Dissipation from noisy Hamiltonians

CDS205 project — Ramon van Handel (June 6, 2003)

1 Introduction

Dissipative systems are (roughly) systems in which the total energy decreases in time: a vector field X is called dissipative if[1] $\langle dE, X\rangle \leq 0$. Obviously energy is not conserved in such a system, so we have to modify the formalism of geometric mechanics somewhat to take such systems into account. After all, Lagrangian and Hamiltonian systems conserve energy.

Generally dissipation is introduced on the Lagrangian side. A dissipative Lagrangian system is a vector field $Z + Y$, where Z is a Lagrangian vector field (with corresponding Lagrangian L) and Y is a dissipative vector field. One then attempts to find a so-called Rayleigh dissipation function $R(q, \dot{q})$ so the equations of motion can be written (in coordinates) as

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}^i}\right) - \frac{\partial L}{\partial q^i} = -\frac{\partial R}{\partial \dot{q}^i}$$

Unfortunately this method of introducing dissipation may not always work; particularly when studying systems on Lie groups one must use a different approach.

In [2], Bloch et al. tackle the problem of dissipation on Lie groups. The particular type of dissipation considered there is known as "double bracket dissipation", as it manifests itself as a nested Lie bracket in the equations of motion. For example, the equation

$$\dot{\Pi} = \Pi \times \Omega + \alpha\, \Pi \times (\Pi \times \Omega)$$

describes a damped rigid body; the first term generates the ordinary rigid body motion while the second term is a dissipative term of the double bracket type. Bloch et al. start their derivations on the Lagrangian side but finally switch to the Hamiltonian side, where they express the dissipation by modifying the Poisson bracket on the Lie group:

$$\dot{F} = \{F, H\} - \{\{F, H\}\}$$

where H is the Hamiltonian, $\{F, H\}$ is the ordinary Lie-Poisson bracket and $\{\{F, H\}\}$ is the (symmetric!) modification which causes the dissipation. We see that for this type of dissipation the flow of the system is still generated by a Hamiltonian, but the bracket which generates the dynamics is no longer Poisson (and hence the flow is not canonical, as expected.)

In this paper I will take a very different point of view. Suppose we have a Hamiltonian system driven by white noise, i.e. some parameter in the Hamiltonian fluctuates randomly. The flow generated by this system will obviously be inherently unpredictable, but by construction each realization of the flow (i.e. the flow generated by a particular history of the random fluctuations) will be canonical: after all, it is generated explicitly by a Hamiltonian. However, when we take an ensemble of particles described by the noisy Hamiltonian and look at their average behavior, we find that the ensemble can behave like a damped system.

An interesting aspect of this third approach to dissipation is that unlike the approaches outlined above, we don't need to introduce extra formalism in order to deal with damping; the flow is generated by an "ordinary" Hamiltonian with the usual symplectic or Poisson structure. The damping is obtained only at the end by averaging the equations of motion over the realizations of the noise. In general, however, we will see that it is impossible to find a closed-form expression for the mean evolution of the stochastic system. Rather, we can write down an (infinite-dimensional) field equation describing the time evolution of the probability density of the system. Whether this picture makes physical sense depends on the system under consideration; we can use the formalism to introduce dissipation into field theories which are derived from microscopic Hamiltonians.

In order to describe noisy processes we must venture into the world of stochastic analysis. In section 2 I briefly review some of the major concepts of stochastic calculus in flat space. Next we must talk about generalizing this to manifolds, which turns out to be somewhat problematic in the usual Itô formulation. The insight gained from this, however, almost directly leads us to dissipation when we apply this to mechanical systems.

I find the mechanics of random systems an extremely fascinating topic—it is unfortunate that there appears to be very little literature about the subject. The theory has actually been developed in depth by J.-M. Bismut in the late 1970s and is presented in his monograph [3]. For some reason the field appears to have been abandoned since then, and his work is hardly cited. The book is remarkably readable, however, despite that it is horribly typeset and written in French.

2 A crash course in stochastic calculus in $\mathbb{R}^n$

Probability spaces, random variables and mathematical expectation

Before embarking on the description of stochastic processes we must formalize the notion of probabilities[1]. Suppose we are describing a system which can take values $\omega \in \Omega$. ω is then called an *elementary event*. In order to describe the random structure of the system we introduce the notion of probability; we could ask, for example, "what is the probability of the system taking the value ω", or "what is the probability of the system taking either the value ω_1 or ω_2", etc. Clearly a probability is associated not only to elements of Ω, but also to sets of elements in Ω (where e.g. $\{\omega_1, \omega_2\}$ denotes "ω_1 *or* ω_2".) We call such subsets *events*.

In order to formalize this notion we introduce two objects, the *σ-algebra* $\mathcal{F}$ and the *probability measure* P associated to Ω. $\mathcal{F}$ is a set of subsets of Ω which are "measurable", i.e. it contains all the events to which we can assign a probability. P is a map from $\mathcal{F}$ to the interval which assigns a probability to each event. Clearly $\mathcal{F}$ and P must satisfy some elementary logic requirements; e.g. if we can assign a probability to ω_1 *or* ω_2 and to ω_3 *or* ω_4, we must also be able to assign a probability to ω_1 *or* ω_2 *or* ω_3 *or* ω_4, and this must be the sum of the two probabilities if none of the elementary events coincide (we say the two events are *independent*.) If all such logic requirements are satisfied we call the triplet $(\Omega, \mathcal{F}, P)$ a *probability space*.

Definition 2.1 *Let Ω be a set of elementary events. A* σ-**algebra** $\mathcal{F}$ *on Ω is a set $\{U_i \subset \Omega\}$ s.t.*

1. $\varnothing \in \mathcal{F}, \Omega \in \mathcal{F}$
2. $U \in \mathcal{F}$ *iff* $\bar{U} \in \mathcal{F}$
3. *For any finite or countably infinite sequence $\{U_k \in \mathcal{F}\}$ we have $\bigcup_k U_k \in \mathcal{F}$*

[1]There is a large amount of literature available on axiomatic probability and stochastic processes. An excellent reference is [4].

Definition 2.2 *The map* $\mu : \mathcal{F} \to \mathbb{R}^+$ *is a* **measure** *on* $(\Omega, \mathcal{F})$ *if*

1. $\mu(\varnothing) = 0$

2. *For any finite or countably infinite sequence* $\{U_k \in \mathcal{F} | U_k \cap U_l = \varnothing \ \forall k \neq l\}$ *we have* $\mu(\bigcup_k U_k) = \sum_k \mu(U_k)$

Definition 2.3 *The map* $P : \mathcal{F} \to [0,1]$ *is a* **probability measure** *on* $(\Omega, \mathcal{F})$ *if it is a measure and* $P(\Omega) = 1$. *Then the triplet* $(\Omega, \mathcal{F}, P)$ *is a* **probability space.**

Now suppose the system expresses itself in some way by determining the value of another variable, for instance a real number. This could correspond to a particular measurement of the system or to a coupling of the random system to some other quantity. We express this notion through the concept of a (real) random variable.

Definition 2.4 *The* **real random variable** $\xi : \Omega \to \mathbb{R}^n$ *on the probability space* $(\Omega, \mathcal{F}, P)$ *is* $\mathcal{F}$**-measurable** *if* $\xi^{-1}(B) \in \mathcal{F} \ \forall B \in \mathcal{B}$, *where* $\mathcal{B}$ *is the* σ*-algebra of Borel sets on* $\mathbb{R}^n$.

Note that we are assuming that we can determine the probability of getting any collection of outputs from the random variables; this corresponds to all the Borel sets. Measurability of a random variable implies that this is consistent with the intrinsic probability structure of the underlying probability space; i.e. it ensures that we cannot extract information from the random variable that cannot be obtained directly from the underlying random system. This is important as the random system directly determines the value of the random variable, so no new information can be gained in this process.

Finally, we introduce the notion of *mathematical expectation*, which simply gives the mean value of a random variable. Recall that the weighted mean of a set of random numbers $\{x_i\}$ is given by $\sum_i x_i P(x_i)$. Expressing this in terms of random variables and taking the continuous limit gives the following definition:

Definition 2.5 *The* **mathematical expectation** $\mathbb{E}\xi$ *of a random variable* ξ *on* $(\Omega, \mathcal{F}, P)$ *is given by the Lebesgue integral*

$$\mathbb{E}\xi = \int_\Omega \xi(\omega) P(d\omega)$$

Stochastic processes and Brownian motion

We now extend the notion of a random variable to a fluctuating random quantity; this is known as a *stochastic process*. It is defined as follows:

Definition 2.6 *A parametrized collection of real random variables* $\{X_t\}_{t \in \mathbb{R}^+}$ *defined on a probability space* $(\Omega, \mathcal{F}, P)$ *is called a* **stochastic process.**

The probability space on which the stochastic process lives is now bigger than the space of a random variable: each $\omega \in \Omega$ gives a particular *realization* (or sample path or trajectory) $X_\cdot(\omega) : \mathbb{R}^+ \to \mathbb{R}^n$ of the system. At each fixed time t, $X_t(\cdot)$ is a random variable.

Measurability is an important concept here. Suppose we perform some measurement of the stochastic process at time t, which we express as a random variable $\xi_T : \Omega \to \mathbb{R}^n$. Obviously a measurement at time T (this could be, for instance, an estimate of some parameter of the system) can only depend on the values of the process at times $t \leq T$, otherwise causality is violated. In order to capture this we introduce a family of σ-algebras $\mathcal{F}_t$ associated to the stochastic process X; each $\mathcal{F}_t$ is constructed in such a way that one cannot distinguish between paths ω of X_t which are the same up to time t, but branch off later (i.e. all such ω are always grouped together in $U \in \mathcal{F}_t$.) For causality to hold, clearly ξ_T must be $\mathcal{F}_T$-measurable. If we repeat the measurement of X_t at different times and group all these measurements into a new process ξ_t, then ξ_t is said to be *$\mathcal{F}_t$-adapted* if ξ_T is $\mathcal{F}_T$-measurable for all T.

Definition 2.7 *Let $\mathcal{F}_t$ be the σ-algebra generated by $X_s(\cdot), s \leq t$. A stochastic process Y_t is* $\mathcal{F}_t$-**adapted** *if Y_T is $\mathcal{F}_T$-measurable $\forall T$.*

Consider the probability distribution

$$p_t(x,y) = (2\pi t)^{-1/2} \exp\left[-\frac{|x-y|^2}{2t}\right] \qquad P(t_1,x_1;t_2,x_2;\cdots) = p_{t_1}(0,x_1)p_{t_2-t_1}(x_1,x_2)\cdots$$

where $x, x_n, y \in \mathbb{R}$. $P(t_1,x_1;t_2,x_2;\cdots)$ is the probability of finding the particle at x_1 at time t_1, x_2 at time t_2, etc., i.e. this is the probability of a particular *discrete* trajectory. Taking the continuous limit and constructing a corresponding probability space, we obtain a stochastic process called *Brownian motion* W_t. As stated here this process isn't very well-defined, but with some additional machinery the existance of this process and its underlying probability space can be proved[4]. It has many nice properties which will be implicitly used in the following, though going into details is beyond the scope of this paper. One suprising property is that its sample paths are continuous (but not differentiable.) Note also that $\mathbb{E}W_t = 0$.

Stochastic differential equations and the Itô integral

In this paper (and in this general subject area) one usually generates a stochastic process by driving a dynamical system with a white noise term, i.e., we want to give meaning to the equation

$$\frac{dX_t}{dt} = b(t, X_t) + \sigma(t, X_t)\xi_t$$

where ξ_t is white noise (zero mean, unit variance). Stochastic processes are generally not differentiable, however, so as it is written this equation doesn't make sense. In order to define what we mean by such a noise-driven dynamical system we write it in integral form:

$$X_t = X_0 + \int_0^t b(t, X_t)dt + \int_0^t \sigma(t, X_t)\xi_t dt$$

This seems more reasonable, as we can give meaning (through Lebesgue or Stieltjes-like constructions) to integrals of all sorts of not-very-well-behaved functions. In this case we want to define the integral over a stochastic process, and we expect the result to be another stochastic process.

We immediately run into a snag, however. It turns out that white noise is *not* a stochastic process; one can prove that there is no way to realize white noise as a continuous stochastic process[4] (one can extend one's notion of a stochastic process to include tempered distributions, but this is not desirable.) Fortunately, there is a way around this problem. Let's replace $\xi_t dt$ by the increment dW_t of an appropiate stochastic process W_t, such that

$$\int_0^t \xi_t dt = W_t$$

where ξ_t is treated as a tempered distribution. It turns out W_t is exactly the brownian motion we defined earlier. Thus, we will write our dynamical system completely in terms of stochastic processes as

$$X_t = X_0 + \int_0^t b(t, X_t)dt + \int_0^t \sigma(t, X_t)dW_t$$

or as a *Stochastic Differential Equation* (SDE)

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dW_t$$

which is just a shorthand notation for the integral equation.

We must now give meaning to the stochastic integral. As usual we do this by replacing the function we're integrating over by a sequence of approximations in terms of elementary functions and taking the limit:

Definition 2.8 *The* **Itô integral** *of a stochastic process is defined as*

$$\int_0^t U_t\, dW_t = \lim_{\Delta t \to 0} \sum_n U_{t_n} \left(W_{t_{n+1}} - W_{t_n}\right)$$

This expression is reminiscent of the Riemann-Stieltjes integral. Note however that we're approximating U_t by its value on the left side of the time interval n. In a Riemann-Stieltjes integral it does not matter at which point $t_* \in [t_n, t_{n+1})$ in the bin the integrand is evaluated, but it turns out that this choice gives very different results for stochastic integrals. This reflects the fact that stochastic processes fluctuate so rapidly that the Riemann-Stieltjes construction doesn't hold.

In the next section we will discuss a different choice of t_*, leading to the *Stratonovich integral* of a stochastic process. We will see that this integral makes much more sense from a geometric point of view. In stochastic analysis, however, one mainly uses the Itô integral due to its nice statistical properties. Note that the increment dX_t is generated by the noise increment in each time step, but it only depends on the value of X_t at the beginning of the time step. This ensures that in a sense the dynamics of the system is uncorrelated with the noise (though the flow is obviously still driven by the noise), and we get the following theorem[4]:

Theorem 2.1 *Let $\mathcal{F}_t$ be generated by $W_{s \leq t}$. Then $\int_0^t U_t\, dW_t$ is $\mathcal{F}_t$-measurable; furthermore, $\mathbb{E} \int_0^t U_t\, dW_t = 0$.*

Finally, the following theorem describes how (one-dimensional) Itô SDE behave under transformations:

Theorem 2.2 (1-dimensional Itô formula) *Let*

$$dX_t = u\, dt + v\, dW_t \quad \text{and} \quad Y_t = g(t, X_t)$$

Then we have

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

where $(dX_t)^2$ is calculated using the **Itô calculus** *$dt^2 = dt\, dW_t = dW_t\, dt = 0$, $dW_t^2 = dt$.*

We see that in order to obtain nice statistical properties of the stochastic integral we have sacrificed the rules of calculus as they apply to ordinary differential equations. Thus, one can't naively interpret the Itô SDE as some noisy limit of an ordinary SDE. This is possible, however, in the Stratonovich picture.

3 Stochastic calculus on manifolds: Stratonovich vs. Itô

From the Itô formula we see immediately that the "Itô differential" dX is not natural under diffeomorphisms: $\phi^*dX \neq d\phi^*X$. Thus such a notion does not make intrinsic sense on a manifold. We can save the day by introducing a new calculus, as follows.

Definition 3.1 *Let $\cdot$ denote the Itô calculus multiplication. We define the* **Stratonovich calculus** *as* $Y \circ dX = Y \cdot dX + \frac{1}{2} dX \cdot dY$, $dX \circ dY = 0$.

In order to make this definition rigorous we have to define exactly what we mean by the Itô algebra; for a formal exposition, see [5]. We proceed using the Itô rule and some intuition to replace the technicalities:

Theorem 3.1 *Let* $Y_t = g(t, X_t)$*; then* $dY_t = \frac{\partial g}{\partial t} dt + \frac{\partial g}{\partial X_t} \circ dX_t$

Proof

$$\frac{\partial g}{\partial X_t} \circ dX_t = \frac{\partial g}{\partial X_t} \cdot dX_t + \frac{1}{2} d\left(\frac{\partial g}{\partial X_t}\right) \cdot dX_t$$

Expanding the second term using Itô's rule gives

$$\frac{1}{2} d\left(\frac{\partial g}{\partial X_t}\right) \cdot dX_t = \frac{1}{2}\left[\frac{\partial^2 g}{\partial X_t \partial t} dt + \frac{\partial^2 g}{\partial X_t^2} \cdot dX_t + \frac{1}{2}\frac{\partial^3 g}{\partial X_t^3} \cdot dX_t^2\right] \cdot dX_t$$

Now use the Itô calculus: $dt \cdot dX_t = dX_t^3 = 0$. We get

$$\frac{\partial g}{\partial t} dt + \frac{\partial g}{\partial X_t} \circ dX_t = \frac{\partial g}{\partial t} dt + \frac{\partial g}{\partial X_t} \cdot dX_t + \frac{1}{2}\frac{\partial^2 g}{\partial X_t^2} \cdot dX_t^2 = dY_t$$

where the last equality is by Itô's rule. □

We see that a *Stratonovich SDE* $dX_t = u\,dt + v \circ dW_t$ transforms naturally under diffeomorphisms, and so we can make intrinsic sense of it on a manifold. The following theorem relates the Stratonovich integral to the Itô integral:

Theorem 3.2

$$\int_0^t U_t \circ dW_t = \lim_{\Delta t \to 0} \sum_n \frac{U_{t_n} + U_{t_{n+1}}}{2} (W_{t_{n+1}} - W_{t_n})$$

Proof

$$\sum_n \frac{U_{t_n} + U_{t_{n+1}}}{2} (W_{t_{n+1}} - W_{t_n}) = \sum_n U_{t_n} (W_{t_{n+1}} - W_{t_n}) + \frac{1}{2} \sum_n (U_{t_{n+1}} - U_{t_n})(W_{t_{n+1}} - W_{t_n})$$

Thus we get

$$\lim_{\Delta t \to 0} \sum_n \frac{U_{t_n} + U_{t_{n+1}}}{2} (W_{t_{n+1}} - W_{t_n}) = \int_0^t U_t \cdot dW_t + \frac{1}{2} \int_0^t dU_t \cdot dW_t = \int_0^t U_t \circ dW_t \qquad \square$$

Once again, this can be made more precise by formalizing what we mean by the Itô and Stratonovich calculus, but the derivation is intuitively clear. For more details, see [5].

The definition of the Stratonovich integral is more symmetric than the Itô integral, averaging the value of the integrand over each bin. From a statistics point of view this leads to very unfortunate results; the integral is no longer $\mathcal{F}_t$-adaped and its mean drifts around instead of being nicely fixed at zero. From a geometric point of view this is very nice however. Particularly, Bismut[3] shows (following a classic work by Wong and Zakai) that if we take *differentiable* approximations to an SDE, i.e. we take a set of smooth vector fields Ξ^n that converge to $\Xi = u + v\xi$ (where ξ is some realization of white noise), then taking the corresponding limit of the ordinary Riemann-Stieltjes integrals of these smooth dynamical systems gives exactly the Stratonovich SDE $dX_t = u\,dt + v \circ dW_t$. Thus in a sense the Stratonovich picture is much more physical, being a singular limit of a "real" physical process. The result is that the system dynamics gets funamentally entangled with the noise and we must do some work to find the statistical properties of the system (i.e., convert to Itô form before doing statistics.)

This result suggests a natural way to extend the Stratonovich picture to manifolds. We already know how to integrate smooth dynamical systems on manifolds; now take the appropiate limit, à la Wong and Zakai, and we obtain a Stratonovich system on a manifold. I have seen various other constructions in the literature, some using Whitney's embedding theorem, others patching together Stratonovich integrals on coordinate charts. The various methods appear to agree and for the purposes of this paper I will assume that the Stratonovich picture is naturally extended to manifolds, and that all such technicalities can be worked out.

It remains to find how to convert between Itô and Stratonovich forms of SDE.

Theorem 3.3 (Itô-Stratonovich equivalence in one dimension)

$$dX_t = u\,dt + v\,dW_t \quad \Longleftrightarrow \quad dX_t = \left(u - \frac{v}{2}\frac{\partial v}{\partial X_t}\right) dt + v \circ dW_t$$

$$dX_t = u\,dt + v \circ dW_t \quad \Longleftrightarrow \quad dX_t = \left(u + \frac{v}{2}\frac{\partial v}{\partial X_t}\right) dt + v\,dW_t$$

Proof Just use the conversion rule: for example,

$$v \circ dW_t = v \cdot dW_t + \frac{1}{2} dv \cdot dW_t$$

Now use the Itô rule

$$dv \cdot dW_t = \left[\frac{\partial v}{\partial t} dt + \frac{\partial v}{\partial X_t} dX_t + \frac{1}{2}\frac{\partial^2 v}{\partial X_t^2}(dX_t)^2\right] \cdot dW_t = \frac{\partial v}{\partial X_t}(v \circ dW_t) \cdot dW_t$$

Now use the conversion rule again,

$$(v \circ dW_t) \cdot dW_t = \left(v \cdot dW_t + \frac{1}{2} dv \cdot dW_t\right) \cdot dW_t = v\,dt$$

The Itô-Stratonovich equivalence follows immediately. Reversing the procedure gives the Stratonovich-Itô equivalence. □

Generally we will work in a dimension higher than one; the derivation of the higher-dimensional analogs of these theorems is more tedious but straightforward. I will just state one result:

Theorem 3.4 (Itô-Stratonovich equivalence in many dimensions) *Let X_t be an n-dimensional process with components X_t^i. Then*

$$dX_t^i = u^i\,dt + v^i\,dW_t \quad\Longleftrightarrow\quad dX_t^i = \left(u^i - \frac{1}{2}\sum_{j=1}^{n} v^j \frac{\partial v^i}{\partial X_t^j}\right) dt + v^i \circ dW_t$$

$$dX_t^i = u^i\,dt + v^i \circ dW_t \quad\Longleftrightarrow\quad dX_t^i = \left(u^i + \frac{1}{2}\sum_{j=1}^{n} v^j \frac{\partial v^i}{\partial X_t^j}\right) dt + v^i\,dW_t$$

The correction term in these expressions is given explicitly in coordinates and does not seem to make intrinsic sense without additional structure. However, one can make sense of this term on a Riemannian manifold[3, 6, 7]; recall that transforming the connection coefficients Γ^i_{jk} gives second derivatives, which is reminiscent of Itô's rule. In fact Itô's rule and the Itô correction term can be expressed in a natural way in terms of covariant derivatives. I will say no more about this here. However, it should be pointed out that the double bracket damping terms of Bloch et al. are defined explicitly using a metric; their construction of the double bracket also doesn't appear to make sense without this additional structure on the manifold. In our case it is the Itô correction which will play the role of the damping term.

4 Stochastic mechanics: a simple linear example

We will take for our configuration manifold $Q = \mathbb{R}$, so the phase space is $T^*Q \simeq \mathbb{R}^2 \ni (q,p)$. The symplectic form in T^*Q is $\Omega = dq \wedge dp$. Consider the equations of motion

$$\begin{pmatrix} \dot q \\ \dot p \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 0 & -\eta \end{pmatrix} \begin{pmatrix} q \\ p \end{pmatrix} = \mathbf{X}(q,p)$$

This describes a linearly damped free particle in one dimension; for example, think of a bead moving on a wire and experiencing friction. Obviously this picture doesn't lend itself very well to a probabilistic interpretation, but it is still a nice demonstration of the origin of the damping term.

The equations of motion are not Hamiltonian; after all,

$$d i_{\mathbf{X}}\Omega = d(p\,dp + \eta p\,dq) = \eta\,dp \wedge dq \neq 0$$

Our goal will be to "reverse-engineer" a noisy Hamiltonian which behaves like this dissipative system on average.

To proceed, we first must know when such a linear system is Hamiltonian. Writing

$$\begin{pmatrix} \dot q \\ \dot p \end{pmatrix} = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} q \\ p \end{pmatrix} = \mathbf{X}'(q,p)$$

and requiring

$$0 = d i_{\mathbf{X}'}\Omega = d((Aq + Bp)dp - (Cq + Dp)dq) = (A + D)dq \wedge dp$$

we see that $\mathbf{X}'$ is Hamiltonian iff $A = -D$ (we know it is globally Hamiltonian as we are working in a vector space.)

Our strategy will be as follows. We assume our given equations of motion are actually the ensemble average of a noisy equation. We postulate

$$\begin{pmatrix} dq_t \\ dp_t \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 0 & -\eta \end{pmatrix} \begin{pmatrix} q_t \\ p_t \end{pmatrix} dt + \begin{pmatrix} A & B \\ C & -A \end{pmatrix} \begin{pmatrix} q_t \\ p_t \end{pmatrix} dW_t$$

in the Itô sense. Taking the mathematical expectation of both sides gives, as everything is linear,

$$\begin{pmatrix} d(\mathbb{E}q_t) \\ d(\mathbb{E}p_t) \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 0 & -\eta \end{pmatrix} \begin{pmatrix} \mathbb{E}q_t \\ \mathbb{E}p_t \end{pmatrix} dt$$

which is exactly what we want. Note that we must necessarily write an Itô equation, as the mathematical expectation of a Stratonovich integral is not zero in general, but for the Itô integral this is guaranteed by theorem 2.1. Thus we are now on the "statistical side" of the picture. In order to find a corresponding Hamiltonian, we must switch to the "geometric side". To this end we convert to the Stratonovich picture using theorem 3.4:

$$\begin{pmatrix} dq_t \\ dp_t \end{pmatrix} = \begin{pmatrix} -\frac{1}{2}(A^2+BC) & 1 \\ 0 & -\eta - \frac{1}{2}(A^2+BC) \end{pmatrix} \begin{pmatrix} q_t \\ p_t \end{pmatrix} dt + \begin{pmatrix} A & B \\ C & -A \end{pmatrix} \begin{pmatrix} q_t \\ p_t \end{pmatrix} \circ dW_t$$

Now that we are in the Stratonovich picture, the concept of a noisy Hamiltonian makes sense. In order for this system to be Hamiltonian we require

$$\frac{1}{2}(A^2+BC) = -\eta - \frac{1}{2}(A^2+BC)$$

One possibility is $A = 0$, $B = -C = \sqrt{\eta}$. We obtain the Stratonovich system

$$\begin{pmatrix} dq_t \\ dp_t \end{pmatrix} = \begin{pmatrix} \eta/2 & 1 \\ 0 & -\eta/2 \end{pmatrix} \begin{pmatrix} q_t \\ p_t \end{pmatrix} dt + \begin{pmatrix} 0 & \sqrt{\eta} \\ -\sqrt{\eta} & 0 \end{pmatrix} \begin{pmatrix} q_t \\ p_t \end{pmatrix} \circ dW_t$$

which is easily verified (by substituting into Hamilton's equations) to come from the noisy Hamiltonian

$$H_t(q,p) = \frac{1}{2}p^2 + \frac{1}{2}\eta qp + \frac{1}{2}\sqrt{\eta}(p^2+q^2)\xi_t$$

where ξ_t is (physical) white noise. Such a solution was also found in [8].

What is the interpretation of this? As the damped system can be derived from a noisy Hamiltonian, we get that each sample path is canonical but the ensemble of particles is damped. In a sense this is sort of obvious. We can see from the Hamiltonian that we add momentum noise into the system; intuitively, at short times the system will "remember" its initial momentum while at long times the momentum will be dominated by the noise, which averages out to zero. Averaging this gives a net damping effect. The Stratonovich-Itô correspondence gives us a way of extracting this information from the noise; the Itô correction term then plays the role of a *deterministic* damping term, while the Itô noise averages out to zero by virtue of theorem 2.1.

Unfortunately, this method doesn't generalize to nonlinear systems. To see the problem, let us try to apply a similar procedure to the damped rigid body

$$\dot{\Pi} = \Pi \times \Omega + \alpha \Pi \times (\Pi \times \Omega)$$

We start by adding a Hamiltonian Itô noise term

$$d\Pi = (\Pi \times \Omega + \alpha \Pi \times (\Pi \times \Omega))\, dt + \Pi \times \nabla F(\Pi)\, dW_t$$

In the previous example we relied on the fact that we regain our original equation of motion if we ensemble-average the stochastic equation. Unfortunately, this procedure fails in this case, because $\mathbb{E}(\Pi_i\Pi_j) \neq \mathbb{E}\Pi_i\,\mathbb{E}\Pi_j$, etc. In order to get the equation for $d(\mathbb{E}\Pi)$ we must obtain an equation for $\mathbb{E}(\Pi\times\Omega)$ in terms of $\mathbb{E}\Pi$. But the equation for $d(\mathbb{E}(\Pi_i\Pi_j))$ (by Itô's rule) depends on *higher-order* moments of Π (i.e. $\mathbb{E}(\Pi_i\Pi_j\Pi_k)$), not on lower-order moments! Thus there is no closed-form equation for $\mathbb{E}\Pi$, but instead we get an infinite hierarchy of coupled equations for the moments $\mathbb{E}\Pi_i$, $\mathbb{E}(\Pi_i\Pi_j)$, $\mathbb{E}(\Pi_i\Pi_j\Pi_k)$, ... We see that the "reverse-engineering" procedure described above only works for linear systems.

5 Field equations from microscopic stochastic Hamiltonians

We get much nicer results if we don't attempt to calculate the mean time evolution of a stochastic system, but we look for the time evolution of the probability density on phase space.

Let's first recall how functions of phase space trajectories evolve in time in a deterministic system. Let $\dot{x} = X_H(x)$ and F_t is the flow of X_H. We wish to calculate the time evolution of $A(x(t))$. We proceed as follows:

$$\frac{d}{dt}A(x(t)) = \frac{d}{dt}F_t^* A(x(0)) = F_t^* \mathscr{L}_{X_H} A(x(0)) = \mathscr{L}_{X_H} A(x(t))$$

where we have used the Lie derivative formula and $\phi^* \mathscr{L}_X f = \mathscr{L}_{\phi^* X}\phi^* f$. But by definition of a Hamiltonian vector field,

$$\frac{d}{dt}A(x(t)) = \mathscr{L}_{X_H} A(x(t)) = \Omega(X_A, X_H) = \{A, H\}$$

Let us now find a corresponding formula for a noisy Hamiltonian. We start with a Stratonovich system

$$dx_t = X_H(x_t)\,dt + X_F(x_t)\circ dW_t$$

which comes from the noisy Hamiltonian $H + F\xi_t$. For a function A we get (using ordinary manifold calculus as we're in the Stratonovich picture)

$$dA_t = \mathscr{L}_{X_H} A_t\,dt + \mathscr{L}_{X_F} A_t \circ dW_t$$

Now we switch to the Itô side. We need to calculate the Itô correction. Suprisingly, in this case we don't need a connection, as we can express the Itô correction completely in terms of Lie derivatives[3, 9]:

$$\frac{1}{2}\left[\frac{\partial}{\partial A_t}(\mathscr{L}_{X_F} A_t)\right](\mathscr{L}_{X_F} A_t) = \frac{1}{2}\mathscr{L}_{X_F}\mathscr{L}_{X_F} A_t = \frac{1}{2}\mathscr{L}^2_{X_F} A_t$$

Thus we get the Itô equation

$$dA_t = \left(\mathscr{L}_{X_H} + \frac{1}{2}\mathscr{L}^2_{X_F}\right) A_t\,dt + \mathscr{L}_{X_F} A_t\,dW_t$$

Replacing the Lie derivatives by Poisson brackets and taking the mathematical expectation, we get

$$\langle \dot{A} \rangle = \left\langle \{A, H\} + \frac{1}{2}\{\{A, F\}, F\} \right\rangle$$

where $\langle A \rangle = \mathbb{E}A$.

We saw earlier that we can't, in general, find a closed form of this expression for the mathematical expectation $\langle A\rangle$ for a general $A(x)$. Instead, we follow the method of Gardiner[10] to obtain an equation of motion for the probability density $\rho(x)$. By definition, for any A

$$\langle A(x(t))\rangle = \int_M A(x)\rho(x,t)dx$$

where M is the phase space and dx is its volume element. Taking the time derivative and using the expression we derived above, we get

$$\langle \dot{A}\rangle = \int_M A(X)\dot{\rho}(x,t)dx = \int_M \left(\mathscr{L}_{X_H}A(x) + \frac{1}{2}\mathscr{L}^2_{X_F}A(x)\right)\rho(x,t)dx$$

In order to simplify this equation we need the following lemma:

Lemma Let M be the phase space and μ its volume form. Assume also that M is boundaryless, or that f, g vanish on ∂M. Let X_H be a Hamiltonian vector field. Then

$$\int_M f(\mathscr{L}_{X_H}g)\,\mu = -\int_M g(\mathscr{L}_{X_H}f)\,\mu$$

Proof

$$\int_M f(\mathscr{L}_{X_H}g)\,\mu = \int_M \mathscr{L}_{X_H}(fg\mu) - \int_M g(\mathscr{L}_{X_H}f)\,\mu - \int_M fg(\mathscr{L}_{X_H}\mu)$$

By Cartan's magic formula, we have

$$\int_M \mathscr{L}_{X_H}(fg\mu) = \int_M (di_{X_H}(fg\mu) + i_{X_H}d(fg\mu))$$

But the last term vanishes as $d(fg\mu) = 0$ ($fg\mu$ is a form of the same dimension as the phase space.) Using Stokes' theorem we get

$$\int_M \mathscr{L}_{X_H}(fg\mu) = \int_M di_{X_H}(fg\mu) = \int_{\partial M} fg\,(i_{X_H}\mu) = 0$$

as f, g vanish on the boundary. Next, we tackle the last term. Note that $\mathscr{L}_{X_H}\mu = \mathrm{div}_\mu X_H = 0$, as the flow of a Hamiltonian vector field is volume-preserving. Thus, the last term vanishes as well, and we obtain the desired result. □

Using this Lemma we rewrite our previous equation as

$$\langle \dot{A}\rangle = \int_M A(x)\dot{\rho}(x,t)dx = \int_M A(x)\left(-\mathscr{L}_{X_H}\rho(x,t) + \frac{1}{2}\mathscr{L}^2_{X_F}\rho(x,t)\right)dx$$

As this equation holds for any $A(x)$, we obtain a closed-form solution for the probability density of the stochastic system:

$$\frac{\partial\rho}{\partial t} = -\{\rho, H\} + \frac{1}{2}\{\{\rho, F\}, F\}$$

In statistics this is known as a Fokker-Planck equation. We see that in a mechanical system, the Fokker-Planck equation is expressed naturally in terms of Poisson brackets. A similar result is found in [9] for non-white noise processes.

We can use this formalism to derive field theories from ensembles of particles for which we specify some microscopic dynamics. An interesting example is the Vlasov-Poisson equation from plasma physics. This field equation is

$$\frac{\partial f}{\partial t} + \mathbf{v}\cdot\frac{\partial f}{\partial \mathbf{x}} - \frac{e}{m}\frac{\partial \phi_f}{\partial \mathbf{x}}\cdot\frac{\partial f}{\partial \mathbf{v}} = 0 \qquad \nabla^2\phi_f = e\left(\int f\,d^3v - 1\right)$$

where f is the plasma particle number density. Bloch et al.[2] show that this equation can be written

$$\frac{\partial f}{\partial t} = -\{f, H_f\} \qquad H_f = \frac{1}{2}m^2\|\mathbf{v}\|^2 + e\phi_f(\mathbf{x})$$

with the canonical Poisson bracket

$$\{g, h\} = \frac{1}{m}\left(\frac{\partial g}{\partial \mathbf{x}} \cdot \frac{\partial h}{\partial \mathbf{v}} - \frac{\partial h}{\partial \mathbf{x}} \cdot \frac{\partial g}{\partial \mathbf{v}}\right)$$

But this is exactly our Fokker-Planck equation without a noise term. This suggests that the Vlasov-Poisson equation actually follows from microscopic dynamics of particles flowing according to the microscopic Hamiltonian H_f. That the Hamiltonian depends on the whole distribution f implies that the particles interact; in order to solve the motion of a single particle we must know the position of all other particles. Thus the microscopic equations of motion can only be solved self-consistently, where the motions of all the particles are used to update their distribution. Obviously one doesn't need to do this as one has the Vlasov-Poisson equation. However, it is conceptually nice to think of it in this way in order to interpret the equation as a Fokker-Planck equation.

Having made this observation, it is now trivial to see what the effect of a noise drive is on the plasma. We can simply write

$$\frac{\partial f}{\partial t} = -\{f, H_f\} + \frac{1}{2}\{\{f, F\}, F\}$$

where F is a suitably chosen noise Hamiltonian.

It is interesting to compare this to the double bracket dissipation found in [2]:

$$\frac{\partial f}{\partial t} = -\{f, H_f\} + \alpha\,\{f, \{f, H_f\}\}$$

The result looks entertainingly similar. It is unclear to me, however, that there should be any similarity between the two expressions other than the aesthetic grouping of the curly brackets.

6 Conclusion

After a brief overview of classical stochastic analysis, I showed (albeit not very rigorously) how these notions can be made sense of on manifolds. The basic lesson to be learned here is that the two "pictures" of Stochastic analysis, the Itô and the Stratonovich picture, are both neccessary in order to get anywhere. Geometry can be done on the Stratonovich side, while statistics takes place on the Itô side. Connecting the two requires (in general) a Riemannian structure on the manifold, but I have glossed over this issue. Finally, I showed how to introduce the notion of a noisy Hamiltonian in mechanics, and the role played by the Itô-Stratonovich equivalence in this case. We also derived a closed-form Fokker-Planck equation in terms of Poisson brackets.

This project originally started as an attempt to make sense of quantum measurement in a geometric mechanics context. When a quantum system is subjected to (continuous) measurement, quantum mechanics prescribes how the noise enters the system. The dynamics of a quantum system under measurement can be written as a (nonlinear!) Itô SDE on Hilbert space[11], much like the SDE I considered in this paper. Unfortunately, however, a quick calculation shows that the quantum measurement SDE is not Hamiltonian.

Still, I think the ideas presented here can be useful in such a context. A particular lesson is that one should do geometry on the Stratonovich side. If one wants to apply methods of geometric control to feedback control of quantum systems this could be an important issue—as far as I know, most of the stochastic control theory is formulated on the Itô side. The observation that one can do geometric control on the Stratonovich side appears to be the essence of a recent article on feedback linearization of stochastic systems[12].

References

[1] J.E. Marsden and T.S. Ratiu, "Introduction to Mechanics and Symmetry", 2nd edition (Springer 1999)

[2] A. Bloch, P.S. Krishnaprasad, J.E. Marsden and T.S. Ratiu, *Comm. Math. Phys.* **175**, 1 (1996)

[3] J.-M. Bismut, "Mécanique Aléatoire", Lecture Notes in Mathematics 866 (Springer-Verlag 1981)

[4] B. Øksendal, "Stochastic Differential Equations", 5th edition (Springer 2000)

[5] N. Ikeda and S. Watanabe, "Stochastic Differential Equations and Diffusion Processes", 2nd edition (North-Holland 1989)

[6] J.R. Norris, "A complete differential formalism for stochastic calculus in manifolds", in Séminaire de Probabilités XXVI, eds. J. Azéma, P.A. Meyer and M. Yor, Lecture Notes in Mathematics 1526 (Springer-Verlag 1992)

[7] D.C. Brody and L.P. Hughston, *Proc. R. Soc. Lond. A* **458**, 1117 (2002)

[8] J. Gough, *J. Math. Phys.* **40**, 2805 (1999)

[9] J.M.R. Parrondo, M. Mañas and F.J. de la Rubia, *J. Phys. A: Math. Gen.* **23**, 2363 (1990)

[10] C.W. Gardiner, "Handbook of Stochastic Methods", 2nd edition (Springer 2002)

[11] H.M. Wiseman, "Quantum Trajectories and Feedback", Ph.D. thesis, Univ. of Queensland (1994)

[12] L. Sládecek, `arXiv:math.OC/0211382` (2002)