A path integral formalism for the closure of autonomous statistical systems.

Richard Kleeman
Courant Institute of Mathematical Sciences, New York, USA.
E-mail: kleeman@cims.nyu.edu

Abstract. Recently a path integral formalism has been proposed by the author which gives the time evolution of moments of slow variables in a Hamiltonian statistical system. That formalism is extended here to deal with a more general autonomous dynamical system. An immediate practical application to forced dissipative turbulent systems is thus now possible.

Keywords: closure, path integral, statistical system

PACS numbers: 05

AMS classification scheme numbers: 82
1. Introduction

The statistical closure of the dynamical systems underlying turbulent phenomenon is a long standing and practically important problem in mathematical physics. Recently the author has proposed a new formalism to deal with Hamiltonian systems using a Wiener path integral [1]. This formalism builds on earlier work by Turkington [2]. A non-equilibrium density for such dynamical systems satisfies the corresponding Liouville equation. The non-equilibrium solutions of this equation are however generally very complex and only approximations might be hoped for using the slow variables and invariants of the system (see [3]). This situation is of course in stark contrast to that applying in equilibrium systems where close to exact Gibbs invariant densities may often be obtained. Despite the approximate nature, one can insist that the temporal sequence of approximations satisfy the Liouville equation “as closely as possible”. Such a “fit metric” can be obtained using information theoretic functionals on the approximating densities. This basic variational principle underlies the approaches of both of the references cited above and is explained for the first reference in more detail below.

The approximation approach been validated numerically using an inviscid truncated Burgers system [4]. Here the approximating densities are of a maximum entropy type with the first moments of the slow, low wavenumber modes constrained. The performance of the closure in simulating the time evolution of these first moments was assessed and found to be generally quite good. This is an encouraging result because the Burgers system does not have a very sharp separation between slow and fast modes (see [5]).

The long term aim of this work is to apply the closure methodology to realistic turbulent system of various kinds. Consequently it would be desirable if the Hamiltonian restriction could be relaxed in order to deal with systems with dissipation and external forcing. This will be shown to be theoretically possible in the present contribution.

Here we briefly review the earlier first mentioned work mainly from a motivational and scene setting perspective before providing a complete derivation of our formalism for a general autonomous dynamical system. The reader is referred to both of the earlier papers mentioned above for more details.

2. Onsager-Machlup Theory

It has been well known since early in the history of statistical physics that the probability of a fluctuation from an equilibrium state at a fixed time is given by the Boltzmann principle

\[ p(\lambda) = C \exp \left( -\sigma(\lambda) \right) \]

where \( \sigma \) is the entropy function for the system. Onsager and Machlup (OM) [6] generalized this principle in the early 1950s and defined a positive likelihood weight \( W \) on temporal paths \( \lambda(t) \) of near equilibrium states

\[ W[\lambda(t)] = \exp \left( -\int_0^t \mathcal{L}[\lambda(t)] \, dt \right) \]  \hspace{1cm} (1)

The “Lagrangian” \( \mathcal{L} \) was assumed to be of the form:

\[ \mathcal{L}[\lambda(t)] = \frac{1}{2} \left( \dot{\lambda} - U\lambda \right)^t g \left( \lambda - U\lambda \right) \]

where \( U \) and \( g \) are constant matrices with the latter importantly assumed to be positive definite.
It is immediately clear from the latter assumption that the path of greatest weight is simply given by
\[ \dot{\lambda} = U \lambda \] (2)
which is, in general, a relaxation to equilibrium for appropriate choices for \( U \). Furthermore the probability of a particular \( \lambda \) at time \( T \) is obtained by \textbf{summing} over the likelihood weights of all paths ending at \( \lambda(T) \):
\[ p(\lambda(T)) = C \int D\lambda W[\lambda] \] (3)
which is a functional or \textbf{path integral} over all possible paths which end at \( \lambda(T) \). As is well known \[7\] this situation corresponds with an Ornstein Uhlenbeck stochastic process and the most likely \( \lambda \) also follows equation (2). Such a trajectory is called thermodynamical since it is most probable.

Many attempts have been made to formulate far from equilibrium versions of OM theory. The approach followed here follows three principles:

(i) The Lagrangian \( \mathcal{L} \) should follow from a fundamental information theoretic argument relating to paths since it effectively generalizes the entropy function underlying the original Boltzmann principle.

(ii) The precise form of the Lagrangian should be derivable from first principles and reflect the statistical properties of the slow or coarse grained variables for the statistical system.

(iii) In appropriate limits thermodynamical relations of the type proposed recently by Öttinger \[8\] should be recovered since such theories work well in practical applications.

3. Trial density manifold

Zubarev \[3\] proposed the concept of a trial density as a way of approximating densities of non-equilibrium states. One specifies a set of moments of slow variables as thermodynamical variables and then uses a maximum entropy principle to associate these with a “trial” density. In general such trial densities \( \hat{\varrho} \) are not exact densities \( \varrho \) for the given non-equilibrium state with the given moments. This situation contrasts with the equilibrium Gibbs density. The philosophical approach taken here therefore differs from that adopted in equilibrium statistical physics in that it tacitly assumed that the \textbf{exact} density for a system will \textbf{never generally} be available and thus one must deal always with an approximation manifold of densities. On such a manifold the “veracity” of each density will be assigned a non-negative weight which we shall term a \textbf{consistency distribution}. It is analogous to a quantum wavefunction as we shall see below.

The set of all trial densities with a prescribed set of moments can be given a Riemannian manifold structure by using the associated Fisher information matrix as a metric tensor. This manifold structure is the essence of the subject of information geometry \[9\] and shall become important to our discussion later.

4. Liouvillian discrepancy

The equation for the density evolution within a general autonomous dynamical equation is simply the Fokker-Planck equation without a diffusion term i.e.
\[ \varrho_t + L \varrho = 0 \] (4)
Path integrals for autonomous dynamics

\[ L \equiv \frac{\partial}{\partial x_i} A_i \]

with summation convention and where the original dynamical system is

\[ \frac{dx_i}{dt} = A_i(x) \]

As usual the formal adjoint of the operator \( L \) is given by

\[ L^* = -A_i \frac{\partial}{\partial x_i} \]

From this we deduce that

\[ L + L^* = L^d \equiv \frac{\partial A_i}{\partial x_i} \quad (5) \]

Consider now a temporal trajectory or path through the trial density manifold. In general this will not satisfy equation (4). A measure of the discrepancy of a given path from such an evolution can be defined as follows: Suppose the trial path \( \hat{\varrho}(t) \) is evolved forward in time an additional \( \Delta t \) according to the above equation resulting in

\[ \varrho'(t + \Delta t) = \exp(-\Delta t L) \hat{\varrho}(t) = \exp(\Delta t T) \hat{\varrho}(t) \]

\[ T \equiv \frac{\partial}{\partial t} \]

According to elementary information theory the “distance” between this evolved density and the assumed trial density \( \hat{\varrho}(t + \Delta t) \) can be measured by their relative entropy i.e. by \( D(\varrho'(t + \Delta t) || \hat{\varrho}(t + \Delta t)) \).

Now we require the time evolution of \( l \equiv \log \varrho \) to calculate the information loss rate defined using the relative entropy. Unlike the Hamiltonian case for which \( \frac{\partial A_i}{\partial x_i} = 0 \) arbitrary functions of the density do not satisfy equation (4) so we need to compute the evolution of \( l \) explicitly using a Taylor series.

A density evolving according to (4) will satisfy

\[ (T - L^*) \varrho = -L^d \varrho \]

Further from the form of \( L^* \) it follows easily using the chain rule that

\[ (T - L^*) F(\rho) = F' (T - L^*) \rho = -F' \frac{\partial A_i}{\partial x_i} \rho \]

and so

\[ (T - L^*) l = -\frac{\partial A_i}{\partial x_i} \]

or

\[ Tl = L^* l - \frac{\partial A_i}{\partial x_i} \quad (6) \]

now assuming that \( A \) does not depend on time explicitly (the autonomous assumption) then we have after iteration

\[ T^n l = (L^*)^n l - (L^*)^{n-1} \frac{\partial A_i}{\partial x_i} \]

and so

\[ l(t + \Delta t) = e^{\Delta t L^*} l - \left( \Delta t \frac{\partial A_i}{\partial x_i} + \frac{(\Delta t)^2}{2} L^* \left( \frac{\partial A_i}{\partial x_i} \right) + \ldots \right) \]

where we are expanding only to second order since this will be appropriate below.
5. Evolution of a Liouville residual

Consider now a trial density \( \hat{\rho} \) which does not evolve according to equation (4) since it is constrained to lie within the trial density manifold. Denote by angle brackets the expectation with respect to this density at a particular time. For a general random variable \( F \) we have

\[
\frac{\partial \langle F \rangle}{\partial t} - \langle LF \rangle = \langle F_t \rangle + \int F (T - L^*) \hat{l}
\]

\[
= \langle F_t \rangle + \int F (T - L^*) \hat{i}
\]

\[
= \langle F_t \rangle + \langle FR' \rangle
\]

(7)

where we are using integration by parts to re-express \( L \) as the adjoint operator on the first line and using the chain rule on the second line. Choosing \( F = 1 \) we obtain

\[
\langle R' \rangle \equiv \langle R \rangle = 0
\]

(8)

where \( R \) is a generalized Liouville residual since it vanishes if a density evolves according to (4) (see equation (6)). We can rewrite (7) as

\[
\frac{\partial \langle F \rangle}{\partial t} - \langle LF \rangle = \langle F_t \rangle + \int F (T - L^*) \frac{\partial A_i}{\partial x_i}
\]

and now set \( F = R \) obtaining

\[
- \langle LR \rangle = \langle TR \rangle + \langle R^2 \rangle - \int F (T - L^*) \frac{\partial A_i}{\partial x_i}
\]

so

\[
- \langle (L + T)R \rangle = \langle R^2 \rangle - \int F (T - L^*) \frac{\partial A_i}{\partial x_i}
\]

or using (4)

\[
- \langle (T - L^*)R \rangle = \langle R^2 \rangle
\]

or using the definition of \( R \) and \( R' \) as

\[
- \langle (T - L^*)^2 \hat{l} \rangle = \langle R^2 \rangle + \int (T - L^*) \frac{\partial A_i}{\partial x_i}
\]

(9)

5.1. Information loss

For one timestep \( \Delta t \) this can be expressed as (see (1))

\[
IL \equiv D(\rho(t + \Delta t)||\hat{\rho}(t + \Delta t)) = \int \rho(t + \Delta t) \left( l(t + \Delta t) - \hat{l}(t + \Delta t) \right)
\]

where \( \rho \) and \( l \) are the Liouville evolved density and log density over the time step with the starting density and log density being those drawn from the trial density manifold at that earlier time i.e. \( \rho(t) = \hat{\rho}(t) \quad l(t) = \hat{l}(t) \).
Using (4) and the log density counterpart (6) this may be re-expressed to second order as

\[ IL = \int e^{-\Delta tL*}\hat{\varrho}(t) \left( \Delta t \frac{\partial A_i}{\partial x_i} - \left( \Delta t \frac{\partial A_i}{\partial x_i} + \frac{(\Delta t)^2}{2} L^* \left( \frac{\partial A_i}{\partial x_i} \right) + \ldots \right) - e^{\Delta tT}\hat{\varrho}(t) \right) \]

Using integration by parts the first operator in the integrand can be shifted to the right and converted to an exponential of a multiple of the adjoint operator so we get

\[ IL = \left\langle e^{-\Delta tL^*} \left( e^{\Delta tL^*} \hat{\varrho}(t) - \left( \Delta t \frac{\partial A_i}{\partial x_i} + \frac{(\Delta t)^2}{2} L^* \left( \frac{\partial A_i}{\partial x_i} \right) + \ldots \right) - e^{\Delta tT}\hat{\varrho}(t) \right) \right\rangle \]

where the angle bracket is an expectation with respect to the trial density at time \( t \) and on the second line we are using the fact that \( T \) and \( L^* \) commute due to the autonomous dynamical system assumption. Expanding out the remaining exponential operators as Taylor series and retaining only terms to second order:

\[ IL = \left\langle -e^{-\Delta tL^*} \left( \Delta t \frac{\partial A_i}{\partial x_i} + \frac{(\Delta t)^2}{2} L^* \left( \frac{\partial A_i}{\partial x_i} \right) \right) \right\rangle \]

where on the second line we are using equations (8) and (9) as well as the autonomous assumption. There is a remarkable number of cancellations and a very simple result which nicely extends the Hamiltonian case. Recall that \( R \) vanishes for time evolution according to the Fokker Planck equation (4).

6. Path integral Lagrangian

Since \( IL \) has a straightforward information theoretic interpretation we set therefore

\[ \mathcal{L} = \frac{1}{2} (\Delta t)^2 \langle R^2 \rangle \hat{\varrho}(t) \]  

(10)

and by analogy with the Onsager and Machlup approach, identify the action \( S \) for this path

\[ S = \frac{\Delta t}{2} \int \mathcal{L} dt \]

as the analog of the entropy for our proposed path Boltzmann principle. To make further progress we need to specify more concretely the trial density. A convenient choice here is to use maximum entropy theory with respect to an equilibrium density obtaining

\[ \hat{\varrho}(\lambda, x) = Z^{-1}(\lambda) \exp \left( \lambda^T H(x) - \beta \psi(x) \right) \]
where the function $\psi$ is the analog of invariant quantities for a Hamiltonian system (often energy in that case). It may only be known approximately as a quadratic function and need not be assumed necessarily to be an invariant of the dynamical system. The vector $H$ is an appropriately chosen set of slow variables for the system whose moments interest us. For practical reasons they are usually chosen to be quadratic functions since calculation of expectations with respect to a trial density often requires a Gaussian density for tractability.

With this choice the Lagrangian is easily evaluated up to an irrelevant additive constant as

$$L = \frac{(\Delta t)^2}{2} \left( \lambda^i g_{ij} \dot{\lambda}^j - 2 \dot{\lambda}^i M + \phi + 2 \dot{\lambda}^i X - 2 \lambda^i Y \right)$$

$$g_{ij} \equiv \langle H_i, H_j \rangle$$

$$M_i \equiv \langle L^* H_i \rangle$$

$$\phi \equiv \lambda^i \langle L^* H_i L^* H_j \rangle \lambda^j$$

$$X_i \equiv \langle (H_i - \langle H_i \rangle) \Gamma \rangle$$

$$Y_i \equiv \langle (L^* H_i) \Gamma \rangle$$

$$\Gamma \equiv \frac{\partial A_i}{\partial x_i} - \beta L^* \psi$$

The non-negative definite matrix $g(\lambda)$ is the Fisher information matrix for the random vector $H$ whose trial densities are the given maximum entropy statistical model family. Thus it is also the metric tensor of the trial density manifold. Note also that the scalar function $\Gamma$ here will vanish identically if $\psi$ happens to be chosen an invariant function for the dynamical system. In the case it is a quadratic approximation to such an invariant then $\Gamma$ will be expected to be small and hence the linear terms in the Lagrangian will be also. Finally it is worth emphasizing that due to (10), the action is always bounded below and so the Euclidean path integral we have proposed here is well defined despite the complexity of equation (11).

It is clear now that our first and second requirements for a path measure from section 2 above have been met since all these fields are specified as functions of $\lambda$ (and hence the moments of the slow variables via a Legendre transformation) and a Lagrangian defined based on a clear information theoretic formulation which generalizes entropy from the equilibrium Boltzmann principle. The third requirement holds in a certain sense when the system is Hamiltonian but the more general autonomous case considered here requires further investigation.

7. Paths and endpoints

We have seen how to associate a weight to a path using the information loss action. It is clear however that we further require a non-negative weight for any trial density at a prescribed endpoint time in order to evaluate it’s significance as an approximation for the true probability density at that time. The path integral proposed by Onsager and Machlup and generalized here, clearly achieves this objective in a natural manner (see equation (3)). The trial density weight at a fixed time is termed a consistency distribution and given it’s formulation as a path integral is analogous to a quantum wavefunction. Indeed it satisfies a (Wick rotated) Schrödinger equation (see [1]).

It is important to emphasize the significance of the path integral here. It is the sum of all path weights with fixed endpoints that is used to define the consistency distribution. Such a path integral is not in general a monotonic function of the extremal action which applies to a path satisfying the Euler-Lagrange equations for $L$. Such an action is termed the classical action $S_{cl}$ in
the path integral literature. In the special case that the Lagrangian $\mathcal{L}$ is a \textbf{quadratic} function of $\lambda$ it may however be proven (see e.g. \cite{10} Chapter 6) that

$$\int D\lambda W[\lambda] = B \exp (-S_{cl})$$

with $B$ not dependent on the value of $\lambda$ at the path endpoint. Clearly in that simple case (which also happens to be the original Onsager-Machlup one) only the extremal actions are important in defining the consistency distribution. In most realistic situations however $\mathcal{L}$ is not quadratic and so this simplifying situation may not apply.

More concretely, one can imagine a scenario when the extremal paths between two sets of endpoints have the same minimal action but the path integral is quite different. This situation is sketched in Figure 1.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure1.png}
\caption{The importance of a path integral versus extremal actions in defining the consistency distribution. In the example sketch we have the situation that $W_3, W_4 \ll W_1, W_2 < W_{max}$ which is when "quantum" effects are significant.}
\end{figure}
References

[1] R. Kleeman. A path integral formalism for non-equilibrium Hamiltonian statistical systems. *J. Stat. Phys.*, 158(6):1271–1297, 2015.

[2] B. Turkington. An optimization principle for deriving nonequilibrium statistical models of Hamiltonian dynamics. *J. Stat. Phys*, 152:569–597, 2013.

[3] D. N. Zubarev. *Nonequilibrium Statistical Thermodynamics*. Plenum Press, New York, 1974.

[4] R. Kleeman and B. E. Turkington. A nonequilibrium statistical model of spectrally truncated Burgers-Hopf dynamics. *Comm. Pure Appl. Math.*, 67(12):1905–1946, 2014.

[5] A. J. Majda and I. Timofeyev. Remarkable statistical behavior for truncated Burgers-Hopf dynamics. *Proc. Nat. Acad. Sci. USA*, 97:12413–12417, 2000.

[6] L. Onsager and S. Machlup. Fluctuations and irreversible processes. *Phys. Rev.*, 91(6):1505–1512, 1953.

[7] H. Risken. *The Fokker-Plank Equation*. Springer Verlag, Berlin, 2nd edition, 1989.

[8] H. C. Oettinger. *Beyond equilibrium thermodynamics*. Wiley-Interscience, Hoboken, New Jersey, 2005.

[9] S. Amari and H. Nagaoka. *Methods of Information Geometry*. Translations of Mathematical Monographs, AMS, Oxford University Press, 2000.

[10] L. S. Schulman. *Techniques and applications of path integration*. Dover, 2005.