_1j_2} \) denotes a double Itô integral:

\[
I_{j_1j_2} = \int_{t_n}^{t_{n+1}} \int_{t_n}^{s_1} dW_{s_2}^j dW_{s_1}^j. \tag{77}
\]

The diagonal terms \( I_{j_1j_2} \) are given by the 1d formula. The off-diagonal elements might be difficult/expensive to simulate.

We can obtain higher order schemes using the **Stochastic Taylor expansion**.

Higher order schemes require the evaluation of a large number of derivatives of the drift and diffusion coefficients.
We can exploit the structure of the SDE to simplify the calculation of the Milstein correction to the EM scheme.

Consider a nonlinear oscillator with friction and noise (example: the noisy Duffing-Van der Pol oscillator):

\[
\ddot{X}_t = b(X_t) - \gamma(X_t)\dot{X}_t + \sum_{j=1}^{d} \sigma^j(X_t) \dot{W}^j. \tag{78}
\]

Write as a first order system:

\[
dX_1^1 = X_2^2 \, dt,
\]
\[
dX_2^2 = (b(X_1^1) - \gamma(X_1^1)X_2^2) \, dt + \sum_{j=1}^{d} \sigma^j(X_t) \, dW^j.
\]

The Milstein discretization for this SDE is

\[
Y_{n+1}^1 = Y_1^n + Y_2^n \Delta t,
\]
\[
Y_{n+1}^2 = Y_2^n - (b(Y_1^n) - \gamma(Y_1^n)Y_2^n) \Delta t + \sum_{j=1}^{d} \sigma^j(Y_n) \, dW^j.
\]
No Milstein correction appears (double Itô integrals vanish). This scheme has strong order of convergence 1.

Solve the first equation for $Y_n^2$ and substitute to the second to obtain a **multistep scheme** for $Y_n^1$:

$$Y_{n+2}^1 = (2 - \gamma(Y_n^1)\Delta t)Y_{n+1}^1 - (1 - \gamma(Y_n^1)\Delta t)Y_n^1 + b(Y_n^1)(\Delta t)^2 + \sum_{j=1}^{d} \sigma^j(Y_n^1)\Delta W_n^j \Delta t.$$  \hspace{1cm} (79)

This is the Lepingle-Ribemont scheme. The first equation in the Milstein scheme can be used as a starting routine and for calculating $Y_n^2$. 


Population dynamics

\[ dX_t = rX_t(K - X_t) \, dt + \beta X_t \, dW_t, \quad X(0) = X_0. \]

Motion in a periodic incompressible velocity field subject to molecular diffusion

\[ dX_t = v(X_t) \, dt + \sqrt{2\kappa} \, dW_t, \quad \nabla \cdot v(x) = 0. \]
% compute effective diffusivities for particles moving ... in periodic velocity
% fields using monte carlo. the velocity field is a ...
Taylor-Green flow
% dt: time step N: number of steps M: number of paths
%
% D: molecular diffusion
%
function [kxx, kyy, tint] = effdiff(N, dt, M, D)
T = dt * N;
tint = [0:dt:T];

u = zeros(M, N);
v = zeros(M, N);
uin = u(:, 1);
vin = v(:, 1);

DDD = num2str(D);
MMM = num2str(M);
ellt = num2str(dt);

% Brownian increments
dW1 = sqrt(dt) * randn(M, N);
dW2 = sqrt(dt) * randn(M,N);

for j = 1:N-1
  u(:,j+1) = u(:,j) + chsowx(u(:,j),v(:,j)) * dt + ...
            sqrt(2*D) * dW1(:,j);
  v(:,j+1) = v(:,j) + chsowy(u(:,j),v(:,j)) * dt + ...
            sqrt(2*D) * dW2(:,j);
end

v = [vin,v];
u = [uin,u];

figure
plot(u(1,:),v(1,:))
grid on

kxx = var(v)./(2*tint);
kyy = var(u)./(2*tint);

figure
plot(tint,kxx,'r','LineWidth',3)
ylabel(['var(x)/2t','Fontsize',16,'Rotation',0])
xlabel('t','Fontsize',16)
title(['var(x)/2t','D = ' DDD ', M = ' MMM ', dt = ...'])
elll

figure
plot(tint,kyy,'r','LineWidth',3)
ylabel('var(y)/2t','FontSize',16,'Rotation',0)
xlabel('t','FontSize',16)
title(['var(y)/2t ', ' D = ' DDD ', M = ' MMM ', dt = ... ' elll '])
We consider the Itô SDE

\[ dX_t = b(X_t) \, dt + \sigma(X_t) \, dW_t, \quad X_0 = x. \]  \hspace{1cm} (80)

We assume that the coefficients are smooth, globally Lipschitz continuous and satisfy the linear growth condition:

\[ |b(x) - b(y)| + |\sigma(x) - \sigma(y)| \leq C|x - y|, \]  \hspace{1cm} (81)

and

\[ |b(x)| + |\sigma(x)| \leq C(1 + |x|). \]  \hspace{1cm} (82)

We also assume that the initial condition is a random variable independent of the Brownian motion \( W_t \) with

\[ \mathbb{E}|X_0|^2 \leq C. \]

Under the above assumptions there exists a unique strong solution to (80) for \( t \in [0, T] \).
Lemma

Under the above assumptions the solution to the SDE satisfies

\[ \mathbb{E} \sup_{t \in [0,T]} |X_t|^2 \leq C(T). \] (83)

Proof.

Use the assumptions on the coefficients \( b(\cdot) \) and \( \sigma(\cdot) \) and the initial conditions, Cauchy-Schwartz and the Burkholder-Gandy-Davis and Gronwall inequalities.

The BDG inequality that we need is

\[ c_2 \mathbb{E} \int_0^t \alpha^2(s) \, ds \leq \mathbb{E} \sup_{t \in [0,T]} \left( \int_0^t \alpha(s) \, dW_s \right)^2 \leq C_2 \mathbb{E} \int_0^t \alpha^2(s) \, ds \] (84)
Let \( \{X_n^{\Delta t}\}_{n=0}^N \) denote the Euler discretisation of the SDE (80) with constant interpolation.

We partition the interval \([0, T]\): \( t_n = n \Delta t, \ n = 0, \ N, \ \Delta t = T/N \).

**Lemma**

*Under the above assumptions we have*

\[
\mathbb{E} \sup_{t \in [t_n, t_{n+1}]} |X_t - X_{t_n}|^2 \leq C \Delta t. \quad (85)
\]

**Proof.**

Use the assumptions on the coefficients, the BDG inequality and Lemma 34.
Theorem

Let \( \{X_n^{\Delta t}\}_{n=0}^N \) denote the Euler discretisation of the SDE (80) with constant interpolation. Under the above assumptions we have

\[
\mathbb{E} \sup_{t \in [0,T]} |X_t - X_t^{\Delta t}|^2 \leq C\Delta t. \tag{86}
\]

For the proof we will need the discrete Gronwall inequality in the following form: let \( \{y_n\}_{n=0}^N \) be a sequence and \( A, B > 0 \) satisfying

\[
y_0 = 0, \quad y_n \leq A + B h \sum_{j=0}^{n-1} y_j, \quad 1 \leq n \leq N, \quad h = 1/N.
\]

Then

\[
\max_{0 \leq i \leq N} y_i \leq Ae^B. \tag{87}
\]
We can write the Euler scheme in the form

\[ X_{n+1}^{\Delta t} = X_n^{\Delta t} + \int_{t_n}^{t_{n+1}} b(X_n^{\Delta t}) \, ds + \int_{t_n}^{t_{n+1}} \sigma(X_n^{\Delta t}) \, dW_s. \]

For the exact solution we have

\[ X_{n+1} = X_n + \int_{t_n}^{t_{n+1}} b(X_n) \, ds + \int_{t_n}^{t_{n+1}} \sigma(X_n) \, dW_s + \varepsilon_n, \]

where

\[ \varepsilon_n = \int_{t_n}^{t_{n+1}} (b(X_s) - b(X_n)) \, ds + \int_{t_n}^{t_{n+1}} (\sigma(X_s) - \sigma(X_n)) \, dW_s. \]
Take the difference between the Euler approximation and the exact solution (use the notation
\[ \delta X_n = X_n - X_n^{\Delta t}, \quad \delta b(X_n) = b(X_n) - b(X_n^{\Delta t}) \text{ etc.} \): 

\[ \delta X_{n+1} = \delta X_n + \int_{t_n}^{t_{n+1}} \delta b_n \, ds + \int_{t_n}^{t_{n+1}} \delta \sigma_n \, dW_s + \varepsilon_n. \]

Sum over \( n \) and use \( \delta X_0 = 0 \) to obtain

\[ \delta X_{n+1} = \sum_{n=0}^{N-1} \int_{t_n}^{t_{n+1}} \delta b_n \, ds + \sum_{n=0}^{N-1} \int_{t_n}^{t_{n+1}} \delta \sigma_n \, dW_s + R_N, \]

where

\[ R_N = \sum_{n=0}^{N-1} \varepsilon_n. \]
Use the Lipschitz continuity, Cauchy-Schwarz, BGD to obtain ($\Delta t \leq 1$)

$$
\mathbb{E}|\delta X_{n+1}|^2 \leq C(T)\Delta t \sum_{n=0}^{N-1} \mathbb{E}|\delta X_n|^2 + C\mathbb{E}|R_N|^2.
$$

Use Lemma 35 and Lipschitz continuity, Cauchy-Schwarz, BGD to obtain

$$
\mathbb{E}|R_N|^2 \leq C\Delta.
$$

Use the discrete Gronwall inequality to conclude the proof of Theorem 36.
Theorem

Let $\{X_n^{\Delta t}\}_{n=0}^N$ denote the Euler discretisation of the SDE (80) with constant interpolation and assume that $b, \sigma \in C_b^4$ and $f \in C_P^4$. Then

$$|\mathbb{E}f(X_T^{\Delta t}) - \mathbb{E}f(X_T)| \leq C \Delta t.$$  

(88)
For the proof of this theorem we will use the backward Kolmogorov equation. Let \( u(t, x) \) denote the solution of the backward Kolmogorov equation with \( u(T, x) = f(x) \). We have

\[
\mathbb{E} f(X_T^{\Delta t}) - \mathbb{E} f(X_T) = \mathbb{E}(u(T, X_T^{\Delta t})) - u(0, X_0)
\]

\[
= \sum_{i=0}^{n-1} \mathbb{E} \left[ u(t_{i+1}, X_{t_i+1}^{\Delta t}) - u(t_i, X_{t_i}^{\Delta t}) \right]
\]

\[
= \sum_{i=0}^{n-1} \mathbb{E} \int_{t_i}^{t_{i+1}} \left[ \partial_t u(s, X_s^{\Delta t}) + \mathcal{L}(X_{t_i}^{\Delta t})u(s, X_s^{\Delta t}) \right] ds
\]

\[
= \sum_{i=0}^{n-1} \mathbb{E} \int_{t_i}^{t_{i+1}} \left[ (\partial_t + \mathcal{L}(X_{t_i}^{\Delta t})) (u(s, X_s^{\Delta t}) - (u(t_i, X_{t_i}^{\Delta t})) \right] ds
\]

We use Itô’s formula:

\[
\mathbb{E} \left| \partial_t u(s, X_s^{\Delta t}) - \partial_t u(t_i, X_{t_i}^{\Delta t}) \right| = \mathbb{E} |\chi(X_{t_i}^{\pi})| O(\Delta t).
\]
when we are interested in weak approximation, we can replace the Brownian increments by different random variables that might be easier to simulate and have the right statistics.

We can use, for example, the two-point distributed random variable \( \sqrt{\Delta t} \zeta_n \) where \( \zeta_n \) are i.i.d. with

\[
\mathbb{P}(\zeta_n = \pm 1) = \frac{1}{2}. \tag{89}
\]

For higher order approximations we can use the 3-point distributed random variable

\[
\mathbb{P}(\zeta_n = \pm \sqrt{3}) = \frac{1}{6}, \quad \mathbb{P}(\zeta_n = 0) = \frac{2}{3}. \tag{90}
\]

This is useful when considering weak convergence for fully implicit schemes.
Itô’s formula can be used to obtain a probabilistic description of solutions to linear partial differential equations of parabolic type. Let $X^x_t$ be a diffusion process with drift $b(\cdot)$, diffusion $\Sigma(\cdot) = \sigma\sigma^T(\cdot)$, and generator $\mathcal{L}$ with $X^x_0 = x$, and let $f \in C^2_0(\mathbb{R}^d)$ and $V \in C(\mathbb{R}^d)$, bounded from below. Then the function

$$u(x, t) = \mathbb{E} \left( e^{-\int_0^t V(X^x_s) \, ds} f(X^x_t) \right)$$

(91)

is the solution to the initial value problem

$$\frac{\partial u}{\partial t} = \mathcal{L}u - Vu, \quad \text{ (92a)}$$

$$u(0, x) = f(x). \quad \text{ (92b)}$$
To derive this result, we introduce the variable

\[ Y_t = \exp \left( - \int_0^t V(X_s) \, ds \right). \]

We rewrite the SDE for \( X_t \) as

\begin{align*}
    dX_t^x &= b(X_t^x) \, dt + \sigma(X_t^x) \, dW_t, \quad X_0^x = x, \quad (93a) \\
    dY_t^x &= -V(X_t^x) \, dt, \quad Y_0^x = 0. \quad (93b)
\end{align*}

The process \( \{X_t^x, Y_t^x\} \) is a diffusion process with generator

\[ \mathcal{L}_{x,y} = \mathcal{L} - V(x) \frac{\partial}{\partial y}. \]

We can write

\[ \mathbb{E} \left( e^{-\int_0^t V(X_s^x) \, ds} f(X_t^x) \right) = \mathbb{E}(\phi(X_t^x, Y_t^x)), \]

where \( \phi(x, y) = f(x)e^y \). We apply Itô’s formula to this function to obtain (92).
The representation formula (91) for the solution of the initial value problem (92) is called the *Feynman–Kac formula*.

We can use the Feynman-Kac formula to develop a numerical scheme for solving parabolic PDEs based on the numerical solution of the underlying SDE.
Stability analysis of numerical schemes

- Goal: solve SDEs accurately over long time intervals.
- Constant in the strong/weak order of convergence estimates grows exponentially fast (Gronwall’s lemma).
- Stability analysis: fix $\Delta t$, study the $T \to +\infty$ limit.
- Test the stability of the scheme on the linear SDE (geometric BM)
  \[ dX_t = \lambda X_t \, dt + \sigma X_t \, dW_t, \quad X(0) = X_0, \]  
  \[ (94) \]
- where $\lambda, \mu \in \mathbb{C}$.
We will say that an SDE is **stable in mean square** if for all \( X_0 \neq 0 \) with prob. 1

\[
\lim_{t \to +\infty} \mathbb{E}|X_t|^2 = 0, \tag{95}
\]

We will say that an SDE is **asymptotically stable** if for all \( X_0 \neq 0 \) with prob. 1

\[
\mathbb{P}\left( \lim_{t \to +\infty} |X_t| = 0 \right) = 1. \tag{96}
\]

For the gBM (94) we have

\[
\lim_{t \to +\infty} \mathbb{E}|X_t|^2 = 0 \iff \text{Re}(\lambda) + \frac{1}{2} |\sigma|^2 < 0. \tag{97}
\]

and

\[
\mathbb{P}\left( \lim_{t \to +\infty} |X_t| = 0 \right) = 1 \iff \text{Re} \left( \lambda - \frac{1}{2} \sigma^2 \right) < 0. \tag{98}
\]
Lemma

The geometric Brownian motion (94) with $\lambda, \mu \in \mathbb{C}$ is stable in mean square provided that

$$\text{Re}(\lambda) + \frac{1}{2}|\sigma|^2 < 0.$$  \hfill (99)
Proof.

We apply Itô’s formula to $|X_t|^2 = (\text{Re}X_t)^2 + (\text{Im}X_t)^2$ to obtain

$$
\frac{d}{dt} |X_t|^2 = \left(2\text{Re}(\lambda) + |\sigma|^2 \right) |X_t|^2 \, dt + dM_t,
$$

where $M_t$ is a martingale. We take the expectation to obtain

$$
\frac{d}{dt} \mathbb{E}|X_t|^2 = \left(2\text{Re}(\lambda) + |\sigma|^2 \right) \mathbb{E}|X_t|^2.
$$

The solution to this equation is

$$
\mathbb{E}|X_t|^2 = \exp \left(2\text{Re}(\lambda) + |\sigma|^2 \right) \mathbb{E}|X_0|^2,
$$

from which (99) follows.
Mean square stability is more stringent than asymptotic stability.

Suppose that the parameters $\lambda$ and $\sigma$ are chosen so that gBM is mean square stable. For what values of $\Delta t$ is the EM also mean square stable?

$$\lim_{j \to +\infty} \mathbb{E}X_j^2 = 0 \iff (1 + \Delta t + \lambda)^2 + \Delta t|\sigma|^2 < 1.$$  \hspace{1cm} (100)

Similar results can be obtained for the Milstein scheme.
Definition

Let $X_t^{\Delta t}$ denote a time-discrete approximation of the SDE

$$dX_t = b(X_t)\, dt + \sigma(X_t)\, dW_t$$  \hspace{1cm} (101)$$

with step size $\Delta t$ starting at $X_0^{\Delta t}$ at $t = 0$ and let $\overline{X}_t^{\Delta t}$ denote the approximation starting at $\overline{X}_0^{\Delta t}$. We will say that $X_t^{\Delta t}$ is **stochastically numerically stable** for the SDE (101) if for any finite interval $[0, T]$ \exists $\Delta t_0$ s.t. \forall $\varepsilon > 0$, $\Delta t \in (0, \Delta t_0)$ we have

$$\lim_{|X_0^{\Delta t} - \overline{X}_0^{\Delta t}| \rightarrow 0} \sup_{t \in [0, T]} \mathbb{P} \left( |X_{nt}^{\Delta t} - \overline{X}_{nt}^{\Delta t}| \geq \varepsilon \right) = 0.$$  \hspace{1cm} (102)$$

We will say that a numerical scheme is stochastically numerically stable if it satisfies (103) for all SDEs for which it is convergent.
Asymptotic stochastic stability: Replace (103) with

\[
\lim_{X_0^\Delta t \rightarrow 0} \lim_{T \rightarrow +\infty} \mathbb{P}\left(\left|X_{n+1}^\Delta t - X_{nT}^\Delta t\right| \geq \varepsilon\right) = 0. \tag{103}
\]

Consider the linear SDE

\[dX_t = \lambda X_t \, dt + dW_t.\]

It is mean square stable for \(\text{Re}(\lambda) < 0\).

Consider numerical schemes that can be written in the form

\[X_{n+1}^\Delta t = X_n^\Delta t G(\lambda \Delta t) + Z_n^\Delta t.\]

The **region of absolute stability** of the scheme is the set of \(\lambda \Delta t\) for which

\[\text{Re}(\lambda) < 1 \quad \text{and} \quad |G(\lambda \Delta t)| < 1.\]

If the region of absolute stability is the left half complex plane the scheme is **A-stable**.
• Just as with ODEs, implicit schemes have better stability properties than explicit schemes.
• The implementation of an implicit scheme requires the solution of an additional algebraic equation at each time step, which can usually be done using the Newton-Raphson algorithm.
• Treating the noise involves reciprocals of Gaussian random variables which do not have finite absolute moments.
• We will treat the drift implicitly and the noise explicitly.
• The implicit Euler scheme is

\[ X_{n+1} = X_n + b(X_{n+1}) \Delta t + \sigma(X_n) \Delta W_n. \]  

(104)

• If we are only interested in weak convergence we can replace \( \sqrt{\Delta t} N(0, 1) \) by random variables that are accurate in the weak sense and are easier to take their reciprocals.
Family of implicit Euler schemes (stochastic theta method, STM):

\[ X_{n+1} = X_n + (\theta b(X_{n+1}) + (1 - \theta)b(X_n)) \Delta t + \sigma(X_n) \Delta W_n, \quad (105) \]

for \( \theta \in [0, 1] \).

Family of implicit Milstein schemes:

\[
X_{n+1} = X_n + (\theta b(X_{n+1}) + (1 - \theta)b(X_n)) \Delta t \\
+ \sigma(X_n) \Delta W_n + \frac{1}{2} (\sigma \sigma')(X_n) ((\Delta W_n)^2 - \Delta t). \quad (107)
\]

For both schemes we have

\[
G(\lambda \Delta t) = \frac{1 + (1 - \theta)\lambda \Delta t}{1 - \theta \lambda \Delta t}.
\]
Test problem: geometric Brownian motion (linear SDE with multiplicative noise), eqn. (94).

For what stepsizes $\Delta t$ does the stochastic $\theta$ method share the stability properties of the test problem?

- Use the concept of mean square stability.
- For the SDE the mean square stability depends on the parameters $\lambda, \sigma$.
- For the numerical scheme it depends on the model parameters $\lambda, \sigma$ and on the numerical scheme parameters $\Delta t, \theta$.
- Our goal is to find for what values of $\Delta t, \theta, \lambda, \sigma$ the stochastic theta method is also mean square stable.
The STM for the geometric Brownian motion becomes

\[ X_{n+1} = X_n + (1 - \theta) \lambda hX_n + \theta \lambda hX_{n+1} + \sigma \sqrt{h} \xi_n X_n, \]  
(108)

with \( h := \Delta t \). We can rewrite this equation as

\[ X_{n+1} = \left( a + b \xi_n \right) X_n, \]  
(109)

with

\[ a = \frac{1 + (1 - \theta)h\lambda}{1 - \theta h\lambda}, \quad b = \frac{\sqrt{h}\sigma}{1 - \theta h\lambda}. \]  
(110)

This is a homogeneous Markov chain with an uncountable state space.

To study the stability properties of the STM we need to study the long time behavior of the Markov Chain (109)
**Definition**

The numerical scheme (Markov Chain) (109) is mean square stable iff

\[
\lim_{n \to +\infty} \mathbb{E}|X_n|^2 = 0.
\]  

(111)

**Theorem**

The MC (109) is mean square stable if and only if

\[
|a|^2 + |b|^2 < 1.
\]  

(112)

In particular, for the STM we have

\[
\frac{|1 + (1 - \theta)h\lambda|^2 + h|\sigma|^2}{|1 - \theta h\lambda|^2} < 1.
\]  

(113)
Define the stability regions for the gBM and for the STM:

\[
S_{SDE} := \{ \lambda, \sigma \in \mathbb{C} : \Re(\lambda) + \frac{1}{2}|\sigma|^2 < 0 \} \tag{114}
\]

and

\[
S_{STM}(\theta, h) := \left\{ \lambda, \sigma \in \mathbb{C} : \frac{|1 + (1 - \theta)h\lambda|^2 + h|\sigma^2|}{|1 - \theta h\lambda|^2} < 1 \right\}. \tag{115}
\]

**Theorem**

*For all* \( h > 0 \) *we have*

- \( S_{STM}(\theta, h) \subset S_{SDE} \) *for* \( \theta \in [0, \frac{1}{2}) \).
- \( S_{STM}(\theta, h) = S_{SDE} \) *for* \( \theta = \frac{1}{2} \).
- \( S_{STM}(\theta, h) \supset S_{SDE} \) *for* \( \theta \in (\frac{1}{2}, 1] \).
Definition
The numerical scheme is (109) (mean square) A-stable provided that whenever the test problem (94) is mean-square stable, then (109) is also mean square stable for all $\Delta t > 0$.

Corollary

*The STM is A-stable for all $\theta \in [\frac{1}{2}, 1]$.*

- If the SDE is unstable, then so is the STM for all $h$.
- If the SDE is stable, then so is the STM for sufficiently small $h$.
- In this case, the resulting stepsize restriction can be arbitrarily severe.
We consider SDEs in $\mathbb{R}^d$

$$
dX_t = b(X_t) \, dt + \sigma(X_t) \, dW_t. \tag{116}
$$

We assume that $X_t$ is ergodic with respect to the invariant measure $\pi(dx) = \rho_\infty(x) \, dx$.

Assuming sufficient regularity, the invariant density $\rho_\infty(x)$ is the unique normalizable solution of the stationary Fokker-Planck equation

$$
\mathcal{L}^* \rho_\infty = 0. \tag{117}
$$

In many applications we need to be able to calculate expectations with respect to the stationary distribution

$$
\mathbb{E}_\pi f(x) = \int_{\mathbb{R}^d} f(x) \rho_\infty(x) \, dx. \tag{118}
$$
In high dimensions it is computationally prohibitive to solve the stationary Fokker-Planck equation.

We also need to calculate the high dimensional integral (118).

For ergodic SDEs time averages equal phase-space averages:

\[
\lim_{T \to +\infty} \frac{1}{T} \int_0^T f(X_s) \, ds = \mathbb{E}_{\pi} f(x). \tag{119}
\]

We can solve numerically the SDE (116) and then calculate the time average in (119) to calculate \( \mathbb{E}_{\pi} f(x) \).

Let \( X_n^h = X^h(nh), h = \Delta t \) denote the solution of the numerical scheme. We need to estimate the difference

\[
\left| \frac{1}{N} \sum_{n=1}^{N} f(X_n^h) - \int_{\mathbb{R}^d} f(x) \rho_\infty(x) \, dx \right| \tag{120}
\]

as a function of \( h \) and \( N \).
Let \( \{X_n^{\Delta t}\}_{n=0}^{nT} \) denote a (long) numerically calculated trajectory. We have

\[
F_T := \int_0^T f(X_s) \, ds \approx \frac{1}{nT} \sum_{n=0}^{nT-1} f(X_n^{\Delta t}) := F_T^{\Delta t}.
\]

Define

\[
F^{\Delta t} = \lim_{T \to +\infty} F_T^{\Delta t}.
\]

We will say that a time discrete approximation \( X_n^{\Delta t} \) converges **with respect to the ergodic criterion** with order \( \beta > 0 \) to the ergodic diffusion process \( X_t \) as \( \Delta t \to 0 \) provided that for every \( f \in C_0^\infty(\mathbb{R}^d, \mathbb{R}) \) there exist a positive constant \( C_f \) independent of \( \Delta t \) and a \( \Delta t_0 \) such that

\[
|F^{\Delta t} - \mathbb{E}_\mu f(x)| \leq C_f(\Delta t)^\beta \tag{121}
\]
This is an extension of the weak convergence criterion to the infinite time horizon $T = \infty$.

The explicit Euler scheme converges with respect to the ergodic criterion with order $\beta = 1.0$.

We expect that most of the numerical schemes with weak order $\beta$ also converge with respect to the ergodic criterion with the same order $\beta$, under the assumptions of the Hasminski criterion.
We need to address the following issues.

1. Does the numerical scheme have an invariant measure (density) $\rho^h_\infty$? How quickly does it converge to it?
2. How close is $\rho^h_\infty$ to $\rho_\infty$?
3. How close is the time averaging estimator to the stationary measure of the SDE?

The standard weak convergence results (e.g. for the explicit Euler method) are valid only over finite time intervals. In order to be able to control the difference between the long time average of the numerical scheme and $\mathbb{E}_\pi f$ we need estimates on the numerical scheme that are uniform in time.

For this we need to use the structural property of (geometric) ergodicity of the SDE.
A very important concept in the study of limit theorems for stochastic processes is that of **ergodicity**.

This concept, in the context of Markov processes, provides us with information on the long–time behavior of a Markov semigroup.

**Definition**

A Markov process is called **ergodic** if the equation

\[ P_t g = g, \quad g \in C_b(E) \quad \forall t \geq 0 \]

has only constant solutions.

Roughly speaking, ergodicity corresponds to the case where the semigroup \( P_t \) is such that \( P_t - I \) has only constants in its null space, or, equivalently, to the case where the generator \( \mathcal{L} \) has only constants in its null space. This follows from the definition of the generator of a Markov process.
Under some additional compactness assumptions, an ergodic Markov process has an *invariant measure* $\mu$ with the property that, in the case $T = \mathbb{R}^+$,

$$\lim_{t \to +\infty} \frac{1}{t} \int_0^t g(X_s) \, ds = \mathbb{E} g(x),$$

where $\mathbb{E}$ denotes the expectation with respect to $\mu$.

This is a physicist’s definition of an ergodic process: *time averages equal phase space averages*.

Using the adjoint semigroup we can define an invariant measure as the solution of the equation

$$P_t^* \mu = \mu.$$

If this measure is unique, then the Markov process is ergodic.
Using this, we can obtain an equation for the invariant measure in terms of the adjoint of the generator $L^*$, which is the generator of the semigroup $P_t^*$. Indeed, from the definition of the generator of a semigroup and the definition of an invariant measure, we conclude that a measure $\mu$ is invariant if and only if

$$\mathcal{L}^* \mu = 0$$

in some appropriate generalized sense ($\langle \mathcal{L}^* \mu, f \rangle = 0$ for every bounded measurable function).

Assume that $\mu(dx) = \rho(x) \, dx$. Then the invariant density satisfies the stationary Fokker-Planck equation

$$\mathcal{L}^* \rho = 0.$$  

The invariant measure (distribution) governs the long-time dynamics of the Markov process.
If \( X_0 \) is distributed according to \( \mu \), then so is \( X_t \) for all \( t > 0 \). The resulting stochastic process, with \( X_0 \) distributed in this way, is **stationary**.

In this case the transition probability density (the solution of the Fokker-Planck equation) is independent of time: \( \rho(x, t) = \rho(x) \).

Consequently, the statistics of the Markov process is independent of time.
Example

The one dimensional Brownian motion is not an ergodic process: The null space of the generator $L = \frac{1}{2} \frac{d^2}{dx^2}$ on $\mathbb{R}$ is not one dimensional!

Example

Consider a one-dimensional Brownian motion on $[0, 1]$, with periodic boundary conditions. The generator of this Markov process $L$ is the differential operator $L = \frac{1}{2} \frac{d^2}{dx^2}$, equipped with periodic boundary conditions on $[0, 1]$. This operator is self-adjoint. The null space of both $L$ and $L^*$ comprises constant functions on $[0, 1]$. Both the backward Kolmogorov and the Fokker-Planck equation reduce to the heat equation

$$\frac{\partial \rho}{\partial t} = \frac{1}{2} \frac{\partial^2 \rho}{\partial x^2}$$

with periodic boundary conditions in $[0, 1]$. Fourier analysis shows that the solution converges to a constant at an exponential rate.
Example

- The one dimensional **Ornstein-Uhlenbeck (OU) process** is a Markov process with generator

\[ \mathcal{L} = -\alpha x \frac{d}{dx} + D \frac{d^2}{dx^2}. \]

- The null space of $\mathcal{L}$ comprises constants in $x$. Hence, it is an ergodic Markov process. In order to calculate the invariant measure we need to solve the stationary Fokker–Planck equation:

\[ \mathcal{L}^* \rho = 0, \quad \rho \geq 0, \quad \| \rho \|_{L^1(\mathbb{R})} = 1. \tag{122} \]
Let us calculate the $L^2$-adjoint of $\mathcal{L}$. Assuming that $f, h$ decay sufficiently fast at infinity, we have:

$$\int_{\mathbb{R}} \mathcal{L} f h \, dx = \int_{\mathbb{R}} \left[ (-\alpha x \partial_x f) h + (D \partial_x^2 f) h \right] \, dx$$

$$= \int_{\mathbb{R}} \left[ f \partial_x (\alpha x h) + f (D \partial_x^2 h) \right] \, dx =: \int_{\mathbb{R}} f \mathcal{L}^* h \, dx,$$

where

$$\mathcal{L}^* h := \frac{d}{dx} (\alpha x h) + D \frac{d^2 h}{dx^2}.$$

We can calculate the invariant distribution by solving equation (122).

The invariant measure of this process is the Gaussian measure

$$\mu(dx) = \sqrt{\frac{\alpha}{2\pi D}} \exp \left( - \frac{\alpha}{2D} x^2 \right) \, dx.$$
Approximation of Invariant Measures and MCMC

Consider the SDE in $\mathbb{R}^d$:

$$dX_t = b(X_t) \, dt + \sigma(X_t) \, dW_t, \quad X_0 = x.$$  \hfill (123)

The generator of $X_t$ is

$$\mathcal{L} = b(x) \cdot \nabla + \frac{1}{2} \sum_{i,j=1}^{d} \Sigma_{ij}(x) \frac{\partial^2}{\partial x_i \partial x_j}.$$ \hfill (124)

The $L^2$-adjoint (Fokker-Planck operator) is

$$\mathcal{L}^* = \nabla \cdot \left( b(x) \cdot + \frac{1}{2} \nabla \cdot (\Sigma(x) \cdot ) \right).$$ \hfill (125)
The expectation $u(x, t) = \mathbb{E}(\phi(X_t) | X_0 = x)$ is the solution of the initial value problem for the backward Kolmogorov equation

$$\frac{\partial u}{\partial t} = Lu, \quad u(x, 0) = \phi(x).$$  \hfill (126)

Using the transition probability density $p(t, x, y)$ we can write

$$u(x, t) = \int_{\mathbb{R}^d} p(t, x, y) \phi(y) \, dy.$$

$p(t, x, y)$ is the solution of the forward Kolmogorov (Fokker-Planck) equation

$$\frac{\partial p}{\partial t} = L^*_y p, \quad p(0, x, y) = \delta(x - y).$$  \hfill (127)
• Connection between SDEs, (elliptic and parabolic) PDEs and probability measures.

• We can use SDEs in order to calculate functional integrals, the solution of elliptic and parabolic PDEs and to sample from a given probability distribution.

• We need to study the ergodic properties of the SDE (123).

• We denote by $P_t$ and $P_t^*$ the semigroups generated by $\mathcal{L}$ and $\mathcal{L}^*$, respectively:

\[
P_t = e^{\mathcal{L}t} \quad \text{and} \quad P_t^* = e^{\mathcal{L}^*t}.
\]

• $P_t$ acts on $L^\infty$ functions and $P_t^*$ on probability measures.
A probability measure $\mu$ is invariant under the dynamics (123) provided that

$$P_t^* \mu = \mu. \quad(128)$$

We will say that the dynamics $X_t$ is **ergodic** if there exists only one invariant measure. In this case

$$\lim_{T \to +\infty} \frac{1}{T} \int_0^T f(X_s) \, ds = \int_{\mathbb{R}^d} f(x) \mu(dx). \quad(129)$$

Let $\mu(dx) = p_\infty(x) \, dx$. We divide (128) by $t$ and pass to the limit $t \to 0$ to obtain the **stationary Fokker-Planck equation**

$$\mathcal{L}^* p_\infty = 0, \quad(130)$$

together with the appropriate boundary conditions.

The SDE (123) is ergodic provided that there exists a unique normalizable solution to (130).
Consider the dynamics (123) with $X_0 \sim p_0(x)$. The law of the process $X_t$ is the solution of the initial value problem for the Fokker-Planck equation

$$\frac{\partial p}{\partial t} = \mathcal{L}^* p, \quad p(0, x) = p_0(x).$$

(131)

When $X_t$ is ergodic we have that

$$\lim_{t \to +\infty} \| p(\cdot, t) - p_\infty(\cdot) \|_{L^1(\mathbb{R}^d)} = 0.$$

In applications to MCMC we need quantitative information on the rate of convergence to equilibrium.
We can study the ergodic properties of an SDE using techniques from stochastic analysis, functional analysis (functional inequalities), Lyapunov function techniques etc.

**Hasminski’s criterion:** if the drift and diffusion coefficients are smooth, the diffusion coefficient is strictly positive definite and bounded and there exists a constant \( \beta > 0 \) and a compact subset \( K \in \mathbb{R}^d \) such that

\[
\langle b(x), x \rangle \leq -\beta |x|^2
\]

for all \( x \in \mathbb{R}^d K \). Then \( X_t \) is ergodic.

Even if we can prove that an SDE is ergodic, we usually do not have an analytic formula for the invariant measure.

Example: nonequilibrium steady states.
We can consider numerical schemes that can be written in the form
\[
X^h_{n+1} = X_n + \hat{b}(X^h_n; h)h + \hat{\sigma}(X^h_n, h)\sqrt{h}\xi_n,
\]
(132)
where \(\hat{b} : \mathbb{R}^d \times (0, 1) \to \mathbb{R}^d\), \(\hat{\sigma} : \mathbb{R}^d \times (0, 1) \to \mathbb{R}^{d \times m}\), and \(\eta_n\) is a collection of iid real-valued r.v. with
\[
\mathbb{E}\eta_{n,i} = \mathbb{E}\eta_{n,i} = 0, \quad \mathbb{E}\eta_{n,i} = 1, \quad \mathbb{E}\eta_{n,i}^{2r} < +\infty,
\]
for \(r\) sufficiently large.
For example, we can have \(\eta_{n,i} \sim \mathcal{N}(0, 1)\) iid or iid two-point r.v (89).
The convergence of the long time average of the solution to the SDE to its invariant measure can be proved by studying the solution of the Poisson equation

\[- \mathcal{L} \phi = f - \mathbb{E}_\pi f, \quad \mathbb{E}_\pi \phi = 0. \tag{133}\]

We apply Itô’s formula to \( \phi \):

\[
d\phi(X_t) = \mathcal{L} \phi(X_t) \, dt + dM_t,
\]

\[
= -(f(X_t) - \mathbb{E}_\pi f) \, dt + dM_t.
\]

where \( M_t = \int_0^t \nabla \phi(X_t) \sigma(X_t) \, dW_t \).

\[
\frac{1}{T} \int_0^T f(X_t) \, dt - \mathbb{E}_\pi f = -\frac{1}{T} \left( \phi(X_T) - \phi(X_0) \right) + \frac{1}{T} M_T.
\]
Assume that solutions of the Poisson equation (133) are bounded (for example, if the diffusion is uniformly elliptic and the state space is compact, e.g. the unit torus $\mathbb{T}^d$). Then:

$$\lim_{T \to +\infty} \mathbb{E} \left( \frac{1}{T} \left( \phi(X_t) - \phi(X_0) \right) \right)^2 = 0,$$

and

$$\lim_{T \to +\infty} \mathbb{E} \frac{1}{T^2} M_t^2 = \lim_{T \to +\infty} \frac{1}{T} \mathbb{E} \langle M_t \rangle = 0,$$

using the law of large numbers and the central limit theorems for martingales.

Consequently:

$$\mathbb{E} \left( \frac{1}{T} \int_0^T f(X_t) \, dt - \mathbb{E}_\pi f \right)^2 \leq \frac{C}{T}.$$
This is a quantitative version of the mean ergodic theorem. We can also prove an almost sure ergodic theorem (strong law of large numbers)

\[ \lim_{T \to +\infty} \frac{1}{T} \int_0^T f(X_t) \, dt = \mathbb{E}_\pi f \quad a.s. \]

The long time average of \( f(X_t) \) is an estimator for the integral \( \mathbb{E}_\pi f \). Its bias and variance are:

\[
\text{Bias} \left( \frac{1}{T} \int_0^T f(X_t) \, dt \right) := \mathbb{E}_\pi \left( \frac{1}{T} \int_0^T f(X_t) \, dt - \mathbb{E}_\pi f \right) = O \left( \frac{1}{T} \right) \tag{134}
\]

and

\[
\text{Var}_\pi \left( \frac{1}{T} \int_0^T f(X_t) \, dt \right) := \mathbb{E}_\pi \left( \frac{1}{T} \int_0^T f(X_t) \, dt - \mathbb{E}_\pi f \right)^2 = O \left( \frac{1}{T} \right). \tag{135}
\]
We have that

$$\lim_{T \to +\infty} T \text{Var}_\pi \left( \frac{1}{T} \int_0^T f(X_t) \, dt \right) = 2\langle \phi, f - \mathbb{E}_\pi f \rangle_\pi,$$

(136)

where $\phi$ denotes the solution of the Poisson equation (133) and $\langle h, g \rangle_\pi = \int hg \, \pi(dx)$ denotes the $L^2(\pi)$ inner product.

This follows from Itô's formula or, equivalently, from (setting $\bar{f} = f - \mathbb{E}_\pi f$)

$$\lim_{T \to +\infty} T \text{Var}_\pi \left( \frac{1}{T} \int_0^T f(X_t) \, dt \right) = 2 \int_0^{+\infty} \mathbb{E}_\pi (\bar{f}(X_t)\bar{f}(X_0)) \, dt \quad (137)$$

and the backward Kolmogorov equation.
We have:

\[
\text{Var}_\pi \left( \frac{1}{T} \int_0^T f(X_t) \, dt \right) = \mathbb{E}_\pi \left( \frac{1}{T} \int_0^T f(X_t) \, dt \right)^2
\]

\[
= \frac{1}{T^2} \int_0^T \int_0^T \mathbb{E}_\pi \left( \overline{f}(X_t) \overline{f}(X_s) \right) \, dt \, ds
\]

\[
=: \frac{1}{T^2} \int_0^T \int_0^T R_{\overline{f}}(t, s) \, dt \, ds
\]

\[
= \frac{2}{T^2} \int_0^T (T - s) R_{\overline{f}}(s) \, ds
\]

\[
= \frac{2}{T} \int_0^T \left( 1 - \frac{s}{T} \right) \mathbb{E}_\pi \left( \overline{f}(X_s) \overline{f}(X_0) \right) \, ds,
\]

from which (137) follows.
Let \( u(x, t) \) denote the solution of the backward Kolmogorov equation with initial condition \( f(x) \). We can formally write

\[
\mathbb{E}(f(X_t) \mid X_0 = x) = u(x, t) = e^{Lt}f(x),
\]

where \( \mathcal{L} \) denotes the generator of \( X_t \).

For the calculation of the time integral of the stationary autocorrelation function we have (\( \mathcal{X} \) denotes the state space):

\[
\int_0^{+\infty} \mathbb{E}_\pi(f(X_t)f(X_0)) \, dt = \int_{\mathcal{X}} \int_0^{+\infty} \left( e^{Lt}f(x) \right) f(x) \, dt \, \pi(dx)
\]

\[
= \int_{\mathcal{X}} \bar{f}(x) \left( \int_0^{+\infty} e^{Lt}f(x) \, dt \right) \pi(dx)
\]

\[
= \int_{\mathcal{X}} \bar{f}(x) \left( \int_0^{+\infty} e^{Lt}f(x) \, dt \right) \pi(dx)
\]

\[
= \int_{\mathcal{X}} \bar{f}(x)(-\mathcal{L})^{-1}f(x) \pi(dx)
\]

\[
= \langle \bar{f}, \phi \rangle_\pi.
\]
Consider the long time average estimator of $\mathbb{E}_\pi f$

$$\hat{f}_N = \frac{1}{N} \sum_{n=0}^{N-1} f(X^h_n), \quad (138)$$

where $\{X^h_n\}_{n=0}^{N-1}$ is the Markov chain obtained from the numerical discretization of the SDE, from example using the EM scheme.

We want to calculate the variance of this estimator by generating a long trajectory:

1. We generate a long trajectory of length $MT$ with $T = Nh$ and we split it into $M$ blocks of length $T$.
2. We evaluate the estimators $\hat{f}_{m,N}$. Assuming that $T$ is large enough and that we have fast decay of correlations then these estimators are close to being uncorrelated.
3. We compute the sampled variance

$$\hat{D} = \frac{1}{M-1} \sum_{m=1}^{M} \hat{f}^2_{m,N} - \left( \frac{1}{M} \sum_{m=1}^{M} \hat{f}_{m,N} \right)^2 \quad (139)$$
For sufficiently large $N, M$ we have the following confidence interval

$$
\mathbb{E}_\pi \hat{f}_N \in \left( \hat{f}_{NM} - c \frac{\sqrt{D}}{\sqrt{M}}, \hat{f}_{NM} + c \frac{\sqrt{D}}{\sqrt{M}} \right),
$$

(140)

with probability 95% for $c = 2$ and probability 99.7% for $c = 3$. 
MARKOV CHAIN MONTE CARLO
Goal: sample from a distribution \( \pi(x) \) that is known only up to a constant. We will sometimes write
\[
\pi(x) = \frac{1}{Z} e^{-V(x)}, \quad Z = \int_{\mathbb{R}^N} e^{-V(x)} \, dx. \tag{141}
\]

We want to calculate integrals of the form
\[
\mathbb{E}_{\pi} f := \int_{\mathbb{R}^d} f(x) \, \pi(dx).
\]

1. **MC approach:** Calculate the normalization constant
\[
Z = \int_{\mathbb{R}^d} \pi(x) \, dx
\]
and the integral
\[
\mathbb{E}_{\pi} f(x) = \int_{\mathbb{R}^d} f(x) \pi(x) \, dx
\]
using MC, together with an appropriate variance reduction technique.

2. **SDE approach:** Construct an appropriate stochastic dynamics whose invariant distribution is \( \pi(x) \).
Construct ergodic stochastic dynamics whose invariant distribution is $\pi(x)$.

There are many different dynamics whose invariant distribution is given by $\pi(x)$.

Different discretizations of the corresponding SDE can behave very differently, even fail to converge to $\pi(x)$.

Computational efficiency:
1. Choose the dynamics that converges to equilibrium as quickly as possible (bias correction).
2. Choose the dynamics that leads to the minimum asymptotic variance (variance reduction).

Consider these problems either for a class of observables or for a specific observable.
We want to calculate the integral

$$\mathbb{E}_\pi f = \int_{\mathbb{R}^d} f(x) \pi(dx),$$  \hspace{1cm} (142)

where $\pi(dx)$ is known up to the normalization constant.

We want to use a diffusion process $X_t$ that is ergodic with respect to $\pi(dx)$:

$$\lim_{T \to +\infty} \frac{1}{T} \int_0^T f(X_s) \, ds = \mathbb{E}_\pi f \quad \text{a.s.}$$  \hspace{1cm} (143)

For all $f \in L^1(\pi)$.

We will consider diffusions that satisfy a functional central limit theorem

$$\lim_{T \to +\infty} \sqrt{T} \left( \frac{1}{T} \int_0^T f(X_s) \, ds - \mathbb{E}_\pi f \right) = \mathcal{N}(0, 2\sigma_f^2),$$  \hspace{1cm} (144)

in distribution for all $f \in L^2(\pi)$.

Our goal is to choose the diffusion process so that we can speed up convergence to equilibrium and minimize the asymptotic variance.
In order for $\pi(x)$ to be the invariant distribution of $X_t$, we require that it is the (unique) solution of the stationary Fokker-Planck equation

$$\nabla \cdot (-b(x)\pi(x) + \nabla \cdot (\Sigma(x)\pi(x))) = 0,$$

(145)

together with appropriate boundary conditions (if we are in a bounded domain).

If the **detailed balance** condition

$$J_s := -b\pi + \nabla \cdot (\Sigma \pi) = 0,$$

(146)

is satisfied, then the process $X_t$ is reversible wrt $\pi(x)$. 

Pavliotis (IC)  CompStochProc  214 / 298
There are (infinitely) many reversible diffusions that can be used in order to sample from $\pi(x)$. A reversible diffusion can be written in the form

$$dX_t = \Sigma(X_t) \nabla \pi(X_t) \, dt + \nabla \cdot \Sigma(X_t) \, dt + \sqrt{2\Sigma(X_t)} \, dW_t. \tag{147}$$

The rate of convergence to equilibrium depends on the tails of the distribution $\pi(x)$ and on the choice of diffusion process $X_t$.

The standard choice is the overdamped Langevin dynamics, $\Sigma = 2I, \ b(x) = \frac{1}{2} \nabla \log \pi(x)$

$$dX_t = \frac{1}{2} \nabla \log \pi(X_t) \, dt + dW_t. \tag{148}$$
Definition

A stationary stochastic process $X_t$ is time reversible if its law is invariant under time reversal: for every $T \in (0, +\infty)$, $X_t$ and the time-reversed process $X_{T-t}$ have the same distribution.

- The processes $X_t$ and $X_{T-t}$ have the same finite dimensional distributions. Equivalently, for each $N \in \mathbb{N}^+$, a collection of times $0 = t_0 < t_1 \cdots < t_N = T$, and bounded measurable functions with compact support $f_j$, $j = 0, \ldots, N$ we have that

$$
\mathbb{E}_\mu \prod_{j=0}^{N} f_j(X_{t_j}) = \mathbb{E}_\mu \prod_{j=0}^{N} f_j(X_{T-t_j}),
$$

(149)

where $\mu(dx)$ denotes the invariant measure of $X_t$ and $\mathbb{E}_\mu$ denotes expectation with respect to $\mu$.

- Reversible diffusion processes can be characterized in terms of the properties of their generator. Time-reversal is equivalent to the selfadjointness of the generator in the Hilbert space $L^2(\mathbb{R}^d; \mu)$. 
Theorem

A stationary Markov process $X_t$ in $\mathbb{R}^d$ with generator $\mathcal{L}$ and invariant measure $\mu$ is reversible if and only if its generator is selfadjoint in $L^2(\mathbb{R}^d; \mu)$. 
Proof.

Assume first (149). We take $N = 1$ and $t_0 = 0$, $t_1 = T$ to deduce that

$$
\mathbb{E}_\mu \left( f_0(X_0)f_1(X_T) \right) = \mathbb{E}_\mu \left( f_0(X_T)f_1(X_0) \right), \quad \forall f_0, f_1 \in L^2(\mathbb{R}^d; \mu).
$$

This is equivalent to

$$
\int \left( e^{\mathcal{L}t}f_0(x) \right) f_1(x) \mu(dx) = \int f_0(x) \left( e^{\mathcal{L}t}f_1(x) \right) \mu(dx),
$$

i.e.

$$
\langle e^{\mathcal{L}t}f_1, f_2 \rangle_{L^2_\mu} = \langle f_1, e^{\mathcal{L}t}f_2 \rangle_{L^2_\mu}, \quad \forall f_1, f_2 \in L^2(\mathbb{R}^d; \rho_s). \tag{150}
$$

Consequently, the semigroup $e^{\mathcal{L}t}$ generated by $\mathcal{L}$ is selfadjoint. Differentiating (150) at $t = 0$ gives that $\mathcal{L}$ is selfadjoint.
Proof.

Conversely, assume that $\mathcal{L}$ is selfadjoint in $L^2(\mathbb{R}^d; \mu)$. We will use an induction argument. Our assumption of selfadjointness implies that (149) is true for $N = 1$

$$
\mathbb{E}_\mu \prod_{j=0}^1 f_j(X_{t_j}) = \mathbb{E}_\mu \prod_{j=0}^1 f_j(X_{T-t_j}),
$$

(151)

Assume that it is true for $N = k$. We have that

$$
\mathbb{E}_\mu \prod_{j=0}^k f_j(X_{t_j}) = \int \ldots \int f_0(x_0) \mu(dx_0) \prod_{j=1}^k f_j(x_j)p(t_j - t_{j-1}, x_{j-1}, dx_j)
$$

$$
= \mathbb{E}_\mu \prod_{n=0}^k f_j(X_{t_{j-1}})
$$

$$
= \int \ldots \int f_k(x_k) \mu(dx_k) \prod_{j=1}^k f_{j-1}(x_{j-1}) p(t_j - t_{j-1}, x_j, dx_{j-1}),
$$
Proof.

Now we show that (149) it is true for \( N = k + 1 \). We calculate, using (151) and (152)

\[
\mathbb{E}_{\mu} \prod_{j=1}^{k+1} f_j(X_{t_j}) = \mathbb{E}_{\mu} \prod_{j=1}^{k} f_j(X_{t_j}) f_{k+1}(X_{t_{k+1}})
\]

\[
= \int \ldots \int \mu(dx_0) f_0(x_0) \prod_{j=1}^{k} f_j(x_j) p(t_j - t_{j-1}, x_{j-1}, dx_j) \times f_{k+1}(x_{k+1}) p(t_{k+1} - t_k, x_k, dx_{k+1})
\]

\[
= \int \ldots \int \mu(dx_k) f_0(x_k) \prod_{j=1}^{k} f_{j-1}(x_{j-1}) p(t_j - t_{j-1}, x_j, dx_{j-1}) \times f_{k+1}(x_{k+1}) p(t_{k+1} - t_k, x_k, dx_{k+1})
\]

\[
= \int \ldots \int \mu(dx_k) f_0(x_k) f_{k+1}(x_{k+1}) p(t_{k+1} - t_k, x_k, dx_{k+1}) \times \prod_{j=1}^{k} f_{j-1}(x_{j-1}) p(t_j - t_{j-1}, x_j, dx_{j-1})
\]

\[
= \int \ldots \int \mu(dx_{k+1}) f_0(x_{k+1}) \prod_{j=1}^{k+1} f_{j-1}(x_{j-1}) p(t_j - t_{j-1}, x_j, dx_{j-1}) \times f_{k+1}(x_{k+1}) p(t_{k+1} - t_k, x_k, dx_{k+1}) = \mathbb{E}_{\mu} \prod_{j=0}^{k+1} f_j(X_{T-t_j}).
\]
In order to sample from $\pi(x)$ using the overdamped Langevin dynamics (148) we first have to discretize the SDE. The Euler discretization gives

$$X_{n+1} = X_n + \frac{1}{2}h\nabla \log \pi(X_n) + \sqrt{h}\xi_n, \quad \xi_n \sim \mathcal{N}(0, 1). \quad (153)$$

This can be written as

$$X_{n+1} \sim \mathcal{N}\left(X_n + \frac{1}{2}h\nabla \log \pi(X_n), hI\right) \quad (154)$$

This defines a Markov Chain $\{X_n\}_{n=0}^N$. It is not clear that this Markov chain has the same ergodic properties as the SDE.

We can choose to either accept or reject the next move $X_{n+1}$ with a certain probability. Introducing this accept-reject step to a given Markov chain leads to the **Metropolis-Hastings** algorithm.

The resulting **Metropolis adjusted** algorithm is always ergodic with respect to the target distribution $\pi(x)$. 

% SDE parameters

dt = 1.0e-2;
kappa = 0.1;
diff = sqrt(2*kappa*dt);

% number of iterations

N = 10^6;

% batch count (N must be divisible by m)
m = floor(sqrt(N));

% Significance value for confidence intervals

Delta = 0.01; % 99% CI

% Derivative of potential

Vprime = @(x) x^3 - x;

% Observable

f = @(x) x;

obs = zeros(1, N);

x = 0;
%simulate SDE
for i=1:N
    x = - Vprime(x) * dt + diff * randn(1);
    obs(i) = f(x);
end;

%The estimator is then given by
av = mean(obs);
disp(['Mean ', num2str(av)]);

%Compute the confidence intervals using a batch-means ... estimator
B = reshape(obs, m, []);

%Compute batch means
bmeans = mean(B, 1);

%Compute batch variance
D = var(bmeans);
disp(['Variance ', num2str(D)]);

%Compute confidence interval
cinv = tinv(1-\Delta/2, m-1) * sqrt(D/m);
%Now do all the
hold on
plot((1:N) * dt, cumsum(obs) ./ (1:N), '-r');
xlim([10 N * dt])

hline = refline(0, av - cinv);
set(hline, 'LineStyle', ':')
hline = refline(0, av + cinv);
set(hline, 'LineStyle', ':')
ylabel('\$f(X_t)$', 'interpreter', 'latex')
xlabel('\$t\$', 'interpreter', 'latex')
title('Estimating $f(x)=x$ from ...
  $e^{-\frac{(x^2-1)^2}{4\kappa}}$', 'interpreter', 'latex');
hold off
The Metropolis-Hastings Algorithm

- We are given a target distribution \( \pi(x) \) on \( \mathbb{R}^d \) that is known up to the normalisation constant.
- We are also given a Markov chain \( \{X_n\} \) with transition kernel density \( q(x, y) = q(x - y) \). Examples include:
  - For the random walk with diffusion coefficient \( \sigma^2 \) we have \( q(x) \sim \mathcal{N}(x, \sigma^2) \).
  - For the Euler discretization of the overdamped Langevin dynamics we consider the Markov chain (154).
- Given \( \pi \) and \( \{X_n\} \), a proposed value \( Y_{n+1} \) is generated from the density \( q(X_n, y) \) and is then accepted with probability
  \[
  \alpha(x, y) = \begin{cases} 
  \min \left\{ \frac{\pi(y)}{\pi(x)} \frac{q(y, x)}{q(x, y)}, 1 \right\} : & \pi(x)q(x, y) > 0, \\
  1 : & \pi(x)q(x, y) = 0.
  \end{cases} \quad (155)
  \]
- If the proposed value is accepted then set \( X_{n+1} = Y_{n+1} \) otherwise set \( X_{n+1} = X_n \).
The transition probability density for the Metropolis adjusted Markov chain is

\[ p(x, y) = q(x, y)\alpha(x, y), \quad \text{for} \ y \neq x \]  

(156)

and with probability of remaining at the same point given by

\[ r(x) = P(x, \{x\}) = \int q(x, y)[1 - \alpha(x, y)] \, dy. \]

With choice of the acceptance probability we have that the resulting Markov chain is ergodic (in fact, reversible) with respect to the target measure \( \pi(dx) \):

\[ \pi(A) = \int \pi(x)P(x, A) \, dx, \quad x \in X, \ A \in B, \]  

(157)

where \( X \) denotes the state space and \( B \) the Borel sets.
function [ X, acc ] = rwmh( X0, Δ, nsamp, ...
    targetDistribution)
% Metropolis Hastings algorithm

    dim = length(X0);
    X = zeros(nsamp, dim);
acc = zeros([1 nsamp]);

%set the initial state
X(1,:) = X0;

for i = 1:nsamp-1
    %Generate proposal
    Y = proposal(X(i,:), Δ);

    %Compute acceptance probability
    pX_given_Y = proposal_prob(Y, X(i,:), Δ);
    pY_given_X = proposal_prob(X(i,:), Y, Δ);

    piY = targetDistribution(Y);
    piX = targetDistribution(X(i,:));
    alpha = min(1, pX_given_Y * piY/(pY_given_X * piX));

    %Accept/Reject sample.
    if (rand < alpha)
        X(i+1,:) = Y;
        acc(i) = 1;
    else
        Pavliotis (IC) CompStochProc

    end
end
44        X(i+1,:) = X(i,:);
        acc(i) = 0;
        end

47        end

49        end

% RWMH proposal function
51        function Y = proposal(X, \Delta)
53            Y = normrnd(X, sqrt(\Delta), [1,length(X)]);
54        end

% RWMH proposal probability
56        function prob = proposal_prob(Y, X, \Delta)
58            d = dot((Y - X), (Y - X));
59            prob = exp(-d/(2*\Delta));
60        end
Given essentially any probability distribution $\pi$ the Metropolis-Hastings algorithm provides us with a method for generating a Markov chain $\{X_n\}$ that has $\pi$ as a stationary distribution.

When the transition kernel density $q(\cdot, \cdot)$ is symmetric then the formula for the acceptance probability simplifies to

$$\alpha(x, y) = \min \left\{ \frac{\pi(y)}{\pi(x)}, 1 \right\}.$$ 

For the **symmetric Random Walk Metropolis algorithm** (RWM), $q(\cdot) = \mathcal{N}(x, \sigma^2)$: Set $X_n = x$, choose $Y_{n+1} \sim \mathcal{N}(x, \sigma^2)$, set

$$X_{n+1} = \begin{cases} 
Y_{n+1} & : \text{with probability } \alpha(x, y), \\
Y_n & : \text{with probability } 1 - \alpha(x, y).
\end{cases}$$  \hspace{1cm} (158)
We consider an MCMC algorithm for sampling from $\pi(x)$, with or without an accept reject step. We consider a Markov chain $\Phi_n$. The $n$-step transition probability is

$$P^n(x, A) = P(\Phi_n \in A \mid \Phi_0 = x),$$

where $x \in \mathcal{X}$ (the state space) and $A \in \mathcal{B}$.

For the RWMH and the Metropolis adjusted Langevin algorithm (MALA) it converges to $\pi$ in total variation (strong law of large numbers)

$$\|P^n(x, \cdot) - \pi\|_{TV} := \frac{1}{2} \sup_{A \in \mathcal{B}} |P^n(x, A) - \pi(A)| \to 0. \quad (159)$$

Under additional assumptions we also have a central limit theorem.

The rate of convergence to the target distribution and the asymptotic variance can be used to measure the efficiency of the MCMC algorithm.
For an MCMC algorithm to be efficient, it has to be geometrically ergodic (exponentially fast convergence to equilibrium).

Without the Metropolis-Hastings step, the (unadjusted) Euler discretization of the Langevin dynamics (ULA) might fail to be (geometrically) ergodic.

This depends on the tails of the target distribution.

We can test the convergence of different MCMC algorithms for a class of 1d distributions $\mathcal{E}(\beta, \gamma)$:

$$
\pi \in \mathcal{E}(\beta, \gamma) \iff \pi(x) \propto e^{-\gamma |x|^\beta} |x| \geq x_0,
$$

for some $x_0$ and some positive constants $\gamma, \beta$.

The ULA Markov chain might fail to be (geometrically) ergodic even when the Langevin dynamics is, depending on the values of $\alpha$ and $\beta$ and on the step size.
Optimal Scaling for the Metropolis-Hastings Algorithm

- For the RWMH algorithm we can choose the diffusion coefficient:
  \[ q(\cdot) = \mathcal{N}(\cdot, \sigma^2). \]

- We want to choose the parameter \( \sigma \) in an optimal way.

- The RWMH can perform arbitrarily badly for \( \sigma \ll 1 \) and for \( \sigma \gg 1 \).
  There exists an optimal value \( \sigma^* \).

- We can measure the efficiency of the MCMC algorithm in terms of
  the asymptotic variance (for all \( f \in L^2(\pi) \)). The efficiency of the
  algorithm depends on the acceptance rate

  \[
  a = \int \alpha(x, y) \pi(x) q(x, y) \, dx dy \quad (161)
  \]

  \[
  = \lim_{n \to +\infty} \frac{1}{n} \text{number\{accepted moves\}}. \quad (162)
  \]

- Choosing the optimal scaling reduces the computational cost.
The optimal acceptance rate for RWMH is

\[ a_{opt} = 0.234. \]

For MALA:

\[ a_{opt} = 0.574. \]

The MALA algorithm asymptotically mixes considerably faster than the RWMH algorithm.
Consider the class $S$ of diffusion processes that are ergodic with respect to $\pi$, the distribution from which we want to sample.

1. Choose $X_t$ so that it converges as quickly as possible to the equilibrium.
2. Choose $X_t$ so that the stationary variance is minimized

$$\sigma_f^2 = \text{Var}_\pi(f) = \langle (-L)^{-1}f, f \rangle_\pi. \quad (163)$$

We can use (2) to introduce a partial ordering in $S$ (Peskun-Tierney ordering).
We can consider the nonreversible dynamics (Hwang et al 1993, 2005).

\[
dX_t^b = \left( -\nabla V(X_t^\gamma) + b(X_t^\gamma) \right) dt + \sqrt{2} dW_t,
\]

where \( \gamma \) is taken to be divergence-free with respect to the invariant distribution \( \pi(dx) = Z^{-1} e^{-V} dx \):

\[
\nabla \cdot (\gamma e^{-V}) = 0.
\]

This ensures that \( \pi(dx) \) is still the invariant measure of the dynamics (164).

We can construct such vector fields by taking

\[
\gamma = J \nabla V, \quad J = -J^T.
\]
The dynamics (164) is non-reversible: \((X_t^\gamma)_{0 \leq t \leq T}\) has the same law as \((X_{T-t}^{-\gamma})_{0 \leq t \leq T}\) and thus not the same law as \((X_{T-t}^\gamma)_{0 \leq t \leq T}\).

Equivalently, the system does not satisfy detailed balance—the stationary probability flux is not zero.

From (166) it is clear that there are many (in fact, infinitely many) different ways for modifying the reversible dynamics without changing the invariant measure.
We consider the nonreversible dynamics

$$dX_t = (-I + \delta J) \nabla V(X_t) \, dt + \sqrt{2\beta^{-1}} \, dW_t,$$

(167)

with $\delta \in \mathbb{R}$ and $J$ the standard $2 \times 2$ antisymmetric matrix, i.e. $J_{12} = 1$, $J_{21} = -1$. For this class of nonreversible perturbations the parameter that we wish to choose in an optimal way is $\delta$.

However, the numerical experiments will illustrate that even a non-optimal choice of $\delta$ can significantly accelerate convergence to equilibrium.

We will use the potential

$$V(x, y) = \frac{1}{4} (x^2 - 1)^2 + \frac{1}{2} y^2.$$  

(168)
Figure: Second moment as a function of time for (167) with the potential (168). We take 0 initial conditions and $\beta^{-1} = 0.1$. 
STATISTICAL INFERENCE FOR STOCHASTIC DIFFERENTIAL EQUATIONS
Many of the stochastic models that are used in applications include unknown parameters that have to be determined from observations.

We can use techniques from statistics to estimate parameters in the drift and diffusion coefficients.

We will consider one dimensional Itô SDEs of the form

\[ dX_t = b(X_t; \theta) \, dt + \sigma(X_t; \theta) \, dW_t, \quad X_0 = x, \]

where \( \theta \in \Theta \subset \mathbb{R}^N \) is a finite set of parameters that we want to estimate from observations.
We can consider either the case where only discrete observations are available, or the case where an entire path $X_t, \ t \in [0, T]$ is observed.

The length of the path can be either fixed or we can also consider the case where the observation interval increases, $T \to +\infty$.

We can also consider noisy observations:

$$Y_{t_j} = X_{t_j} + \xi \epsilon_{t_j}, \quad \epsilon_{t_j} \sim \mathcal{N}(0, 1), \ iid.$$
Examples

- Estimate the diffusion coefficient of Brownian motion
  \[ dX_t = \sqrt{2\sigma} \, dW_t. \]

- Estimate the drift and diffusion coefficients of the (stationary) OU process
  \[ dX_t = -\alpha X_t \, dt + \sqrt{2\sigma} \, dW_t. \]

- Estimate the drift and diffusion coefficients \( \theta = (A, B, \sigma_a, \sigma_b) \) in the Landau-Stuart equation with additive and multiplicative noise
  \[ dX_t = (AX_t - BX_t^3) \, dt + \sqrt{\sigma_a^2 + \sigma_b^2 X_t^2} \, dW_t. \] (170)

- Estimate the volatility in geometric Brownian motion:
  \[ dX_t = \mu X_t \, dt + \sigma X_t \, dW_t, \]
Estimate the integrated stochastic volatility in the Heston model:

\[ dX_t = \mu X_t \, dt + \sqrt{\nu_t} X_t \, dW_1^t, \]
\[ d\nu_t = \kappa (\theta - \nu_t) \, dt + \sigma \sqrt{\nu_t} \, dW_2^t, \]

where \( \mathbb{E}(W_1^t W_2^t) = \rho \).

The integrated stochastic volatility is

\[ \sigma(T) = \int_0^T \nu_t X_t^2 \, dt. \]
To estimate parameters in the diffusion coefficient we can use the **quadratic variation** of the process $X_t$

$$\langle X_t, X_t \rangle := \int_0^t \sigma^2(X_s; \theta) \, ds = \lim_{\Delta t_k \to 0} \sum_{t_k \leq t} |X_{t_{k+1}} - X_{t_k}|^2.$$

(171)

where the limit is in probability.

When the diffusion coefficient is constant the convergence in (171) becomes almost sure:

$$\lim_{n \to +\infty} \sum_{i=1}^{n} \left[ X_{iT2^{-n}} - X_{(i-1)T2^{-n}} \right]^2 = \sigma^2 T, \quad \text{a.s.} \quad (172)$$

If we fix the length of the observation $[0, T]$ and we let the number of observations become infinite, $n \to +\infty$ we can in fact determine (not only estimate) the diffusion coefficient.

This is called the **high frequency limit**.

$T$ can be (arbitrarily) small.

For the estimation of the diffusion coefficient we do not need to assume that the process $X_t$ is stationary.
Theorem

Let \( \{X_j\}_{j=0}^J \) be a sequence of equidistant observations of (??) with timestep \( \Delta t = \delta \) and \( J\delta = T \) fixed. Assume that the drift \( b(x; \theta) \) is bounded and define

\[
\hat{\sigma}_J^2 = \frac{1}{J\delta} \sum_{j=0}^{J-1} (X_{j+1} - X_j)^2,
\]

Then

\[
|\mathbb{E}\hat{\sigma}_J^2 - \sigma^2| \leq C(\delta + \delta^{1/2}).
\]

In particular,

\[
\lim_{J \to +\infty} |\mathbb{E}\hat{\sigma}_J^2 - \sigma^2| = 0.
\]
We have

$$X_{j+1} - X_j = \int_{j\delta}^{(j+1)\delta} b(X_s; \theta) \, ds + \sigma \Delta W_j,$$

where $\Delta W_j = W_{(j+1)\delta} - W_{j\delta} \sim \mathcal{N}(0, \delta)$. We substitute this into (173)

$$\hat{\sigma}_j^2 = \sigma^2 \frac{1}{\delta J} \sum_{j=0}^{J-1} (\Delta W_j)^2 + \frac{2}{\delta J} \sum_{j=0}^{J-1} I_j M_j + \frac{1}{\delta J} \sum_{j=0}^{J-1} I_j^2,$$

where

$$I_j := \int_{j\delta}^{(j+1)\delta} b(X_s; \theta) \, ds$$

and $M_j := \sigma \Delta W_j$.

Note that $\mathbb{E}(\Delta W_n)^2 = \delta$. 
From the boundedness of $b(x; \theta)$ and using the Cauchy-Schwarz inequality we get

$$\mathbb{E}I_j^2 \leq \delta \int_{j\delta}^{(j+1)\delta} \mathbb{E}(b(X_s; \theta))^2 \, ds \leq C\delta^2.$$ 

Consequently:

$$\left| \mathbb{E}\hat{\sigma}_j^2 - \sigma^2 \right| \leq \frac{1}{\delta} \mathbb{E}I_j^2 + \frac{2}{\delta} \mathbb{E}|I_jM_j|$$

$$\leq C\delta + \frac{C}{\delta} \left( \frac{1}{\alpha} \mathbb{E}I_j^2 + \alpha \mathbb{E}M_j^2 \right)$$

$$\leq C \left( \delta + \delta^{1/2} \right).$$

In the above we used Cauchy’s inequality with $\alpha = \delta^{1/2}$. 

In the above we used Cauchy’s inequality with $\alpha = \delta^{1/2}$. 

In the above we used Cauchy’s inequality with $\alpha = \delta^{1/2}$. 

In the above we used Cauchy’s inequality with $\alpha = \delta^{1/2}$. 

In the above we used Cauchy’s inequality with $\alpha = \delta^{1/2}$. 

In the above we used Cauchy’s inequality with $\alpha = \delta^{1/2}$. 

In the above we used Cauchy’s inequality with $\alpha = \delta^{1/2}$. 

In the above we used Cauchy’s inequality with $\alpha = \delta^{1/2}$. 

In the above we used Cauchy’s inequality with $\alpha = \delta^{1/2}$. 

In the above we used Cauchy’s inequality with $\alpha = \delta^{1/2}$. 

In the above we used Cauchy’s inequality with $\alpha = \delta^{1/2}$. 

In the above we used Cauchy’s inequality with $\alpha = \delta^{1/2}$. 

In the above we used Cauchy’s inequality with $\alpha = \delta^{1/2}$. 

In the above we used Cauchy’s inequality with $\alpha = \delta^{1/2}$. 

In the above we used Cauchy’s inequality with $\alpha = \delta^{1/2}$. 

In the above we used Cauchy’s inequality with $\alpha = \delta^{1/2}$. 

In the above we used Cauchy’s inequality with $\alpha = \delta^{1/2}$. 

In the above we used Cauchy’s inequality with $\alpha = \delta^{1/2}$. 

In the above we used Cauchy’s inequality with $\alpha = \delta^{1/2}$. 

In the above we used Cauchy’s inequality with $\alpha = \delta^{1/2}$. 

In the above we used Cauchy’s inequality with $\alpha = \delta^{1/2}$. 

In the above we used Cauchy’s inequality with $\alpha = \delta^{1/2}$. 

In the above we used Cauchy’s inequality with $\alpha = \delta^{1/2}$. 

In the above we used Cauchy’s inequality with $\alpha = \delta^{1/2}$. 

In the above we used Cauchy’s inequality with $\alpha = \delta^{1/2}$. 

In the above we used Cauchy’s inequality with $\alpha = \delta^{1/2}$. 

In the above we used Cauchy’s inequality with $\alpha = \delta^{1/2}$. 

In the above we used Cauchy’s inequality with $\alpha = \delta^{1/2}$. 

In the above we used Cauchy’s inequality with $\alpha = \delta^{1/2}$. 

In the above we used Cauchy’s inequality with $\alpha = \delta^{1/2}$. 

In the above we used Cauchy’s inequality with $\alpha = \delta^{1/2}$. 

In the above we used Cauchy’s inequality with $\alpha = \delta^{1/2}$. 

In the above we used Cauchy’s inequality with $\alpha = \delta^{1/2}$. 

In the above we used Cauchy’s inequality with $\alpha = \delta^{1/2}$. 

In the above we used Cauchy’s inequality with $\alpha = \delta^{1/2}$. 

In the above we used Cauchy’s inequality with $\alpha = \delta^{1/2}$. 

In the above we used Cauchy’s inequality with $\alpha = \delta^{1/2}$. 

In the above we used Cauchy’s inequality with $\alpha = \delta^{1/2}$. 

In the above we used Cauchy’s inequality with $\alpha = \delta^{1/2}$. 

In the above we used Cauchy’s inequality with $\alpha = \delta^{1/2}$. 

In the above we used Cauchy’s inequality with $\alpha = \delta^{1/2}$. 

In the above we used Cauchy’s inequality with $\alpha = \delta^{1/2}$. 

In the above we used Cauchy’s inequality with $\alpha = \delta^{1/2}$. 

In the above we used Cauchy’s inequality with $\alpha = \delta^{1/2}$. 

In the above we used Cauchy’s inequality with $\alpha = \delta^{1/2}$. 

In the above we used Cauchy’s inequality with $\alpha = \delta^{1/2}$. 

In the above we used Cauchy’s inequality with $\alpha = \delta^{1/2}$. 

In the above we used Cauchy’s inequality with $\alpha = \delta^{1/2}$. 

In the above we used Cauchy’s inequality with $\alpha = \delta^{1/2}$. 

In the above we used Cauchy’s inequality with $\alpha = \delta^{1/2}$. 

In the above we used Cauchy’s inequality with $\alpha = \delta^{1/2}$. 

In the above we used Cauchy’s inequality with $\alpha = \delta^{1/2}$. 

In the above we used Cauchy’s inequality with $\alpha = \delta^{1/2}$. 

In the above we used Cauchy’s inequality with $\alpha = \delta^{1/2}$. 

In the above we used Cauchy’s inequality with $\alpha = \delta^{1/2}$. 

In the above we used Cauchy’s inequality with $\alpha = \delta^{1/2}$. 

In the above we used Cauchy’s inequality with $\alpha = \delta^{1/2}$. 

In the above we used Cauchy’s inequality with $\alpha = \delta^{1/2}$. 

In the above we used Cauchy’s inequality with $\alpha = \delta^{1/2}$. 

In the above we used Cauchy’s inequality with $\alpha = \delta^{1/2}$. 

In the above we used Cauchy’s inequality with $\alpha = \delta^{1/2}$. 

In the above we used Cauchy’s inequality with $\alpha = \delta^{1/2}$. 

In the above we used Cauchy’s inequality with $\alpha = \delta^{1/2}$. 

In the above we used Cauchy’s inequality with $\alpha = \delta^{1/2}$. 

In the above we used Cauchy’s inequality with $\alpha = \delta^{1/2}$. 

In the above we used Cauchy’s inequality with $\alpha = \delta^{1/2}$. 

In the above we used Cauchy’s inequality with $\alpha = \delta^{1/2}$. 

In the above we used Cauchy’s inequality with $\alpha = \delta^{1/2}$. 

In the above we used Cauchy’s inequality with $\alpha = \delta^{1/2}$. 

In the above we used Cauchy’s inequality with $\alpha = \delta^{1/2}$. 

In the above we used Cauchy’s inequality with $\alpha = \delta^{1/2}$. 

In the above we used Cauchy’s inequality with $\alpha = \delta^{1/2}$. 

In the above we used Cauchy’s inequality with $\alpha = \delta^{1/2}$. 

In the above we used Cauchy’s inequality with $\alpha = \delta^{1/2}$. 

In the above we used Cauchy’s inequality with $\alpha = \delta^{1/2}$. 

In the above we used Cauchy’s inequality with $\alpha = \delta^{1/2}$. 

In the above we used Cauchy’s inequality with $\alpha = \delta^{1/2}$.
Assume that we have already estimated the diffusion coefficient. Consider the SDE
\[ dX_t = b(X_t; \theta) \, dt + dW_t. \] (176)

Our goal is to estimate the unknown parameters in the drift \( \theta \in \Theta \) from discrete observations.

We denote the true value by \( \theta_0 \).

We will use the maximum likelihood approach which is based on maximizing the likelihood function.

We first study the problem of estimating parameters in the distribution function of random variables using observations.
We consider a random variable $X$ whose probability distribution function $f(x|\theta)$ is known up to parameters $\theta$ that we want to estimate from observations.

When $X$ is a Gaussian random variable, in which case the parameters to be estimated are the mean and variance, $\theta = (\mu, \sigma)$.

Suppose that we have $N$ independent observations of the random variable $X$.

We define the **likelihood function**

$$L(\{x_i\}_{i=1}^N | \theta) = \prod_{i=1}^N f(x_i|\theta). \quad (177)$$

The likelihood function is essentially the probability density function of the random variable $X$, viewed as a function of the parameters $\theta$. The maximum likelihood estimator (MLE) is then

$$\hat{\theta} = \text{argmax} \; L(x|\theta), \quad (178)$$

with $x = \{x_i\}_{i=1}^N$. 

"Pavliotis (IC) CompStochProc 244 / 298"
The MLE is a random variable that depends on the observations \( \{x_i\}_{i=1}^N \). When \( X \) is a Gaussian random variable, \( X \sim \mathcal{N}(\mu, \sigma^2) \) the likelihood function takes the form

\[
L(\{x_i\}_{i=1}^N | \theta) = \left( \frac{1}{2\pi\sigma^2} \right)^{N/2} \exp \left( -\frac{\sum_{i=1}^{N} (x_i - \mu)^2}{2\sigma^2} \right).
\]

Maximizing (the log likelihood function) then with respect to \( \mu \) and \( \sigma^2 \) we obtain the maximum likelihood estimators

\[
\hat{\mu} = \frac{1}{N} \sum_{i=1}^{N} x_i, \quad \hat{\sigma}^2 = \frac{1}{N} \sum_{i=1}^{N} (x_i - \hat{\mu})^2. \tag{179}
\]

Notice that

\[
\mathbb{E}\hat{\mu} = \mu \quad \text{and} \quad \mathbb{E}\hat{\sigma}^2 = \frac{N - 1}{N} \sigma^2.
\]
Under appropriate assumptions the maximum likelihood estimator is consistent: We can estimate the true values of the parameters as accurately as we wish if we have a sufficiently large number of observations $N$.

Use the LLN and the CLT to study the asymptotic (large $N$) properties of the MLE.

In particular, we can show that

$$\lim_{N \to +\infty} \hat{\theta} = \theta_0,$$

either in probability or almost surely and

$$\lim_{N \to +\infty} \sqrt{N}(\hat{\theta} - \theta_0) = \mathcal{N}(0, D^2),$$

in distribution, for an appropriate constant $D^2$. 
We want to use this idea in order to estimate the parameters in the drift of the SDE (176).

The (independent) observations of the random variable $X$ are now replaced by observations of the process $X_t, \{X_i\}_{i=1}^N$ with $X_i = X_ih$ and $hN = T$.

We have $N$ equidistant observations, or with an entire path $X_t, t \in [0, T]$.

Assume then that a path of the process $X_t$ is observed.

The analogue of the likelihood function (177) is the law of the process on path space.

To obtain a formula for the likelihood function we need to use Girsanov’s theorem.
The law of $X_t$, denoted by $\mathbb{P}_X$, is absolutely continuous with respect to the Wiener measure, the law of Brownian motion.

The density of $\mathbb{P}_X$ with respect to the Wiener measure is given by the Radon-Nikodym derivative:

$$
\frac{d\mathbb{P}_X}{d\mathbb{P}_W} = \exp \left( \int_0^T b(X_s; \theta) \, dX_s - \frac{1}{2} \int_0^T (b(X_s; \theta))^2 \, ds \right) =: L(\{X_t\}_{t\in[0,T]}; \theta, T).
$$

The maximum likelihood estimator MLE is defined as

$$
\hat{\theta} = \arg\max_{\theta \in \Theta} L(\{X_t\}_{t\in[0,T]}; \theta).
$$

The MLE estimator is a random variable that depends on the path $\{X_t\}_{t\in[0,T]}$. 
Assume that the diffusion process (176) is stationary.

The MLE (181) is **asymptotically unbiased**: in the limit as the window of observation becomes infinite, $T \to +\infty$, the MLE $\hat{\theta}$ converges to the true value $\theta_0$.

Assume that there are $N$ parameters to be estimated, $\theta = (\theta_1, \ldots, \theta_N)$. The MLE is obtained by solving the (generally nonlinear) system of equations

$$\frac{\partial L}{\partial \theta_i} = 0, \quad i = 1, \ldots, N. \quad (182)$$

The solution of this system of equations can be expressed in terms of functionals (e.g. moments) of the observed path $\{X_t\}_{t \in [0,T]}$

$$\hat{\theta} = \mathcal{F}(\{X_t\}_{t \in [0,T]}).$$
(MLE for the stationary OU process). Consider the stationary OU process

\[ dX_t = -\alpha X_t \, dt + dW_t \]  \hspace{1cm} (183)

with \( X_0 \sim \mathcal{N}(0, \frac{1}{2\alpha}) \). The log Likelihood function

\[ \log L = \alpha \int_0^T X_t \, dX_t - \frac{\alpha^2}{2} \int_0^T X_t^2 \, dt. \]

We solve the equation \( \frac{\partial \log L}{\partial \alpha} = 0 \) from which we obtain

\[ \hat{\alpha} = -\frac{\int_0^T X_t \, dX_t}{\int_0^T X_t^2 \, dt} = -\frac{B_1(\{X_t\}_{t \in [0,T]\})}{M_2(\{X_t\}_{t \in [0,T]\})} \]  \hspace{1cm} (184)

where we have used the notation

\[ B_n(\{X_t\}_{t \in [0,T]\}) := \int_0^T X_t^n \, dX_t, \quad M_n(\{X_t\}_{t \in [0,T]\}) := \int_0^T X_t^n \, dt. \]  \hspace{1cm} (185)
Given a set of discrete equidistant observations \( \{X_j\}_{j=0}^J \),  
\( X_j = X_j \Delta t \), \( \Delta X_j = X_{j+1} - X_j \), formula (184) can be approximated by

\[
\hat{\alpha} = - \frac{\sum_{j=0}^{J-1} X_j \Delta X_j}{\sum_{j=0}^{J-1} |X_j|^2 \Delta t}.
\]  

(186)

The MLE (184) becomes asymptotically unbiased in the large sample limit \( J \to +\infty \), \( \Delta t \) fixed.
Figure: MLE for the OU process.
Using Itô’s formula we can obtain an alternative formula for the maximum likelihood estimator of the drift coefficient for the OU process.

The numerator in (184) can be written as

\[ \int_0^t X_s dX_s = -\alpha \int_0^t X_s^2 ds + \int_0^t X_s dW_s. \]

We apply Itô’s formula to the function \( V(x) = \frac{1}{2}x^2 \) to obtain

\[ dV(X_t) = -\alpha X_t^2 dt + \frac{1}{2} dt + X_t dW_t. \]

We combine the above two equations to obtain

\[ \int_0^t X_s dX_s = \frac{X_t^2 - X_0^2 - t}{2}. \]

The formula for the MLE now becomes

\[ \hat{\alpha} = -\frac{X_T^2 - X_0^2 - T}{2 \int_0^T X_t^2 dt}. \]
Example

Consider the following generalization of the previous example:

\[ dX_t = \alpha b(X_t) \, dt + dW_t, \]  

(188)

where \( b(x) \) is such that the equation has a unique ergodic solution. The log Likelihood function is

\[ \log L = \alpha \int_0^T b(X_t) \, dX_t - \frac{\alpha^2}{2} \int_0^T b(X_t)^2 \, dt. \]

The MLE is

\[ \hat{\alpha} = \frac{\int_0^T b(X_t) \, dX_t}{\int_0^T (b(X_t))^2 \, dt}. \]
Example

MLE for a stationary bistable SDE

Consider the SDE

$$dX_t = (\alpha X_t - \beta X_t^3) \, dt + dW_t$$  \hspace{1cm} (189)$$

This SDE is of the form $dX_t = -V'(X_t) \, dt + dW_t$ with

$V(x) = \frac{\alpha}{2} x^2 - \frac{\beta}{4} x^4$ and is ergodic with invariant distribution

$\rho(x) = Z^{-1} e^{-\frac{1}{2} V(x)}$.

Our goal is to estimate the coefficients $\alpha$ and $\beta$ from observations using the maximum likelihood approach. The log likelihood functions reads

$$\log L = \int_0^T (\alpha X_t - \beta X_t^3) \, dX_t - \frac{1}{2} \int_0^T (\alpha X_t - \beta X_t^3)^2 \, dt$$

$$=: \quad \alpha B_1 - \beta B_3 - \frac{1}{2} \alpha^2 M_2 - \frac{1}{2} \beta^2 M_6 + \alpha \beta M_4,$$

using the notation (185).
Example

Equations (182) become

$$\frac{\partial \log L}{\partial \alpha}(\hat{\alpha}, \hat{\beta}) = 0, \quad \frac{\partial \log L}{\partial \beta}(\hat{\alpha}, \hat{\beta}) = 0,$$

This leads to a linear system of equations

$$\begin{pmatrix} M_2 & -M_4 \\ M_4 & -M_6 \end{pmatrix} \begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix} = \begin{pmatrix} B_1 \\ B_3 \end{pmatrix},$$

The solution of which is

$$\hat{\alpha} = \frac{B_1M_6 - B_3M_4}{M_2M_6 - M_4^2}, \quad \hat{\beta} = \frac{B_1M_4 - B_3M_2}{M_2M_6 - M_4^2}.$$ (190)
Figure: MLE estimators for a bistable potential.
The rigorous justification of the MLE is based on Girsanov’s theorem.

Here we present a heuristic derivation of (180) that is based on the Euler-Marayama discretization of (176):

$$X_{n+1} - X_n = b(X_n; \theta) \Delta t + \Delta W_n,$$

(191)

where $X_n = X(n\Delta t)$ and $\Delta W_n = W_{n+1} - W_n =: \xi_n$.

We have that $\xi_n \sim \mathcal{N}(0, \sqrt{\Delta t})$, mutually independent.

Our goal is to calculate the Radon-Nikodym derivative of the law of the discrete-time process $\{X_n\}_{n=0}^{N-1}$ and the discretized Brownian motion.

In the discrete case this derivative becomes the ratio between the distribution functions of the two processes.
We rewrite (191) in the form

$$\Delta X_n = b_n \Delta + \xi_n,$$  \hspace{1cm} (192)

where $b_n := b(X_n; \theta)$, $\Delta := \Delta t$. Our goal is to calculate the Radon–Nikodym derivative of the law of the discrete-time process $\{X_n\}_{n=0}^{N-1}$ and the discretized Brownian motion.

In the discrete case, this derivative becomes the ratio between the distribution functions of the two processes.

We rewrite (??) in the form

$$\Delta X_n = b_n \Delta t + \xi_n \sqrt{\Delta t},$$  \hspace{1cm} (193)

where $b_n := b(X_n; \theta)$. The distribution function of the discretized Brownian motion is

$$p^N_W = \prod_{i=0}^{N-1} \frac{1}{\sqrt{2\pi \Delta t}} \exp \left( - \frac{1}{2 \Delta t} (\Delta W_i)^2 \right)$$

$$= \frac{1}{(\sqrt{2\pi \Delta t})^N} \exp \left( - \frac{1}{2 \Delta t} \sum_{i=0}^{N-1} (\Delta W_i)^2 \right).$$  \hspace{1cm} (194)
Similarly, for the law of the discretized process \( \{X_n\}_{n=0}^{N-1} \), using the fact that \( p(X_{i+1}|X_i) \sim \mathcal{N}(X_i + b_i \Delta t, \Delta t) \), we can write

\[
p_X^N = \frac{1}{(\sqrt{2\pi\Delta t})^N} \exp \left( -\sum_{i=0}^{N-1} \left( \frac{1}{2\Delta t} (\Delta X_i)^2 + \frac{1}{2} (b_i)^2 \Delta t - b_i \Delta X_i \right) \right).
\]

Now we can calculate the ratio of the laws of the two processes, evaluated at the path \( \{X_n\}_{n=0}^{N-1} \):

\[
\frac{d\mathbb{P}_X^N}{d\mathbb{P}_W^N} = \exp \left( -\frac{1}{2} \sum_{i=0}^{N-1} (b_i)^2 \Delta t + \sum_{i=0}^{N-1} b_i \Delta X_i \right).
\]

Passing now (formally) to the limit as \( N \to +\infty \) while keeping \( \Delta \) fixed, we obtain (180).
Lamperti’s Transformation and Girsanov’s Theorem

- For SDEs in one dimension it is possible to map multiplicative noise to additive noise.
- Consider a one dimensional Itô SDE with multiplicative noise

\[
dX_t = f(X_t) \, dt + \sigma(X_t) \, dW_t. \tag{196}
\]

- We ask whether there exists a transformation \( z = h(x) \) that maps (196) into an SDE with additive noise.
- We apply Itô’s formula to obtain

\[
dZ_t = \mathcal{L}h(X_t) \, dt + h'(X_t)\sigma(X_t) \, dW_t,
\]

- where \( \mathcal{L} \) denotes the generator of \( X_t \).
In order to obtain an SDE with unit diffusion coefficient we need to impose the condition

\[ h'(x)\sigma(x) = 1, \]

from which we deduce that

\[ h(x) = \int_{x_0}^{x} \frac{1}{\sigma(x)} \, dx, \tag{197} \]

where \( x_0 \) is arbitrary. We have that

\[ \mathcal{L}h(x) = \frac{f(x)}{\sigma(x)} - \frac{1}{2} \sigma'(x). \]

Consequently, the transformed SDE has the form

\[ dY_t = f_Y(Y_t) \, dt + dW_t \tag{198} \]

with

\[ f_Y(y) = \frac{f(h^{-1}(y))}{\sigma(h^{-1}(y))} - \frac{1}{2} \sigma'(h^{-1}(y)). \]

This is called the Lamperti transformation.
Consider the **Cox-Ingersoll-Ross (CIR) SDE**

\[
dX_t = (\mu - \alpha X_t) \, dt + \sigma \sqrt{X_t} \, dW_t, \quad X_0 = x > 0, \tag{199}
\]

where \( \mu, \alpha, \sigma > 0 \). From (197) we deduce \( h(x) = \frac{2}{\sigma} \sqrt{x} \).

The generator of the CIR process is

\[
\mathcal{L} = (\mu - \alpha x) \frac{d}{dx} + \frac{\sigma^2}{2} x \frac{d^2}{dx^2}.
\]

We have that

\[
\mathcal{L} h(x) = \left( \frac{\mu}{\sigma} - \frac{\sigma}{4} \right) x^{-1/2} - \frac{\alpha}{\sigma} x^{1/2}.
\]

The CIR SDE becomes, for \( Y_t = \frac{2}{\sigma} \sqrt{X_t} \),

\[
dY_t = \left( \frac{\mu}{\sigma} - \frac{\sigma}{4} \right) \frac{1}{\sqrt{X_t}} \, dt - \frac{\alpha}{\sigma} \sqrt{X_t} \, dt + dW_t
\]

\[
\quad = \left[ \left( \frac{2\mu}{\sigma^2} - \frac{1}{2} \right) \frac{1}{Y_t} - \frac{\alpha}{2} Y_t \right] \, dt + dW_t.
\]

When \( \mu = \frac{\sigma^2}{4} \) the equation above becomes the **Ornstein-Uhlenbeck process** for \( Y_t \).
The Lamperti transformation is very useful when estimating parameters for an SDE model using the MLE approach.

Such a transformation does not exist for arbitrary SDEs in higher dimensions. In particular, it is not possible, in general, to transform a multidimensional Itô SDE with multiplicative noise to an SDE with additive noise.

Multidimensional diffusion processes for which the reduction from multiplicative to additive noise is possible are called reducible.

Conditions on the coefficients of an SDE so that it is reducible are obtained in Ait-Sahalia, *Closed form likelihood expansions for multivariate diffusions, the Annals of Statistics (2008)* **36**(2) 906–937.
Conversely, it is also sometimes possible to remove the drift term from an SDE and obtain an equation where only (multiplicative) noise is present.

Consider the one dimensional SDE:

$$dX_t = b(X_t) \, dt + dW_t.$$  \hfill (200)

We introduce the following functional of $X_t$

$$M_t = \exp \left( - \frac{1}{2} \int_0^t b^2(X_s) \, ds + \int_0^t b(X_s) \, dW_s \right).$$  \hfill (201)

We can write $M_t = e^{-Y_t}$ where $Y_t$ is the solution to the SDE

$$M_t = e^{-Y_t}, \quad dY_t = \frac{1}{2} b^2(X_t) \, dt + b(X_t) \, dW_t, \quad Y_0 = 0.$$  \hfill (202)

We can now apply Itô’s formula to obtain the SDE

$$dM_t = -M_t b(X_t) \, dW_t.$$  \hfill (203)

Notice that this is an equation without drift.
Under appropriate conditions on the drift $b(\cdot)$, it is possible to show that the law of the process $X_t$, denoted by $\mathbb{P}$, which is a probability measure over the space of continuous functions, is absolutely continuous with respect to the Wiener measure $\mathbb{P}_W$, the law of the Brownian motion $W_t$.

The **Radon-Nikodym** derivative between these two measures is the inverse of the stochastic process $M_t$ given in (202):

$$
\frac{d\mathbb{P}}{d\mathbb{P}_W}(X_t) = \exp \left( \frac{1}{2} \int_0^t b^2(X_s) \, ds + \int_0^t b(X_s) \, dW_s \right). \quad (204)
$$

This is a form of **Girsanov’s theorem**.

Using now equation (200) we can rewrite (204) as

$$
\frac{d\mathbb{P}}{d\mathbb{P}_W}(X_t) = \exp \left( \int_0^t b(X_s) \, dX_s - \frac{1}{2} \int_0^t |b(X_s)|^2 \, ds \right). \quad (205)
$$
A form of Girsanov’s theorem that is very useful in statistical inference for diffusion processes is the following: consider the two SDEs

\[
\begin{align*}
    dX_t &= b_1(X_t) \, dt + \sigma(X_t) \, dW_t, \quad X_0 = x^1, \quad t \in [0, T], \quad (206a) \\
    dX_t &= b_2(X_t) \, dt + \sigma(X_t) \, dW_t, \quad X_0 = x^2, \quad t \in [0, T], \quad (206b)
\end{align*}
\]

where \( \sigma(x) > 0 \). We assume that we have existence and uniqueness of strong solutions for both SDEs.

Assume that \( x^1 \) and \( x^2 \) are random variables with densities \( f_1(x) \) and \( f_2(x) \) with respect to the Lebesgue measure which have the same support, or nonrandom and equal to the same constant.

Let \( P_1 \) and \( P_2 \) denote the laws of these two SDEs. Then these two measures are equivalent and their Radon-Nikodym derivative is

\[
\frac{dP_2}{dP_1}(X) = \frac{f_2(X_0)}{f_1(X_0)} \exp \left( \int_0^T \frac{b_2(X_t) - b_1(X_t)}{\sigma^2(X_t)} \, dX_t - \frac{1}{2} \int_0^T \frac{b_2^2(X_t) - b_1^2(X_t)}{\sigma^2(X_t)} \, dt \right). \quad (207)
\]
Example: estimation of the diffusion and the drift for the OU process

```matlab
1  % Estimate Parameters in the OU process.
2  % Use MLE and quadratic variation
3  % Input: a path of the SDE and the observation times
4  %
5  function [alpha, lambda, t, x] = param_est(q)
6  t = q(:,1); x = q(:,2);
7  figure; plot(t,x,'Linewidth',2)
8  hold
9  N = length(t); dt = t(2) - t(1); T = N*dt;
10  lambda = (1/(2*T))*sum(diff(x).^2)
11  alpha = - sum(x(1:end-1).*diff(x))/(sum(x.^2)*dt)
```
The MLE (181) depends on the path \( \{X_t\}_{t \in [0,T]} \) and consequently it is random variable.

We have to prove that, in the large sample limit \( J \rightarrow +\infty \), \( \Delta t \) fixed, and for appropriate assumptions on the diffusion process \( X_t \), the MLE converges to the true value \( \theta_0 \) and to also obtain information about the fluctuations around the limiting value \( \theta_0 \).

Assuming that \( X_t \) is stationary we can prove that the MLE \( \hat{\theta} \) converges in the limit as \( T \rightarrow +\infty \) (assuming that the entire path \( \{X_t\}_{t \in [0,T]} \) is available to us) to \( \theta_0 \).

Furthermore, we can prove asymptotic normality of the maximum likelihood estimator,

\[ \sqrt{T}(\hat{\theta} - \theta_0) \rightarrow \mathcal{N}(0, \sigma^2), \quad (208) \]

for a variance \( \sigma^2 \) that can be calculated.
Theorem

Let $X_t$ be the stationary OU process

\[ dX_t = -\alpha X_t \, dt + dW_t, \quad X_0 \sim \mathcal{N} \left( 0, \frac{1}{2\alpha} \right) \]

and let $\hat{\alpha}$ denote the MLE (184). Then

\[ \lim_{T \to +\infty} \sqrt{T} |\hat{\alpha} - \alpha| = \mathcal{N} \left( 0, 2\alpha \right) \]

(209)

in distribution.
For the proof of this theorem we will need the following result from probability theory.

**Theorem**

**Slutsky** Let \( \{X_n\}_{n=1}^{+\infty} \), \( \{Y_n\}_{n=1}^{+\infty} \) be sequences of random variables such that \( X_n \) converges in distribution to a random variable \( X \) and \( Y_n \) converges in probability to a constant \( c \neq 0 \). Then

\[
\lim_{n \to +\infty} Y_n^{-1} X_n = c^{-1} X,
\]

in distribution.
Proof of Theorem 57.

- First we observe that

\[ \hat{\alpha} = -\frac{\int_0^T X_t \, dX_t}{\int_0^T X_t^2 \, dt} = \alpha - \frac{\int_0^T X_t \, dW_t}{\int_0^T X_t^2 \, dt}. \]

- Consequently:

\[ \hat{\alpha} - \alpha = -\frac{\int_0^T X_t \, dW_t}{\int_0^T X_t^2 \, dt} = -\frac{1}{\sqrt{T}} \frac{1}{\frac{1}{T} \int_0^T X_t^2 \, dt} \]

\text{Law} \quad \equiv \quad -\frac{1}{\sqrt{T}} \frac{W \left( \frac{1}{T} \int_0^T X_t^2 \, dt \right)}{\frac{1}{T} \int_0^T X_t^2 \, dt},

where the scaling property of Brownian motion was used.

\[ \int_0^T f(s) \, dW(s) = W \left( \int_0^T f^2(s) \, ds \right). \quad (210) \]
• The process $X_t$ is stationary. We use the ergodic theorem for stationary Markov processes to obtain

$$\lim_{T \to +\infty} \frac{1}{T} \int_0^T X_t^2 \, dt = \mathbb{E}X_t^2 = \frac{1}{2\alpha},$$

(211)

• Let now $Y = \mathcal{N}(0, \sigma^2)$ with $\sigma^2 = \frac{1}{2\alpha}$. We can write $Y = W(\sigma^2)$.

• We use the Hölder continuity of Brownian motion to conclude that, almost surely,

$$\left| \frac{1}{\sqrt{T}} \int_0^T X_t^2 \, dt - Y \right| = \left| W \left( \frac{1}{T} \int_0^T X_t^2 \, dt \right) - W \left( \frac{1}{2\alpha} \right) \right| \leq \text{Hö l}(W) \left| \frac{1}{T} \int_0^T X_t^2 \, dt - \frac{1}{2\alpha} \right|^\frac{1}{2-\epsilon},$$

(212)

(213)

where $\text{Hö l}(W)$ denotes the Hölder constant of Brownian motion.
We have (in distribution)

$$\lim_{T \to +\infty} \frac{1}{T} \int_0^T X_t \, dW_t = \mathcal{N} \left( 0, \frac{1}{2\alpha} \right),$$

(214)

We combine (211) with (214) and use Slutsky’s theorem to conclude that (in distribution)

$$\lim_{T \to +\infty} \sqrt{T} |\hat{\alpha} - \alpha| = \mathcal{N} \left( 0, 2\alpha \right)$$

(215)
**Econometrics: Market Microstructure Noise**

Zhang, Mykland, Ait-Sahalia J. American Statistical Association 2005 **100**(472) pp. 1394–1411

- $S_t$ is the price process of a security. $X_t = \log(S_t)$ is the solution of

\[
dX_t = \mu_t \, dt + \sigma_t \, dW_t^1.
\]

(216)

- $\mu_t$, $\sigma_t$ are stochastic processes.
- For example, we can take (Heston model)

\[
\mu_t = (\mu - \nu_t/2), \quad \nu_t = \sigma_t^2.
\]

- and

\[
d\nu_t = \kappa (\alpha - \nu_t) \, dt + \gamma \nu_t^{1/2} \, dW_t^2,
\]

(217)

- with $\mathbb{E}(dW_t^1 \, dW_t^2) = \rho \, dt$.
- Goal: Estimate the integrated stochastic volatility of $X_t$ from noisy observations $Y_t$. 
Assume that we have \textit{market microstructure (observation error)}:

\[ Y_{ti} = X_{ti} + \varepsilon_{ti}, \quad i = 0, \ldots, N \quad (218) \]

where

\[ \varepsilon_{ti} \text{ iid } \mathbb{E}\varepsilon_{ti} = 0, \quad \mathbb{E}\varepsilon_{ti}^2 = \sigma^2_{\varepsilon}. \]

Continuous time modelling of discrete time data:

The quadratic estimator of the volatility fails when the data is sampled at the highest frequencies.

In the absence of market microstructure:

\[ \lim_{\Delta t \to 0} [X,X]_T = \int_0^T \sigma_t^2 \, dt, \]

in probability, where

\[ [X,X]_T = \sum_{t_i} (X_{t_{i+1}} - X_{t_i})^2. \quad (219) \]
On the other hand, for the noisy measurements:

\[
\lim_{\Delta t \to 0} [Y,Y]_T = 2N\sigma^2 + O(N^{1/2}).
\]

\[
\frac{1}{2N} \lim_{\Delta t \to 0} [Y,Y]_T
\]
provides us with an estimate for the variance of the noise!

The standard estimator for the integrated volatility gives us the wrong result.

We need to ignore high frequency data.

This is called the "fifth best estimator" in Ait-Sahalia et al.
Fourth best estimator: sample at an arbitrary sparse frequency \( n_s \):

\[
[Y, Y]_T^{sp} \approx \langle X, X \rangle_T + 2n_s \sigma^2_\varepsilon + C(\varepsilon, \sigma_t) Z, \quad Z \sim \mathcal{N}(0, 1).
\]

Third best estimator: sample at the optimal sparse frequency by minimizing the mean squared error:

\[
n_{opt} = \left( \frac{T}{4(\sigma^2_\varepsilon)^2} \int_0^T \sigma^4_t \, dt \right)^{1/3}
\]

We are using small sample sizes, the variance can be quite large (use one out every 300 observations).
Second best estimator: average over the data (moving average). Averaging over $K$ grids of average size $\bar{n}$ we have

$$[Y, Y]_{T}^{avg} \approx \langle X, X \rangle_{T} + 2\bar{n}\sigma_{\varepsilon}^{2} + C(\varepsilon, \sigma_{t}) Z, \quad Z \sim \mathcal{N}(0, 1).$$

The optimal choice of $\bar{n}$ is

$$n_{opt} = \left(\frac{T}{6(\sigma_{\varepsilon}^{2})^{2}} \int_{0}^{T} \sigma_{t}^{4} dt\right)^{1/3}.$$ 

The first best estimator: subsampling, averaging and bias-correction. Use the fifth best estimator to estimate the noise and subtract it off from the fourth best estimator.

$$\langle X, \hat{X} \rangle_{T} = [Y, Y]_{T}^{avg} - \frac{\bar{n}}{n} [Y, Y]_{T}^{all}.$$ 

This estimator is unbiased

$$\langle X, \hat{X} \rangle_{T} = \langle X, X \rangle_{T} + n^{-1/6} C(\varepsilon, \sigma_{t}) Z, \quad Z \sim \mathcal{N}(0, 1).$$
Thermal Motion in a Two-Scale Potential


Consider the SDE

\[ dx^\varepsilon(t) = -\nabla V \left( x^\varepsilon(t), \frac{x^\varepsilon(t)}{\varepsilon}; \alpha \right) dt + \sqrt{2\sigma} dW(t), \]

Separable potential, linear in the coefficient \( \alpha \):

\[ V(x, y; \alpha) := \alpha V(x) + p(y). \]

\( p(y) \) is a mean-zero smooth periodic function.

\[ x^\varepsilon(t) \Rightarrow X(t) \] weakly in \( C([0, T]; \mathbb{R}^d) \), the solution of the homogenized equation:

\[ dX(t) = -\alpha K \nabla V(X(t)) dt + \sqrt{2\sigma K} dW(t). \]
Figure: Bistable potential with periodic fluctuations
The coefficients $A$, $\Sigma$ are given by the standard homogenization formulas.

Goal: fit a time series of $x^\varepsilon(t)$, the solution of (??), to the homogenized SDE.

Problem: the data is not compatible with the homogenized equation at small scales.

Model misspecification.

Similar difficulties when studying inverse problems for PDEs with a multiscale structure.
In one dimension

\[ dx(\varepsilon(t)) = -\alpha V'(x(\varepsilon(t)))dt - \frac{1}{\varepsilon} p'(x(\varepsilon(t))) dt + \sqrt{2\sigma} dW(t). \]

The homogenized equation is

\[ dX(t) = -AV'(X(t))dt + \sqrt{2\Sigma} dW(t). \]

\((A, \Sigma)\) are given by

\[ A = \frac{\alpha L^2}{\hat{Z} \hat{Z}}, \quad \Sigma = \frac{\sigma L^2}{\hat{Z} \hat{Z}}, \quad Z = \int_0^L e^{-\frac{p(y)}{\sigma}} dy, \quad \hat{Z} = \int_0^L e^{\frac{p(y)}{\sigma}} dy. \]

\(A\) and \(\Sigma\) decay to 0 exponentially fast in \(\sigma \to 0\).

The homogenized coefficients satisfy (detailed balance):

\[ \frac{A}{\alpha} = \frac{\Sigma}{\sigma}. \]
We are given a path of

\[ dx^\epsilon(t) = -\alpha V'(x^\epsilon(t)) \, dt - \frac{1}{\epsilon} p' \left( \frac{x^\epsilon(t)}{\epsilon} \right) \, dt + \sqrt{2\sigma} \, d\beta(t). \]

We want to fit the data to

\[ dX(t) = -\hat{A} V'(X(t)) \, dt + \sqrt{2\hat{\Sigma}} \, d\beta(t). \]

It is reasonable to assume that we have some information on the large-scale structure of the potential \( V(x) \).

We do not assume that we know anything about the small scale fluctuations.
We fit the drift and diffusion coefficients via maximum likelihood and quadratic variation, respectively.

For simplicity we fit scalars $A, \Sigma$ in

$$dx(t) = -A \nabla V(x(t)) dt + \sqrt{2\Sigma} dW(t).$$

The Radon–Nikodym derivative of the law of this SDE wrt Wiener measure is

$$\mathbb{L} = \exp \left( -\frac{1}{\Sigma} \int_0^T A \nabla V(x) \, dx(s) - \frac{1}{2\Sigma} \int_0^T |A \nabla V(x(s))|^2 \, ds \right).$$

This is the maximum likelihood function.
Let $x$ denote $\{x(t)\}_{t \in [0,T]}$ or $\{x(n\delta)\}_{n=0}^{N}$ with $n\delta = T$.

Diffusion coefficient estimated from the quadratic variation:

$$\hat{\Sigma}_{N,\delta}(x) = \frac{1}{dN\delta} \sum_{n=0}^{N-1} |x_{n+1} - x_n|^2,$$

Choose $\hat{A}$ to maximize $\log \mathbb{L}$:

$$\hat{A}(x) = -\frac{\int_{0}^{T} \langle \nabla V(x(s)), dx(s) \rangle}{\int_{0}^{T} |\nabla V(x(s))|^2 \, ds}$$
In practice we use the estimators on discrete time data and use the following discretisations:

\[
\hat{\Sigma}_{N, \delta}(x) = \frac{1}{N\delta} \sum_{n=0}^{N-1} |x_{n+1} - x_n|^2,
\]

\[
\hat{A}_{N, \delta}(x) = -\frac{\sum_{n=0}^{N-1} \langle \nabla V(x_n), (x_{n+1} - x_n) \rangle}{\sum_{n=0}^{N-1} |\nabla V(x_n)|^2 \delta},
\]

\[
\tilde{A}_{N, \delta}(x) = \hat{\Sigma}_{N, \delta} \frac{\sum_{n=0}^{N-1} \Delta V(x_n) \delta}{\sum_{n=0}^{N-1} |\nabla V(x_n)|^2 \delta},
\]
No Subsampling

- Generate data from the unhomogenized equation (quadratic or bistable potential, simple trigonometric perturbation).
- Solve the SDE numerically using Euler–Marayama for a single realization of the noise. Time step is sufficiently small so that errors due to discretization are negligible.
- Fit to the homogenized equation.
- Use data on a fine scale $\delta \ll \varepsilon^2$ (i.e. use all data).
- Parameter estimation fails.
Figure: $\hat{A}, \hat{\Sigma}$ vs $\varepsilon$ for quadratic potential.
Figure: $\hat{A}$, $\hat{\Sigma}$ vs $\sigma$ for quadratic potential with $\varepsilon = 0.1$. 
Subsampling

- Generate data from the unhomogenized equation.
- Fit to the homogenized equation.
- Use data on a coarse scale $\varepsilon^2 \ll \delta \ll 1$.
- More precisely

$$\delta := \Delta t_{sam} = 2^k \Delta t, \quad k = 0, 1, \ldots.$$

- Study the estimators as a function of $\Delta t_{sam}$.
- Parameter Estimation Succeeds.
Figure: $\hat{A}$, $\hat{\Sigma}$ vs $\Delta t_{sam}$ for quadratic potential with $\varepsilon = 0.1$. 
Figure: $\hat{A}, \hat{B}$ vs $\Delta t_{sam}$ for bistable potential with $\sigma = 0.5$, $\varepsilon = 0.1$. 
Figure: $\hat{B}_{ij}, i,j = 1, 2$ vs $\Delta t_{sam}$ for 2d quadratic potential with $\sigma = 0.5, \varepsilon = 0.1$. 
Conclusions From Numerical Experiments

- Parameter estimation fails when we take the small–scale (high frequency) data into account.
- $\hat{A}$, $\hat{\Sigma}$ become exponentially wrong in $\sigma \to 0$.
- $\hat{A}$, $\hat{\Sigma}$ do not improve as $\varepsilon \to 0$.
- Parameter estimation succeeds when we subsample (use only data on a coarse scale).
- There is an optimal sampling rate which depends on $\sigma$.
- Optimal sampling rate is different in different directions in higher dimensions.
Theorem (No Subsampling)

Let \( x^\varepsilon(t) : \mathbb{R}^+ \mapsto \mathbb{R}^d \) be generated by the unhomogenized equation. Then

\[
\lim_{\varepsilon \to 0} \lim_{T \to \infty} \hat{A}(x^\varepsilon(t)) = \alpha, \quad \text{a.s.}
\]

Fix \( T = N\delta \). Then for every \( \varepsilon > 0 \)

\[
\lim_{N \to \infty} \Sigma_{N,\delta}(x^\varepsilon(t)) = \sigma, \quad \text{a.s.}
\]

Thus the unhomogenized parameters are estimated – the wrong answer.
Theorem (With Subsampling)

Fix $T = N\delta$ with $\delta = \varepsilon^\alpha$ with $\alpha \in (0, 1)$. Then

$$\lim_{\varepsilon \to 0} \hat{\Sigma}_{N,\delta}(x^\varepsilon) = \Sigma \quad \text{in distribution}.$$ 

Let $\delta = \varepsilon^\alpha$ with $\alpha \in (0, 1)$, $N = \lfloor \varepsilon^{-\gamma} \rfloor$, $\gamma > \alpha$. Then

$$\lim_{\varepsilon \to 0} \hat{A}_{N,\delta}(x^\varepsilon) = A \quad \text{in distribution}.$$ 

Thus we get the right answer provided subsampling is used.