STOCHASTIC AVERAGING
WITH A FLATTENED HAMILTONIAN
A MARKOV PROCESS ON A LOLLIPOP

Richard Sowers
Department of Mathematics
University of Illinois
Urbana-Champaign
Outline

(1) Classical problem
(2) Graph-valued problem
(3) Lollipop problem
(4) Stability of Averaging
Let’s start with the equation of a 1-dimensional particle in a force field:

\[ \ddot{x}_t = F(x_t) \]

where \( F \in C^\infty(\mathbb{R}) \). This can be written as a second-order ODE

\[
\begin{align*}
\dot{x}_t^0 &= y_t^0 \\
\dot{y}_t^0 &= F(x_t^0).
\end{align*}
\]

We can use the idea of conservation of energy to define

\[
H(x, y) \overset{\text{def}}{=} U(x) + \frac{y^2}{2}
\]

where \( \dot{U}(x) = -F(x) \). Then

\[
\begin{align*}
\dot{x}_t^0 &= \frac{\partial H}{\partial y}(x_t^0, y_t^0) \\
\dot{y}_t^0 &= -\frac{\partial H}{\partial x}(x_t^0, y_t^0)
\end{align*}
\]

and

\[
\frac{d}{dt} H(x_t^0, y_t^0) = \frac{\partial H}{\partial x} \frac{\partial H}{\partial y} - \frac{\partial H}{\partial y} \frac{\partial H}{\partial x}(x_t^0, y_t^0) = \{H, H\}(x_t^0, y_t^0) = 0.
\]

Thus the energy \( H(x_t^0, y_t^0) \) is conserved.
Let’s assume for simplicity that the potential $U$ has a single well, viz.

$$U(x) = \frac{1}{2}x^2;$$

thus the force is $F(x) = -x$.

**Single-Well Potential**

Thus

$$H(x, y) = \frac{x^2}{2} + \frac{y^2}{2}$$
We can understand the dynamics of \((x_t^0, y_t^0)\) by looking at the graph of the energy \(H\).

\[
H(x, y) = \frac{x^2}{2} + \frac{y^2}{2}
\]

The ODE for \((x_t^0, y_t^0)\) simply moves around the level sets of \(H\). Let \(x_t^0(x, y)\) and \(y_t^0(x, y)\) be the solution of the particle dynamics with \(x_0^0(x, y) = x\) and \(y_0^0(x, y) = y\). We note that if we set \(h \overset{\text{def}}{=} H(x, y)\), then the measure \(\mu\) given by

\[
\langle f, \mu(x, y) \rangle \overset{\text{def}}{=} \lim_{T \to \infty} \frac{1}{T} \int_0^T f(x_s^0(x, y), y_s^0(x, y))ds
\]

\[
= \frac{\int_{(x', y') : H(x', y') = h} f(x', y') \|\nabla H(x', y')\| \mathcal{H}^1(dx', dy')} \int_{(x', y') : H(x', y') = h} \|\nabla H(x', y')\| \mathcal{H}^1(dx', dy')
\]

depends only on \(h\); i.e., \(\mu(x, y) = \mu_H(x, y)\).
In reality, the world is noisy; the acceleration is subject to random jerk. A more accurate model for the particle is

\[
\begin{align*}
    dx_t^\varepsilon &= \frac{\partial H}{\partial y}(x_t^\varepsilon, y_t^\varepsilon)dt \\
    dy_t^\varepsilon &= -\frac{\partial H}{\partial x}(x_t^\varepsilon, y_t^\varepsilon)dt + \varepsilon dW_t,
\end{align*}
\]

where \( W \) is a Wiener process, and \( \varepsilon \) is a small parameter. Now energy is not conserved, but is slowly-varying; by Ito’s formula,

\[
dH(x_t^\varepsilon, y_t^\varepsilon) = \varepsilon \frac{\partial H}{\partial y}(x_t^\varepsilon, y_t^\varepsilon) dW_t + \varepsilon^2 \frac{1}{2} \frac{\partial^2 H}{\partial y^2}(x_t^\varepsilon, y_t^\varepsilon)dt \tag{Ito correction term}
\]

Thus, we will see a change in energy in time of order \( \varepsilon^{-2} \); let’s look at \((X_t^{\varepsilon^2}, Y_t^{\varepsilon^2})\). This rescaling gives us a new SDE:

\[
\begin{align*}
    dX_t^\varepsilon &= \frac{1}{\varepsilon^2} \frac{\partial H}{\partial y}(X_t^\varepsilon, Y_t^\varepsilon)dt \\
    dY_t^\varepsilon &= -\frac{1}{\varepsilon^2} \frac{\partial H}{\partial x}(X_t^\varepsilon, Y_t^\varepsilon)dt + dW_t.
\end{align*}
\]

Then

\[
dH(X_t^\varepsilon, Y_t^\varepsilon) = \frac{\partial H}{\partial y}(Y_t^\varepsilon, Y_t^\varepsilon) dW_t + \frac{1}{2} \frac{\partial^2 H}{\partial y^2}(X_t^\varepsilon, Y_t^\varepsilon)dt.
\]

What happens to energy dynamics as \( \varepsilon \to 0 \)?
Think of a **tornado**. If we put a tracer particle in a tornado, then we can’t keep track of the angular position of the particle (it moves too fast), but we can see the vertical dynamics of the particle. We can try to find a **reduced model** for the dynamics of the vertical position. Somehow, we should **average** out the angular dynamics to get the vertical dynamics.

---

**Geometry of Classical Case**
Although we could directly look at the dynamics of $H(X_t^\varepsilon, Y_t^\varepsilon)$, let’s instead look at the dynamics of a function of $H(X_t^\varepsilon, Y_t^\varepsilon)$; i.e., we will look at convergence of the **martingale problem** associated with $(X_t^\varepsilon, Y_t^\varepsilon)$ (the martingale problem is analogous to the weak theory of solutions of PDE’s). Fix $f \in C^2_c(\mathbb{R})$. Then

$$f(H(X_t^\varepsilon, Y_t^\varepsilon)) = \int_0^t \left\{ \frac{1}{2} \ddot{f}(H(X_s^\varepsilon, Y_s^\varepsilon)) \left( \frac{\partial H}{\partial y}(X_s^\varepsilon, Y_s^\varepsilon) \right)^2 + \dot{f}(H(X_s^\varepsilon, Y_s^\varepsilon)) \left( \frac{1}{2} \frac{\partial^2 H}{\partial y^2}(X_s^\varepsilon, Y_s^\varepsilon) \right) \right\} ds + M_t^{f,\varepsilon} \quad (*)$$

where $M_t^{f,\varepsilon}$ is a martingale. Two facts help us. First, that (roughly) martingales are stable under limits. Second, that for “nice” functions $\phi$ and $\psi$,

$$\int_0^t \phi(H(X_s^\varepsilon, Y_s^\varepsilon)) \psi(X_s^\varepsilon, Y_s^\varepsilon) ds \approx \int_0^t \phi(H(X_s^\varepsilon, Y_s^\varepsilon)) \langle \phi, \mu_{H(X_s^\varepsilon, Y_s^\varepsilon)} \rangle ds$$

(this is the **averaging lemma**). We can now argue as follows. **Step 1.** We can assume that

$$H(X_t^\varepsilon, Y_t^\varepsilon) \to \underline{h_t}$$

(in a weak functional sense); some easy tightness calculations ensure this at least along a subsequence. **Step 2.** From $(*)$ and the averaging lemma, we have that

$$f(\underline{h_t}) = \int_0^t \left\{ \frac{1}{2} \ddot{f}(h_s) \sigma^2(h_s) + \dot{f}(h_s) b(h_s) \right\} ds + M_t^f \quad (**)$$

where $M$ is a martingale and where

$$\sigma^2(h) \overset{\text{def}}{=} \left\langle \left( \frac{\partial H}{\partial y} \right)^2, \mu_h \right\rangle \quad \text{and} \quad b(h) \overset{\text{def}}{=} \left\langle \frac{1}{2} \frac{\partial^2 H}{\partial y^2}, \mu_h \right\rangle.$$

**Step 3.** The solution of $(**)$ is unique in law and satisfies

$$dh_t = b(h_t) dt + \sigma(h_t) dZ_t$$

where $Z$ is some Brownian motion.
The foregoing is essentially **classical** and was understood in the 1960’s. A natural next problem, solved by Freidlin and Wentzell (1996) and Freidlin and Weber (1998) is that of a **double-well** potential;

\[ U(y) = x^2(x - c_1)(x - c_2) \]

(where \( c_1 \) and \( c_2 \) are positive).

**Double-Well Potential**

The graph of the total energy thus has two wells.

**Double-Well Energy**

If \( h = H(x, y) > 0 \), \( \mu(x, y) \) still depends only \( h \), but if \( h < 0 \), it depends on \( h \) and which well we are interested in (i.e., whether \( x < 0 \) or \( x > 0 \)).

Define an **equivalence relation** on \( \mathbb{R}^2 \); we say that \( (x, y) \sim (x', y') \) if \( \mu(x, y) = \mu(x', y') \). Then set \( \Gamma \equiv \mathbb{R}^2 / \sim \).
Geometry of Double-Well Case

Freidlin-Wentzell and Freidlin-Weber identified a limiting graph-valued process \( \xi_t \overset{\text{def}}{=} \lim_{\varepsilon \to 0} [(X_t^\varepsilon, Y_t^\varepsilon)] \) (where \( [(x, y)] \) is the equivalence class of \( (x, y) \) under \( \sim \)). The limit \( \xi \) is given by the classical calculations as long as \( \xi_t \) does not touch the vertex. When it does, its behavior is governed by some glueing conditions, which are restrictions on the domain of the generator of the limiting process. These glueing conditions are roughly and imprecisely as follows. If \( \xi \) starts at the vertex, one flips a three-sided coin (with statistics given by the glueing conditions) to decided which leg to go to. Then \( \xi \) evolves according to the classical dynamics until it again hits the vertex, where the process is repeated again (with a new coin). Equivalently, the coins describes which of three excursion measures govern the individual excursions from the vertex (a heuristic description); this is similar to skew Brownian motion. It is also related to spider martingales.
Let’s now discuss the new problem. Let’s make two changes to the system.

(1) In both the single-well and double-well problems, the critical points of $H$ are nondegenerate; i.e., if $dH(x, y) = 0$, then $D^2 H(x, y)$ is full rank. We want to violently remove this assumption. We will assume that $H$ has a flat.

(2) Let’s assume that the noise is in both components (to avoid complications). Our “canonical” Hamiltonian will be

$$H(z) \overset{\text{def}}{=} (\max\{||z|| - 1, 0\})^n \quad z \in \mathbb{R}^2$$

where $n > 2$ (so that $H$ has a locally Lipshitz derivative). Then

$$\mathcal{Z} \overset{\text{def}}{=} \{z \in \mathbb{R}^2 : H(z) = 0\} = \{z \in \mathbb{R}^2 : ||z|| \leq 1\}$$

![Flattened Hamiltonian (from below)](image)

As usual, define

$$\tilde{\nabla} H(x, y) \overset{\text{def}}{=} \left( \frac{\partial H}{\partial y}(x, y), -\frac{\partial H}{\partial x}(x, y) \right)$$

and consider the SDE

$$dZ_t^\varepsilon = \frac{1}{\varepsilon^2} \tilde{\nabla} H(Z_t^\varepsilon) dt + dW_t$$

where $W$ is now a two-dimensional Brownian motion.
We consider the orbits
\[ \dot{Z}^0_t(z) = \nabla H(Z^0_t(z)) \]
\[ Z^0_0(z) = z \]  
(*)
and define the measures \( \mu_z \) by
\[ \langle f, \mu_z \rangle \overset{\text{def}}{=} \lim_{T \to \infty} \frac{1}{T} \int_0^T f(Z^0_s(z)) ds. \]

As in the classical case, if \( H(z) > 0 \), then \( \mu_z = \mu_{z'} \) if and only if \( H(z) = H(z') \). On the other hand, if \( H(z) = 0 \), then \( \mu_z = \mu_{z'} \) if and only if \( z = z' \) (each element of \( \mathfrak{g} \) is a critical point of (*)).

**Contour Plot of Flattened Hamiltonian**

Thus, we should collapse each level
\[ \{ z \in \mathbb{R}^2 : H(z) = h \} \]
(***)
to a point if \( h > 0 \). We should also keep the structure of the (open) disk \( \mathfrak{g}_0 \). We should define a point \( * \) as the edge of the disk and the limit of the sets of (**) as \( h \to 0 \). This gives us a downward-pointing lollipop.
Geometry of Flattened Hamiltonian
Our result is

**Theorem.** *The process* \([Z^\xi_t]\) *tends to a lollipop-valued Markov process* \(\xi\) *with a computible generator*

This is geometrically interesting; the lollipop is a **stratified space**.

**Stratified Space**

Locally, the lollipop can look like a line, a plane or the union of a line and plane (a flagpole); the lollipop has a **dimensional discontinuity**. This seems to be one of the first instances of a Markov process whose state space is stratified (there is come coeval work by Burdzy and Bass on a “fiber” Brownian motion).
Before specifying our result, we observe that a similar equation, namely

$$dY_t^\varepsilon = \frac{1}{\varepsilon^2} \nabla H(Y_t^\varepsilon) \circ dW_t^1 + dW_t^2$$

(where $W^1$ and $W^2$ are independent two-dimensional Brownian motions), is a reasonable model for some microscopic phenomena seen in thin-film growth (fast adatom-diffusion on highly-terraced surfaces).

**Thickness = 25 nm**

<table>
<thead>
<tr>
<th>T = 750°C</th>
<th>T = 650 °C</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Image 1" /></td>
<td><img src="image2.png" alt="Image 2" /></td>
</tr>
</tbody>
</table>

**STM images of single-crystal TiN(001)/MgO(001) films**

**Courtesy of Thin Film Group, UIUC**
Let’s be a bit more precise about our result. The lollipop can be written as

\[ L \overset{\text{def}}{=} \beta^c \cup \star \cup (0, \infty) \]

we map \( \{ z \in \mathbb{R}^2 : H(z) = h \} \) (for \( h > 0 \)) into \( h \in (0, \infty) \). The standard averaging results give us that as long as \( Z^\varepsilon \) stays outside of \( \beta \), the limiting process has generator

\[ \mathcal{L}_1 f(h) \overset{\text{def}}{=} \frac{1}{2} \sigma^2(h) \ddot{f}(h) + b(h) \dot{f}(h) \quad f \in C^2(0, \infty) \]

where

\[ \sigma^2(h) = \left\langle \| \nabla H \|^2, \mu_h \right\rangle \quad \text{and} \quad b(h) = \left\langle \frac{1}{2} \Delta H, \mu_h \right\rangle. \]

When the limiting process is inside \( \beta \), it has the same dynamics as the original process (there is no fast drift on \( \beta \)); the generator inside \( \beta \) is thus

\[ \mathcal{L}_2 f(h) = \frac{1}{2} \Delta f(h). \quad f \in C^2(\beta^c) \]

We need a **glueing condition** at the junction \( \star \). We can motivate (but not prove) the glueing condition via the relevant equation for the density (the forward Kolmogorov equation) of the lollipop-valued process (which we for the moment assume to exist). For explanation, we assume that

\[ \mathbb{E}[f(\xi_t)] = \int_{(0, \infty) \subset \mathbb{R}^1} f_1(h)p_1(t, h)dh + \int_{\beta \subset \mathbb{R}^2} f_2(z)p_2(t, z)dz \]

Roughly, the flux through the junction must sum to zero. The flux entering the junction from the handle of the lollipop is one-dimensional. The flux leaving the junction into \( \beta \) is the integral of the flux at the edge of the disk. Thus we should have

\[ -\frac{\partial p_1}{\partial h}(t, 0) = \int_{z \in \partial \beta} \frac{\partial p_2}{\partial n}(t, z)dz \quad (*) \]

(\( \nu \) the inward-pointing unit normal at \( \partial \beta \)). It also seems reasonable that the density is continuous at the junction; i.e.,

\[ p_2(t, \cdot) \big|_{\partial \beta} = p_1(t, 0). \quad (**) \]
In fact, (the adjoint of) these two conditions (*) and (**) are almost sufficient. The domain of the generator of the lollipop-valued process consists of those functions $f$ such that $f_1$ and $f_2$ are $C^2$ up to the junction and such that

1. $f$ is continuous at the vertex
2. we have that
   
   \[-\frac{\partial f_1}{\partial h}(0) = \int_{z \in \partial_3} \frac{\partial f_2}{\partial \nu}(z) dz\]

3. $L_2 f_2(\cdot) \big|_{\partial_3} = L_1 f_1(0)$

For such a function, we define $L f$ as $L_1 f_1$ on $(0, \infty)$ and as $L_2 f_2$ on $\bar{3}$; by (3), continuity allows us to uniquely define $L f(*)$. 
There seem to be many interesting unsolved problems in stochastic averaging. For example, one can ask

**How stable is stochastic averaging?**

We can look for a *Lyapunov exponent* for averaging (related work by Baxendale). Remove critical points of $H$ and look at a **twist map** on the cylinder.

**Twist Map**

Then we can study a “canonical” problem

$$
\begin{align*}
    dX_t^\varepsilon(x) &= \frac{1}{\varepsilon^2} \alpha(Z^\varepsilon_t(x)) \frac{\partial}{\partial \theta}(X^\varepsilon_t(x)) dt + \sigma(X^\varepsilon_t(x)) \circ dW_t \\
    X_0^\varepsilon(x) &= x
\end{align*}
$$

Here $X_0^\varepsilon = (\theta_t^\varepsilon, Z_t^\varepsilon) \in S^1 \times \mathbb{R}$, $\alpha$ is a “nice” coefficient, $\sigma$ is a vector field tangential to the cylinder, and $W$ is a Wiener process (of the appropriate dimension). If we start this SDE at two close points, how will the two points evolve? If the two points are infinitesimally close, we should look at the **tangent flow**

$$
TX_t^\varepsilon(x) = DX_t^\varepsilon(x).
$$

It can be shown (using some results of Pinsky and Wihstutz) that under some nondegeneracy conditions,

$$
\|TX_t^\varepsilon(x)\| \sim \exp \left[ \varepsilon^{-2/3} \int_0^t \lambda(Z_s^\varepsilon(x)) ds \right].
$$