Stochastic Multiscale Modeling

lecture notes by Gil Ariel
Summer School on Multiscale Modeling and Analysis
University of Texas at Austin
August, 2008

Important note (disclaimer)

The purpose of theses lectures is to introduce students to some of the interesting phenomena observed with stochastic processes which are evolving on two widely separated time scales. The study, and even definition of such processes involve a wide range of necessary concepts and tools from the theory of probability and stochastic processes. With the short time at our disposal, these notes are aimed to highlight a line of reasoning so that a student with some basic knowledge of elementary probability can get a feeling, and perhaps some intuition, on the complicated subject of multiscale stochastic processes. As such, these notes are not to be regarded as a summary or introductory probability course. I apologize for any inaccuracies. Some may be unintentional. However, some are, in order to facilitate a smoother presentation. None, I hope, are too gross. I will do my best to point out whenever I allowed myself some leniencies which cannot be justified, at least formally.

Finally, these notes should be regarded as a list of key definitions, theorems and practice problems and are supplementary to the class. They were not written as a self-study guide or review.

Lecture 1 - Basic definitions

Definition: The *power set* of a given set A, denoted as $\mathcal{P}(A)$, is the set of all subsets of A, i.e., $\mathcal{P}(A) = \{B | B \subset A\}$.
Definition: A *probability space* is a triplet $(\Omega, \mathcal{F}, P)$, where Ω is a set of "outcomes", $\mathcal{F} \subset \mathcal{P}(\Omega)$ is a set if events, and $P : \mathcal{F} \to [0, 1]$ is a function that assigns probabilities to events. P satisfies:

1. For all $A \in \mathcal{F}, P(A) \geq 0$
2. If $A_i \in \mathcal{F}$ is a finite or countable sequence of disjoint sets, then $P(\cup_i A_i) = \sum_i P(A_i)$.
3. $P(\phi) = 0$ and $P(\Omega) = 1$.

where ϕ is the empty set.

Note: to be precise, $\mathcal{F}$ needs to be a σ-field ($\phi \in \mathcal{F}$, if $A \in \mathcal{F}$ then $A^c \in \mathcal{F}$, if $A_i \in \mathcal{F}$ for $i \geq 1$ then $\cup_{i \geq 1} A_i \in \mathcal{F}$). This requirement will not be discussed any further.
Examples:

- Finite space. $\Omega = \{1, 2, 3\}$, $\mathcal{F} = \mathcal{P}(\Omega)$, some P.
- Discrete space. $\Omega = \mathcal{N}$ (natural numbers), $\mathcal{F} = \mathcal{P}(\mathcal{N})$, $P(A) = \sum_{k \in A} 6/(\pi^2 k^2)$.
- Continuous space. $\Omega = \mathcal{R}$ (real numbers), $\mathcal{F}$ the Borel σ-field (includes countable unions of all open and close segments) and P is defined using $P([a, b]) = (2\pi)^{-1/2} \int_a^b e^{-x^2/2} dx$. *Normal distribution.*

Definition: We say that an event happens *almost surely* if it has probability one.
Ad hoc definition: For this class, we will think of a *random variable* (abbreviated RV) as a probability on $\mathcal{R}$, i.e., $\Omega = \mathcal{R}$. We will not discuss what $\mathcal{F}$ is (the Borel σ-field).
Definition: We say that a continuous RV has *probability density* $p(x)$ if $P([a, b]) = \int_a^b p(x) dx$ for all $a < b \in \mathcal{R}$
Example: Gaussian distribution.
$X \sim G(m, \sigma^2)$ has density $p(x) = (2\pi\sigma)^{-1/2} e^{-(x-m)^2/2\sigma^2}$.
Notation: The *average*, or *mean* of a RV X is denoted $\mathbb{E}X$. For a continuous RV with density $p(x)$, $\mathbb{E}X = \int_{\mathcal{R}} x p(x) dx$.

Definition: The Variance of a RV is defined as $\mathrm{Var}X = \mathbb{E}[X^2] - (\mathbb{E}X)^2$. For a continuous RV with density $p(x)$, $\mathrm{Var}X = \int_{\mathcal{R}} x^2 p(x)dx - (\mathbb{E}X)^2$.
Ad hoc definition: Two random variables X and Y are called *independent* if the outcome of X does not depend on the outcome of Y, and vice-versa. In particular, for all (continuous) function f and g, $\mathbb{E}[f(X)g(Y)] = \mathbb{E}[f(X)]\mathbb{E}[g(Y)]$.
Homework: Show that a linear combination of independent Gaussian RVs is a Gaussian RV (use characteristic functions). Calculate the mean and variance of the linear combination.
Definition: *Multivariate RVs* are probabilities on R^n.
Example: multivariate Gaussian distribution.
$\mathbf{X} \sim G(\mathbf{m}, \Sigma)$, $\mathbf{m} \in \mathcal{R}^n$, Σ a symmetric, positive definite $n \times n$ matrix, has density $p(\mathbf{x}) = ((2\pi)^n |\det \Sigma|)^{-1/2} e^{-(\mathbf{x}-\mathbf{m})^T \Sigma^{-1} (\mathbf{x}-\mathbf{m})/2)}$. Σ is called the covariance matrix.
Definition: The *joint probability* of RVs $X_1, \ldots, X_n$ is the probability distribution of the random vector $(X_1, \ldots, X_n)$.
Definition: The *covariance* matrix of a multivariate RV $X = (X_1, \ldots, X_n)^T$ is $\mathrm{Cov}_X = \mathbb{E}XX^T - \mathbb{E}X\mathbb{E}X^T$.
Homework: Calculate the mean and covariance of a Gaussian RV $G(m, \Sigma)$.
Homework: Let U and V be IID (independent, identically distributed) RVs with uniform distribution on $(0, 1)$. Let $\theta = 2\pi U$, $R = \sqrt{-2 \log V}$, $X = R \cos\theta$ and $Y = R \sin\theta$. Write a program that evaluates moments of the joint distribution of X and Y to order 6 ($\mathbb{E}[X^i Y^j]$, $i + j \leq 4$). Make sure that the relative error of your calculations is smaller than 1%. Show that your calculations confirm that X and Y are actually IID normal. Can you prove that?

Lecture 2 - Stochastic processes and Brownian motion

Definition: Let $(\Omega, \mathcal{F}, P)$ denote a probability space. A *stochastic process* on Ω in $[0,T]$ is a collection of RVs $\{X_t | t \in [0,T]\}$.

Note that X_t at different times do not have to be independent.

Definition: A *sample trajectory*, or *sample path*, of a stochastic process X_t is a particular realization of the process, i.e., pick one possible outcome of X_t for all $t \in [0,T]$. Hence, every trajectory is a function from $[0,T]$ to Ω.

With these definitions we can look at stochastic processes from two different, yet equivalent points of view. The first, following our initial definition, is as a collection of inter-dependent RVs defined on some time segment. The second, which is often more intuitive, is as a probability distribution which is defined on some function space. It is as if we are assigning a probability weight for every single trajectory. Of course, since the set of possible trajectories is usually uncountable, we actually need a probability density. Unfortunately, a proper definition of such a density is beyond the scope of these lectures.

Definition: Let $t_1, \ldots, t_k \in [0,T]$ denote k sample times. A multivariate RV $(X_{t_1}, \ldots, X_{t_k})$ is called a *marginal* of the process X_t.

Definition: The mean of a process X_t is a function $\mathbb{E}X_t : [0,T] \to \mathcal{R}$ given by the mean at time t.

Definition: The covariance of a process X_t is a function $\mathrm{Cov}_X : [0,T] \times [0,T] \to \mathcal{R}$ given by $\mathrm{Cov}_X(t,s) = \mathbb{E}[X_t X_s] - \mathbb{E}[X_t]\mathbb{E}[X_s]$.

Definition: The variance of a process X_t is a function $\mathrm{Var}_X : [0,T] \to \mathcal{R}$ given by $\mathrm{Var}_X(t) = \mathrm{Cov}_X(t,t) = \mathbb{E}[X_t^2] - (\mathbb{E}[X_t])^2 = \mathrm{Var}X_t$.

Definition: We say that two *stochastic processes are equal* (more precisely, they have the same distribution) if all their marginals have the save distribution. Note that we need to check all possible combinations of any finite number of sample times.

Definition: A stochastic process is called *Gaussian* if all its marginal are Gaussian RVs.

Example: (not very useful) X_t are IID normal for all $t \in [0,1]$.

Definition: We say that a stochastic process X_t has *independent increments* if, for all $0 < t_1 < \cdots < t_n$, the RVs $X_{t_1} - X_0, \ldots X_{t_n} - X_{t_{n-1}}$ are independent.

Lemma: A Gaussian process with independent increments is completely characterized by its mean and covariance (this means that if you know the mean and covariance then you can calculate all marginals).

Homework: Prove the previous Lemma (show that if two Gaussian processes have the same mean and covariance functions then they are equal).
Definition: *Brownian motion* (BM), also called the *Wiener process*, is a stochastic process B_t that satisfies the following properties:

1. $B_0 = 0$.
2. B_t has independent increments.
3. For all $s \geq t \geq 0$, $B_s - B_t \sim G(0, s-t)$.
4. Trajectories are almost surely continuous.

From now on we will denote BM by B_t.
Theorem: There exists a BM process.
Exercise: Prove that BM is a Gaussian process. Calculate its mean (zero) and covariance ($\min\{t, s\}$). By our definition of equality, this proves uniqueness.
Lemma: Let X_t denote a stochastic process satisfying

1. $X_0 = 0$.
2. X_t is Gaussian.
3. X_t has zero mean and covariance $\mathrm{Cov}_X(s,t) = \min\{s,t\}$.
4. Trajectories are almost surely continuous.

Then, $X_t = B_t$.
Homework: Prove that $B_t = -B_t$.
Homework: Let, $X_0 = 0$ and $X_t = tB_{1/t}$ for $t > 0$. Prove that $X_t = B_t$.
How to simulate BM: For example, on $[0,1]$. Choose a grid spacing h. $t_k = hk$. Define $B^h_{t_k}$ inductively as $B^h_0 = 0$, $B^h_{t_k} = B^h_{t_{k-1}} + h\xi_k$, when ξ_k are IID normal. Finally, do a linear interpolation between points $B^h_{t_k}$.
Lemma: $\mathbb{E}[B_t - B^h_t] = 0$ and $\mathrm{Var}[B_t - B^h_t] \leq Ch$ for some constant $C > 0$. In other words, the two processes: exact BM – B_t and the discrete approximation – B^h_t, are close. Warning: ξ_k have to be constructed appropriately from B_t.
Homework: Plot three sample paths of BM on $[0,1]$ using $h = 10^{-3}$. Plot $\mathbb{E}B^h_t$ and $\mathrm{Var}B^h_t$. How many samples do you need to take? How do you evaluate the errors.

Theorem: With probability 1, Brownian paths are not Lipschitz continuous (and therefore not differentiable) at any point. An elementary proof can be found in Durrett, Chapter 7, Theorem 1.7.

Markov property: BM satisfied the Markov property. Loosely speaking, this means that given the present, the future is independent of the past. This is due to the independent increments property.

Lecture 3 - Stochastic differential equations

Definition: A *stochastic differential equations* is a differential equation of the form

$$dX_t = a(t, B_t, X_t)dt + \sigma(t, B_t, X_t)dB_t. \tag{1}$$

We define the process X_t satisfying (1) using discretization: approximate the differential dt by a finite step size h and dB_t by $\Delta B_t = B_{t+h} - B_t$. Let $t_n = hn$, $X_n = X_{t_n}$, $B_n = B_{t_n}$ and $\Delta B_n = B_{n+1} - B_n$. Then, a forward Euler approximation of (1) is

$$X_{n+1} = X_n + a(t_n, B_n, X_n)h + \sigma(t_n, B_n, X_n)\Delta B_n, \tag{2}$$

with initial condition $X_0 = x_0$. Recalling the properties of BM, ΔB_n are IID $G(0, h)$. Hence, (2) can be rewritten as

$$X_{n+1} = X_n + ha(t_n, B_n, X_n) + \sqrt{h}\sigma(t_n, B_n, X_n)\xi_n, \tag{3}$$

where ξ_n are IID normal.

Theorem: Under appropriate conditions for a and σ, solutions of (3) converge (with the L^2 norm) in the limit $h \to 0$. The (unique) limit is defined to be the solution of (1). For a proof see Oksendal, Chapter 3.

Theorem: The process X_t defined by (1) satisfies the Markov property (because ξ_n are independent).

Example: The *Ornstein-Uhlenbeck* (OU) process is defined as the solution of the linear SDE

$$dX_t = -aX_t dt + bdB_t. \tag{4}$$

with initial condition $X_0 = x_0$. Here, $a > 0$ and b are constants.

Homework: Plot two sample trajectories of (4) in $[0, 3]$ with $a = b = 1$ for each of the initial conditions $x_0 = 0, 1, -1$.

Exercise: Explain why OU is a Gaussian process.

Alternative notation for SDE:

$$\dot{X}_t = a(X_t) + \sigma(X_t)\dot{B}_t. \tag{5}$$

$\dot{B}_t$ is called *white noise*.

Alternative notation for SDE: *Ito integral*

$$X_t = \int_0^t a(X_\tau)d\tau + \int_0^t \sigma(X_\tau)dB_\tau. \tag{6}$$

We need to define the second integral. Write as a Riemann sum but use only the left most point in each segment. Why left? So it is Markov. This definition is identical to the previous one for SDEs using forward Euler (2).

Ito's Lemma (Ito's isometry) 1:

$$\mathbb{E}\left[\int_s^t f(\tau, B_\tau) dB_\tau\right] = 0. \tag{7}$$

Formally, this reads $\mathbb{E}dB_t = 0$.

Proof: write the Riemann sum using the leftmost point.

Ito's Lemma (Ito's isometry) 2:

$$\mathbb{E}\left[\left(\int_s^t f(\tau, B_\tau) dB_\tau\right)^2\right] = \int_s^t \mathbb{E}f^2(t, B_t) d\tau. \tag{8}$$

Formally, this reads $\mathbb{E}dB_t dB_s = \delta(t-s)dtds$.

Proof: write the Riemann sum using the leftmost point.

Exercise: Calculate the mean and covariance of OU.

Homework: Plot the mean and variance of OU in $[0, 3]$ with $a = b = 1$ for each of the initial conditions $x_0 = 0, 1, -1$. Estimate your errors.

Exercise: Show that OU converges as $t \to \infty$ to an invariant (equilibrium) distribution. Find it.

What does invariant distribution means. Ergodicity.

Homework: Plot three trajectories, mean and variance of $dX_t = (1+B_t)dt + e^{B_t}dB_t$.

Theorem: *(the Ito formula).* Let $dX_t = udt + vdB_t$ and $Y_t = g(t, X_t)$. Then, Y_t satisfies

$$dY_t = \frac{\partial g(t, X_t)}{\partial t}dt + \frac{\partial g(t, X_t)}{\partial x}dX_t + \frac{1}{2}\frac{\partial^2 g(t, X_t)}{\partial x^2}(dX_t)^2, \tag{9}$$

where $dt^2 = 0$, $dtdB_t = 0$ and $dB_t^2 = dt$.

Formal proof: expand Y_t in a Taylor series up to order two. Use the Ito isometries.

Example: Solve $dX_t = B_t dB_t$.

Solution: take $Y_t = B_t$ and find $g(x)$ such that $X_t = g(Y_t)$ satisfies the required SDE.

Lecture 4 - Multiscale SDEs: deterministic limit example.

Let $0 < \epsilon \ll 1$ and consider the SDE

$$\begin{cases} dX_t^\epsilon = & -Y_t^\epsilon dt \\ dY_t^\epsilon = & -\frac{1}{\epsilon}(Y_t^\epsilon - X_t^\epsilon)dt + \frac{1}{\sqrt{\epsilon}}dB_t, \end{cases} \tag{10}$$

with initial conditions $X_0 = x$ and $Y_0 = y$.

Called a singular perturbation form.

Change variables $t = \epsilon s$ to see that the dynamics of Y_t^ϵ is on the ϵ time scale. Accordingly, we say that Y_t^ϵ is fast while X_t^ϵ is slow.

What does Y_t^ϵ do? Suppose X_t^ϵ is fixed. Without the noise, Y_t relaxes to a value close to X_t^ϵ exponentially fast on a time scale of order ϵ. To this we add a (not so small) noise.

More precisely, with fixed X_t^ϵ, Y_t^ϵ is an OU process. No matter what the initial condition is, Y_t^ϵ quickly converges (on an ϵ time scale) to an asymptotic distribution, which is $G(X_t^\epsilon, 1/2)$. As a result, it makes sense to assume that Y_t^ϵ is always (to order ϵ) at this asymptotic distribution, $Y_t \sim G(X_t^\epsilon, 1/2)$.

Now look at X_t^ϵ. We discretize the equation using forward Euler with a macroscopic step size H that is independent of ϵ. This yields

$$\begin{aligned} X_{n+1} =& X_n - HY_n = X_n - H(X_n + \frac{1}{\sqrt{2}}\xi_n) \\ =& (1-H)X_n + H\frac{1}{\sqrt{2}}\xi_n, \end{aligned} \tag{11}$$

where ξ_n are IID normal. Note that ξ_n is multiplied by H rather than the usual $\sqrt{H}$ obtained in the discretization of SDEs. This is not a typo. Now, equation (11) is linear. Its solution is

$$X_n = x(1-H)^n + \frac{H}{\sqrt{2}}\sum_{j=1}^{n-1}(1-H)^j\xi_{n-j}. \tag{12}$$

We show that the sum is small:

$$\begin{aligned}
\mathbb{E}[X_n - x(1-H)^n] =& \frac{h}{\sqrt{2}} \sum_{j=1}^{n-1} (1-H)^j \mathbb{E}\xi_{n-j} = 0 \\
\mathrm{Var}[X_n - x(1-H)^n] =& \mathbb{E}[(X_n - x(1-H)^n)^2] \\
=& \frac{H^2}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} (1-H)^{i+j} \mathbb{E}[\xi_{n-i}\xi_{n-j}] \\
=& \frac{H^2}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} (1-H)^{i+j} \delta_{ij} \leq \frac{H^2}{2} n \leq TH/2,
\end{aligned} \tag{13}$$

where we used $n \leq T/H$. Therefore, on a fixed time segment of length T, independent of ϵ, the solution for X_t is close to $X_n = x(1-H)^n$.
Exercise: Show that X_n above is the forward Euler approximation of the ordinary differential equation $\dot{X}_t = -X_t$, $X_0 = x$.

The above derivation can be summarized by saying that in the limit $\epsilon \to 0$, the slow stochastic process X_t^ϵ is well approximated by an effective *deterministic* equation $\dot{X}_t = -X_t$, $X_0 = x$.
Homework: (a) Plot three trajectories of X_t^ϵ along with the effective approximation X_t in $[0,1]$ with $x = 1$ and $\epsilon = 10^{-2}$. (b) Plot the mean and variance of the error $E_t^\epsilon = X_t^\epsilon - X_t$ (you need to run many trajectories and calculate the mean and variance of the errors). (c) Plot $\mathrm{Var}E_t^\epsilon$ as a function of ϵ (should be decreasing linearly).

The above example was simple, since the asymptotic distribution of Y_t with fixed X_t could be calculated analytically. This is not usually the case. However, there is still a lot we can do as long as Y_t has an asymptotic distribution, even though we may not know what it is.

The idea is to simulate Y_t for a while, keeping X_t fixed. Once Y_t reached it's equilibrium distribution, we simply pick one instance of Y_t and use it in the slow equation for X_t. Let us consider a general equation of the form

$$\begin{cases} dX_t = & a(X_t, Y_t)dt + b(X_t)dB_t \\ dY_t = & \frac{1}{\epsilon} c(X_t, Y_t)dt + \frac{1}{\sqrt{\epsilon}} d(X_t, Y_t)dB_t, \end{cases} \tag{14}$$

with initial conditions $X_0 = x$ and $Y_0 = y$. The algorithm can be summarized as follows

Stochastic HMM:

1. initialization: $X_0 = x$, $Y_0 = y$.

2. micro-solver: Let $X = X_t$. Advance Y_t a total of k time steps with size h. The step size h is of order ϵ. Using forward Euler, this implies solving
$$Y_{n+1} = Y_n + \frac{h}{\epsilon} c(X, Y_n) + \sqrt{\frac{h}{\epsilon}} d(X, Y_n)\xi_n, \tag{15}$$
for $k = 0, \ldots, k-1$. k needs to be large enough so that Y_0 and Y_k are practically independent. Denote $Y = Y_k$.

3. Macro-solver: Fix $Y_t = Y$. Advance X_t a single time step with size H. The step size H is independent of ϵ. Using forward Euler, this implies solving
$$X_{n+1} = X_n + Ha(X_n, Y) + \sqrt{H} b(X_n)\xi_l. \tag{16}$$

4. Repeat steps 2 and three with the new values for X_{n+1} and Y_k as initial conditions.

Homework: (a) Plot three trajectories of the HMM approximation for X_t for example (10) in $[0, 1]$ with $x = 1$ and $\epsilon = 10^{-2}$. (b) Plot the mean and variance of the error $E_t^\epsilon = X_t^\epsilon - X_t$ in the HMM approximation. (c) Show that the variance of the error decreases with decreasing H.

Lecture 5 - Multiscale SDEs: diffusive limit example.

Consider the SDE

$$\begin{cases} dX_t^\epsilon = & \frac{1}{\epsilon} Y_t^\epsilon dt \\ dY_t^\epsilon = & -\frac{1}{\epsilon^2} Y_t^\epsilon dt + \frac{\sqrt{2}}{\epsilon} dB_t, \end{cases} \tag{17}$$

with initial conditions $X_0 = x$ and $Y_0 = 0$.

Change variables $t = \epsilon^2 s$ to see that the dynamics of Y_t^ϵ is on the ϵ^2 time scale. The equilibrium measure for Y_t is normal.

What about X_t? At a first glance, it may seem like the RHS of the equation for X_t is of order $1/\epsilon$ because Y_t is of order one. Following the same line of reasoning from the previous lecture, we discretize the equation for X_t^ϵ using forward Euler with macroscopic step size H independent of ϵ,

$$X_{n+1} = X_n + \frac{H}{\epsilon} Y_n, \tag{18}$$

where $X_n = X_{nH}^\epsilon$ and $Y_n = Y_{nH}^\epsilon$. Since Y_n and Y_{n+1} are separated by a macroscopic time difference, it is safe to assume they are IID with the equilibrium distribution of Y_t. Hence, we have

$$X_{n+1} = X_n + \frac{H}{\epsilon} \xi_n, \tag{19}$$

where ξ_n are IID normal. As a result,

$$X_n = x + \frac{H}{\epsilon} \sum_{i=1}^{n} \xi_i. \tag{20}$$

The mean and variance for X_n is

$$\mathbb{E} X_n = x + \frac{H}{\epsilon} \sum_{i=1}^{n} \mathbb{E}\xi_i = x, \tag{21}$$

and

$$\mathrm{Var} X_n = \mathbb{E}[X_n^2] - x^2 = \frac{H^2}{\epsilon^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \mathbb{E}\xi_i \xi_j = \frac{H^2}{\epsilon^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \delta_{ij} = \frac{H^2 n}{\epsilon^2} \leq \frac{HT}{\epsilon^2}, \tag{22}$$

while diverges as $\epsilon \to 0$.

What is wrong with this picture? Our mistake was that X_n does not actually grow in one big jump of order H, but rather in $O(\epsilon^{-2})$ small steps which are correlated. This fact could be ignored with the scaling considered in the previous lecture since fluctuations were not important anyway. The difference here is the big $1/\epsilon$ pre-factor.

Let us concentrate on a single macroscopic step from X_n to X_{n+1}. Denote the $O(\epsilon^{-2})$ microscopic steps we take with Y as $Y_0^n, \ldots, Y_k^n$, where $kh = H$, $h = O(\epsilon^2)$. Hence,

$$X_{n+1} = X_n + \frac{h}{\epsilon} \sum_{i=1}^{k} Y_i^n. \tag{23}$$

The mean and covariance are

$$\mathbb{E}X_{n+1} = \mathbb{E}X_n + \frac{H}{\epsilon} \sum_{i=1}^{k} \mathbb{E}Y_i^n = \mathbb{E}X_n = x, \tag{24}$$

and

$$\begin{aligned} \mathrm{Var}X_{n+1} =& \mathbb{E}[X_{n+1}^2] - x^2 = \mathbb{E}\left[\left(X_n + \frac{H}{\epsilon}\sum_{i=1}^{k} Y_i^n\right)^2\right] - x^2 \\ =& \mathrm{Var}X_n + \frac{h^2}{\epsilon^2} \sum_{i=1}^{k}\sum_{j=1}^{k} \mathbb{E}Y_i^n Y_j^n, \end{aligned} \tag{25}$$

where we assumed that X_n and Y_i^n are independent (why?). Approximating the sum with integrals,

$$\begin{aligned} \mathrm{Var}X_{n+1} - \mathrm{Var}X_n =& \frac{1}{\epsilon^2} \int_0^H \int_0^H \mathrm{Cov}_Y(t,s) ds dt \\ =& \frac{2}{\epsilon^2} \int_0^H \int_t^H \left(e^{-(s-t)/\epsilon^2} - e^{-(s+t)/\epsilon^2}\right) ds dt. \end{aligned} \tag{26}$$

The exponents above are all small, except in a few segments with length of order ϵ^2. Hence, the inner integral is of order ϵ^2. The outer integral is proportional to H. As a result, the variance is proportional to H, which is independent of ϵ. In conclusion, the steps of X_n are IID Gaussian with zero mean and variance given by (26).

Homework: Calculate the integrals in (26).

Homework: Show numerically that the mean and covariance of X_t^ϵ is indeed close to the approximations (24) and (26).

The discussion above suggests a numerical method for approximating the equation for X_t^ϵ even if it is not possible to obtain the invariant measure of Y_t^ϵ analytically. The general idea is to approximate the mean and covariance of the RHS of the equation for X by a short time simulation of the fast process Y_t. Approximating the variance is simple since the integrals in (25) decay rapidly. Hence, a short simulation of Y_t^ϵ suffices. However, numerical evaluation of the average is more complicated. The reason is that the ϵ^{-1} appearing in (24) requires us to evaluate the average of Y_t^ϵ to order ϵ, which is highly inefficient. A method for overcoming this difficulty is suggested in references 3 and 5. Such methods are beyond the scope of these lectures. Instead, we will make a simplifying assumption, which does not necessarily hold in the general case, that $\mathbb{E}Y_t^\epsilon = 0$.

To summarize, consider the SDE

$$\begin{cases} dX_t^\epsilon = & a(X_t^\epsilon, Y_t^\epsilon)dt + b(X_t^\epsilon)dB_t \\ dY_t^\epsilon = & \frac{1}{\epsilon^2}c(X_t^\epsilon, Y_t^\epsilon)dt + \frac{1}{\epsilon}d(X_t^\epsilon, Y_t^\epsilon)dB_t, \end{cases} \tag{27}$$

with initial conditions $X_0 = x_0$ and $Y_0 = y_0$. We assume that for all x, $\mathbb{E}_y^x[a(x,y)] = 0$, where $\mathbb{E}_y^x[\cdot]$ denotes averaging over the equilibrium measure for y with X_t^ϵ held fixed at x. Then, to order ϵ, the equation for X_t^ϵ is approximated as

$$X_{n+1} = X_n + H\Delta X_n, \tag{28}$$

Where ΔX_n are IID $G(0, \sigma^2)$ with

$$\sigma^2 = \frac{2}{\epsilon^2}\int_0^H \int_t^H \mathrm{Cov}_Y(t,s)dsdt + b^2(X_n). \tag{29}$$

Finally, we note that the inner integral is practically constant. Hence,

$$\sigma^2 = \frac{2H}{\epsilon^2}\int_0^H \mathrm{Cov}_Y(0,s)ds + b^2(X_n). \tag{30}$$

An HMM algorithm:

1. initialization: $X_0 = x$, $Y_0 = y$.
2. micro-solver:

(a) Pick some random initial condition for Y.

(b) Advance Y for L small $h = O(\epsilon^2)$ steps to make sure it is close to the equilibrium distribution.

(c) Advance Y for additional M small $h = O(\epsilon^2)$ steps to obtain a short sample trajectory.

(d) Repeat (a)-(c) to obtain many trajectories.

(e) Calculate the covariance of Y.

Notes: Since the initial condition for Y does not matter, you can simply pick the final state of one simulation as initial condition for the next. Then, there is no need for relaxation except for the first time in every run of the micro solver.

3. Parameter estimation: integrate the covariance numerically to approximate σ^2 of (30).

4. Macro-solver: generate a Gaussian RV with zero mean and variance σ^2. Propagate X_n according to (28).

5. Repeat steps 2-4 to time T.

Efficiency: Since L, M, T and H are all independent of ϵ, the efficiency of the algorithm is also independent of ϵ.
Homework: (a) Plot three trajectories of the HMM approximation for X_t^ϵ for example (17) in $[0, 1]$ with $x = 1$ and $\epsilon = 10^{-2}$. (b) Calculate the mean and variance of X_t in the HMM approximation. Compare with (24) and (26)).

Final assignment

1. Effective *deterministic* dynamics emerging from multiscale *stochastic* systems:

   Consider the system of equations:

$$\begin{cases} dX_t = & \left[-(Y_t + Y_t^3) + \cos(\pi t) + \sin(\sqrt{2}\pi t)\right] dt \\ dY_y = & -\epsilon^{-1}(Y_y + Y_y^3 - X_t)dt + \epsilon^{-1/2} dB_t, \end{cases} \tag{31}$$

   with initial conditions $(X_0, Y_0) = (0, 1)$.

   Integrate the system using the HMM algorithm as explained in lecture 4. Plot the mean and covariance of X_t with $\epsilon = 10^{-4}$.

2. Effective *stochastic* dynamics emerging from multiscale *stochastic* systems:

   Consider

$$\begin{cases} dX_t = & \epsilon^{-1} Y_t \\ dY_y = & -\epsilon^{-2} Y_t^3 dt + \epsilon^{-1}(1 + X_t^2) dB_t, \end{cases} \tag{32}$$

   with initial conditions $(X_0, Y_0) = (0, 1)$. Solve the equation using the HMM algorithm.

3. Effective *stochastic* dynamics emerging from multiscale *deterministic* systems:

   Consider the system of ordinary differential equations:

$$\begin{cases} \dot{x} = & x - x^3 + \frac{4}{90}\epsilon^{-1} y_2 \\ \dot{y}_1 = & 10\epsilon^{-2}(y_2 - y_1) \\ \dot{y}_2 = & \epsilon^{-2}(28 y_1 - y_2 - y_1 y_3) \\ \dot{y}_3 = & \epsilon^{-2}(y_1 y_2 - \frac{8}{3} y_3). \end{cases} \tag{33}$$

   The equations for (y_1, y_2, y_3) are called the Lorenz equations. In the example above, the solution is chaotic. To get a feeling what chaotic means, plot the trajectory of (y_1, y_2, y_3) with any non-zero initial condition and some small ϵ.

   Now, consider the SDE

$$dX_t = (X_t - X_t^3)dt + \sigma dB_t. \tag{34}$$

where $\sigma = 0.126$. We wish to show that, for small ϵ, X_t approximates trajectories of $x(t)$. It is still not clear what this statement means since, for any given initial conditions, $x(t)$ is a particular deterministic solution, while X_t is a stochastic process.

Let P denote any distribution on $\mathcal{R}^3$ with a continuous density, for example, each coordinate is IID normal. Plot several sample trajectories of $x(t)$ with $\epsilon = 10^{-3}$. Initial conditions are $x(0) = 1$ and $(y_1(0), y_2(0), y_3(0))$ drawn from the distribution P. Plot the mean and variance of $x(t)$ on a macroscopic time scale independent of ϵ. Here, randomness comes from initial conditions. Compare to that of X_t.

Repeat this example with two additional distributions on $\mathcal{R}^3$ as initial conditions for (y_1, y_2, y_3). Show (through numerical examples) that after an initial relaxation time, the statistics of $x(t)$ does not depend on the initial distribution.

Try to explain how come the dynamics of x appears random. In particular, explain why we have to assume that the initial conditions for (y_1, y_2, y_3) are random although it is not too important what this distribution is.

4. Effective *deterministic* dynamics emerging from multiscale *deterministic* systems:

   Richard Tsai's class.

References and further reading

1. Probability: Theory and Examples, by Richard Durrett.

2. Stochastic Differential Equations - an Introduction with Applications, by Bernt Oksendal.

3. Weinan E, Di Liu and Eric Vanden-Eijnden, Analysis of multiscale methods for stochastic differential equations, *Commun. Pure Appl. Math.* **63**, 1544 (2005).

4. Dror Givon, Raz Kupferman and Andrew Stuart, Extracting macroscopic dynamics: model problems and algorithms, *Nonlinearity* **17**, R55 (2004).

5. Eric Vanden-Eijnden, Numerical Techniques for multi-scale dynamical systems with stochastic effects, *Comm. Math. Sci.* **1**, 385 (2003).