Hamiltonian Systems with Noise: Glue, Spiders, and Lollipops

Richard B. Sowers
Department of Mathematics
University of Illinois
Urbana-Champaign
http://www.math.uiuc.edu/~r-sowers

June 22, 2000
Outline

• Classical problem
• Graph-valued problem
• Lollipop problem
• Directions

Issues

• Convergence of Markov Processes/Semigroups
• Markov processes/PDE’s on stratified spaces
Classical Problem

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 coupled pair of **first-order ODE’s**

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

We can use the idea of **conservation of energy** to define

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

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 \textbf{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 \((x^0_t, y^0_t)\) by looking at the graph of the energy \(H\).

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

The ODE for \((x^0_t, y^0_t)\) simply moves around the level sets of \(H\).

Let \(x^0_t(x, y)\) and \(y^0_t(x, y)\) 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

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

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

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

\[
dx_t^\varepsilon = \frac{\partial H}{\partial y}(x_t^\varepsilon, y_t^\varepsilon)dt
\]

\[
dy_t^\varepsilon = -\frac{\partial H}{\partial x}(x_t^\varepsilon, y_t^\varepsilon)dt + \varepsilon dW_t,
\]

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,

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

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

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

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

Then

\[
dH(X_t^\varepsilon, Y_t^\varepsilon) = \frac{\partial H}{\partial y}(X_t^\varepsilon, Y_t^\varepsilon)dW_t + \frac{1}{2} \frac{\partial^2 H}{\partial y^2}(X_t^\varepsilon, Y_t^\varepsilon)dt.
\]
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
To understand the essence of the proof, let’s 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 $H(X_t^\varepsilon, Y_t^\varepsilon)$. Fix $f \in C^2_c(\mathbb{R})$. Then

$$f(H(X_t^\varepsilon, Y_t^\varepsilon)) = f(H(X_0^\varepsilon, Y_0^\varepsilon)) + \int_0^t \left\{ \frac{1}{2} \ddot{f}(H(X_s^\varepsilon, Y_s^\varepsilon)) \left( \frac{\partial H}{\partial y}(X_s^\varepsilon, Y_s^\varepsilon) \right)^2 + \dot{f}(H(X_s^\varepsilon, Y_s^\varepsilon)) \left( \frac{1}{2} \frac{\partial^2 H}{\partial y^2}(X_s^\varepsilon, Y_s^\varepsilon) \right) \right\} \, ds + M_t^{f,\varepsilon} \quad (1)$$

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

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

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

\[ H(X^e_t, Y^e_t) \to h_t \]

(in a weak functional sense); some easy tightness calculations ensure this at least along a subsequence. **Step 2.** From (1) and the averaging lemma, we have that

\[
f(h_t) = \int_0^t \left\{ \frac{1}{2} \dot{f}(h_s) \sigma^2(h_s) + \dot{f}(h_s) b(h_s) \right\} ds + M_t^f
\]

(2)

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 (2) is unique in law and satisfies

\[
dh_t = b(h_t) dt + \sigma(h_t) dZ_t
\]

where \( Z \) is some Brownian motion.
Graph-valued Problem

The foregoing is essentially classical and was understood in the 1960’s (Khasminskii). A natural next problem, solved by Freidlin and Wentzell (1996) and Freidlin and Weber (1998) (see also Neishtadt (1991)) is that of a double-well potential;

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

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

Double-Well Potential

The graph of the total energy thus has two wells.
Double-Well Potential

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 \( \Gamma \overset{\text{def}}{=} \mathbb{R}^2 / \sim \).
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 (more anon).

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 decide 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). This is similar to skew Brownian motion. It is also related to spider martingales.
In our case,

Each leg has its own elliptic operator $\mathcal{L}_i$. Domain of $\Gamma$-valued generator consists of triplets of $C^2$ functions $f_1$, $f_2$, and $f_3$ such that (here are the glueing conditions)

- $f_i$'s agree at vertex
- $\beta_1 \dot{f}_1(0) = \beta_2 \dot{f}_2(0) + \beta_3 \dot{f}_3(0)$ (a flux-type condition on the $f_i$'s)
- $\mathcal{L}_i f_i$'s agree at vertex
One way (not the original way) to understand the glueing conditions is as follows. The process \((X^\varepsilon, Y^\varepsilon)\) moves between the different wells due to the diffusive kicks (i.e., the martingale part) of \(H(X^\varepsilon, Y^\varepsilon)\). The bracket of this martingale is \((dH, dH)\). Define an angular coordinate via PDE (due to Khasminskii)

\[
(\nabla \Theta(x, y), \nabla H)(x, y) = (dH, dH)(x, y).
\]

Equal increments in \(\Theta\) correspond to equal amounts of \(H\)-bracket. This allows us to change pictures and make a boundary-layer expansion. The glueing conditions are exactly what is needed to solve a certain PDE on this space.

Khasminskii Coordinates for Graph-Valued Problem
New Problem

Let’s make two changes to the system.

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

- 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)

As usual, define

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

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

\[ \dot{Z}_t^0(z) = \nabla H(Z_t^0(z)) \]

\[ Z_0^0(z) = z \]

and define the measures \( \mu_z \) by

\[ \langle f, \mu_z \rangle \overset{\text{def}}{=} \lim_{T \to \infty} \frac{1}{T} \int_0^T f(Z_s^0(z)) \, ds. \]

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 (3)).

Bottom view of \text{ is } \text{ Contour Plot of Flattened Hamiltonian}
Thus, we should collapse each level

$$\{ z \in \mathbb{R}^2 : H(z) = h \}$$

(4)

to a point if $h > 0$. We should also keep the structure of the (open) disk $\mathbb{D}^\circ$. We should define a point $\star$ as the edge of the disk and the limit of the sets of (4) as $h \to 0$. This gives us a downward-pointing lollipop.

Geometry of Flattened Hamiltonian
Our result is

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

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

![Stratified Space](image)

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 some 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).

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

\[ L \overset{\text{def}}{=} \mathcal{J} \circ \ast \cup (0, \infty); \]

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 \( \mathcal{J} \), 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 \( \mathcal{J} \), it has the same dynamics as the original process (there is no fast drift on \( \mathcal{J} \)); the generator inside \( \mathcal{J} \) is thus

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

We need a \textbf{glueing condition} at the junction \( \ast \). 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

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

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 \( \mathbb{R} \) is the integral of the flux at the edge of the disk. Thus we should have

\[ -\frac{\partial p_1}{\partial h}(t, 0) = \int_{z \in \partial_3} \frac{\partial p_2}{\partial \nu}(t, z)dz \quad (5) \]

(\( \nu \) the inward-pointing unit normal at \( \partial_3 \)). It also seems reasonable that the density is continuous at the junction; i.e.,

\[ p_2(t, \cdot)|_{\partial_3} = p_1(t, 0). \quad (6) \]
In fact, (the adjoint of) these two conditions (5) and (6) 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

- \( f \) is continuous at the vertex
- we have that

\[
- \frac{\partial f_1}{\partial h}(0) = \int_{z \in \partial_3} \frac{\partial f_2}{\partial \nu}(z) dz
\]

- \( \mathcal{L}_2 f_2(\cdot) \big|_{\partial_3} = \mathcal{L}_1 f_1(0) \)

For such a function, we define \( \mathcal{L} f \) as \( \mathcal{L}_1 f_1 \) on \((0, \infty)\) and as \( \mathcal{L}_2 f_2 \) on \( \partial_3 \); by (3), continuity allows us to uniquely define \( \mathcal{L} f(\ast) \).

The particle diffuses according to different generators on the stick and ball. At the junction, a coin flip decides where the next excursion should be. An entrance law is required to start on the ball/disk.
Problems

A. General connection between Hamiltonian/classical dynamics and noise.

- Arnol’d Diffusion
- Resonances
- Higher-dimensional systems (★).

B. Stratified Spaces

- Diffusions
- Natural PDE’s (parabolic and elliptic)
- PDE/Probability estimates with a separation of scales (★).
Higher-Dimensional Problems

In the classical and graph-valued problem, we were essentially looking at Newtonian dynamics of a single quantity. Often, several Newtonian quantities can be \textit{coupled}. Energy can then be \textit{transferred} between the variables. Usually there is an obvious energy function which is slowly-varying. Often there is also another conserved quantity (i.e., a symmetry). In this case, the reduced state space will locally have dimension

\[
\sqrt{4} - \sqrt{2} = 2.
\]

![Diagram](image-url)
Probability Estimates with a Separation of Scales

We can ask: How stable is stochastic averaging? We 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.

![Diagram of a twist map]

Twist Map

Then we can study a “canonical” problem

$$dX_t^\varepsilon(x) = \frac{1}{\varepsilon^2} \alpha(Z_t^\varepsilon(x)) \frac{\partial}{\partial \theta}(X_t^\varepsilon(x)) dt + \sigma(X_t^\varepsilon(x)) \circ dW_t$$

$$X_0^\varepsilon(x) = x$$
Here $X_t^\varepsilon = (\theta_t^\varepsilon, Z_t^\varepsilon) \in S^1 \times \mathbb{R}$, $\alpha$ 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,

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

**Harder Problem**

What is corresponding result for double-well problem?