Richard Kleeman

A path integral formalism for non-equilibrium Hamiltonian statistical systems

Received: date / Accepted: date

Abstract  A path integral formalism for non-equilibrium systems is proposed based on a manifold of quasi-equilibrium densities. A generalized Boltzmann principle is used to weight manifold paths with the exponential of a multiple of the information discrepancy of a particular manifold path with respect to full Liouvillean evolution. The likelihood of a manifold member at a particular time is termed a consistency distribution and is analogous to a quantum wavefunction. The Lagrangian here is not of Onsager-Machlup form, however at large times approaches one with the thermodynamics being of a generic Öttinger form. The proposed path integral has connections with those occurring in the quantum theory of a particle in an external electromagnetic field. It is however entirely of a Wiener form and so practical to compute. Finally it is shown that providing certain reasonable conditions are met then there exists a unique equilibrium consistency distribution.

Keywords  Non-equilibrium · Path Integral · Closure

PACS  05.20.-y · 03.65.-w

1 Introduction

Around sixty years ago Onsager and Machlup (OM) \cite{24} proposed a near equilibrium variational principle for determining the likelihood of time dependent fluctuations in statistical systems in equilibrium. This principle is formulated as a Wiener path integral and the associated stochastic process
is easily shown to be Ornstein Uhlenbeck. Formally the path amplitudes $W$ are given by

$$W(\lambda(t)) = C \exp \left[ -\eta \int_0^t \left( \dot{\lambda} - U\lambda \right)^T g \left( \dot{\lambda} - U\lambda \right) dt \right] \equiv C \exp \left[ -\eta \int_0^t \mathcal{L}(\dot{\lambda}, \lambda) dt \right]$$  \hspace{1cm} (1)$$

where the vector path $\lambda(t)$ lies in an appropriate vector space of thermodynamical variables and the $g$ and $U$ are constant matrices with the former non-negative definite. A particularly attractive feature of the formulation in (1) is that a positive definite $g$ implies that a trajectory which is a solution of

$$\dot{\lambda} = U\lambda$$  \hspace{1cm} (2)$$
is the most probable path and thus serves to determine thermodynamical paths. Random variations from this trajectory are controlled by the above specified path amplitude and constitute a theory of near equilibrium fluctuations.

Two natural questions arise from this seminal formulation

1. Can this principle be extended to far from equilibrium systems and if so how exactly? One might hope that a Lagrangian of the same general form might be possible with the vector $U$ and matrix $g$ being generalized to being state dependent. This question occupied the attention of those concerned with general (as opposed to Ornstein Uhlenbeck) Markov stochastic processes in the 1970s. It was discovered (e.g. [15] and [14]) however that the Lagrangian given in (1), as well as requiring state dependent $g$ and $U$, also required the addition of several other terms. The nature of those terms depended crucially on the time discretization procedure used to rigorously define the path integral. An attractive feature of the contribution by [14] was that the Lagrangian could be cast into a covariant form with the various quantities becoming tensors for a Riemannian manifold determined by regarding $g$ as a metric tensor. The author and collaborators showed that the additional terms required in the Lagrangian had a natural manifold interpretation (one was a multiple of the Ricci scalar) and moreover that a natural time discretisation procedure could be specified to give the derived Lagrangian using renormalization theory (see [9]).

2. How should the thermodynamical variables be selected? This problem of definition appears to have no unique solution in general but is always connected to issues of slow and fast timescales. In general one expects the predicted probability density for the thermodynamical variables to be a good approximation in the complete system given a time average of the system of sufficient length.

Unsurprisingly the above two questions have received considerable attention in the literature where many other approaches have been proposed. Examples of this work include [21], [28] and [3]. In this contribution we shall propose an answer to these questions based upon the classical approach to non-equilibrium statistical systems of Zubarev [31] and its recent extension by Turkington (BT) [29]. In that work non-equilibrium marginal distributions of system slow variables are approximated with maximum entropy trial
densities. Such densities are obtained, as suggested by Jaynes [18] and others, by specifying a finite set of slow variable moments and assuming maximal ignorance with respect to other moments. This implies that the densities belong to the manifold of a general exponential family. Co-ordinates on such a manifold can be specified either using the original constraining moments or the associated Lagrange multipliers for each constraint. The implicit assumption behind such approaches is firstly that if a sufficiently long time average of the system is taken then the resulting system density will be close in some sense to a marginal distribution of slow variables multiplied by an equilibrium (Gibbs) density. Secondly it is assumed that the system (true) density can be adequately approximated by a member of a suitable exponential family.

The question then arises as to how they evolve from one to another on the particular slow time averaging scale. For Hamiltonian (and other) dynamical systems the full densities evolve according to the Liouville equation. Applying the Liouville propagator to the trial densities results in generally a density outside the chosen manifold of trial densities. One can measure the discrepancy between this evolved density and the trial manifold points using a distance function on the density space with natural choices deriving from information theory. Consider now a continuous sequence of such trial densities through time. The time integral of the information discrepancies are a measure of the consistency of a particular path within the trial manifold with the underlying Hamiltonian dynamics. BT [29] showed that the discrepancy tendency at a particular time can be formulated as a Lagrangian function of the manifold co-ordinates \( \lambda \) and their time derivatives \( \dot{\lambda} \). A thermodynamical path through the manifold was then determined using this Lagrangian and an extremization principle in a manner similar to classical Hamiltonian mechanics.

This closure was tested numerically by Kleeman and Turkington (KT) [19] in a dynamical system which has often served as a simple model of turbulence: A spectrally truncated Burgers-Hopf (TBH) model which obeys Hamiltonian dynamics. TBH has the attractive property that the equilibrium statistical density has been shown numerically to be given by a simple Gaussian Gibbs measure. The system is also a rather stringent test of the formalism because the decorrelation timescales of the spectral modes vary inversely with wavenumber which means that there is not a clean separation between fast and slow variables. Nevertheless the formalism developed performed reasonably well in predicting the time evolution of the means of the slow (low wavenumber) spectral modes both in a situation close to equilib-

---

1 These two coordinate systems are related to each other as usual by a Legendre transform.
2 An important exception of course being the equilibrium Gibbs density which is an invariant measure.
3 It is important to emphasize that the Hamiltonian mechanics here refers to the coarse grained coordinates or moments of the approximating distributions rather than the fine grained variables of the original complete Hamiltonian system.
4 The energy function for TBH is simply the sum of the squares of the spectral mode amplitudes meaning the Gibbs measure is a Gaussian with uncorrelated modes and equal variances proportional to the conserved energy of the system.
rium and moderately removed from it. In particular after initialisation with a member of the trial density family, the closure predicted two qualitative features of the relaxation with high accuracy:

1. The relaxation time to equilibrium is proportional to the inverse wavenumber which as noted is proportional to the spectral mode decorrelation time.
2. The modal relaxation is characterised by an initial “plateau” period in which dissipation increases followed by an exponential decay to equilibrium via an asymptotic dissipation. The plateau period occupied the same significant fraction of the relaxation time for all modes.

The second property has fundamental implications for the macrostate description of the system. If the system is restarted at a particular time after the original start time using the trial density implied by the optimal \( \lambda \) co-ordinates at that time then, in general, it will follow a different path to the original experiment. Such behaviour occurs in both the direct numerical simulations and in the theoretical solutions. It occurs theoretically because a period of increasing dissipation is always evident for a system initialised with a trial density. Consequently the macrostate co-ordinates \( \lambda \) of the system at this evolved later time are insufficient to fully specify the future slow random variable evolution. This reflects the fact that at this evolved time, the identified exponential family member is only ever an approximation to the true density of time averaged variables of the system. It is rather curious that for the system to equilibrate maximally this approximate behaviour appears essential. The kind of non-Markovian behaviour just noted is also an intrinsic part of other non-equilibrium theories such as that of Mori-Zwanzig (see [32] and [3]).

In this contribution we shall propose that the macrostate is better specified using a (positive) consistency distribution\(^5\) of the exponential manifold co-ordinates. When such a distribution is given at a particular time, the future macrostate evolution of the system can be computed uniquely. In some respects this approach is analogous to quantum mechanics where a wave function at a given time is sufficient via the Schrödinger equation to specify the future state of the system. Indeed as we shall see, the mentioned consistency distribution may be derived in a natural way from a path integral in the same basic way that a quantum wave function is derived from a Feynman path integral. The Lagrangian involved is that derived previously by BT and discussed above. As usual in statistical mechanics this path integral is of a Wiener rather than complex Feynman type.

In the next section we derive the information loss implied in the choice of a particular time sequence of trial approximating densities. This loss has an interesting decomposition due to information geometry into pieces related to reversible and irreversible paths within the manifold.

\(^5\) Note that we use the terminology distribution here to avoid confusion with the approximating trial densities. The consistency distribution is a function (or distribution) of the coarse grained variables which specify the position within the manifold of trial densities. The densities are defined on the original fine grained variables of the problem.
In section 3 we use this derived information loss Lagrangian to propose a path integral formulation for the problem at hand using a generalized Boltzmann principle. A very simple pedagogical example is given to illustrate fundamental behavior.

An essential part of the path integral derivation is a clear distinction between manifold paths and their endpoints: For the Lagrangian derived here the extremal or classical trajectories for the problem depend on the temporal endpoint chosen in a non-trivial fashion. This dependency implies that the locus of extremal endpoints is not itself an extremal. Such behavior distinguishes the Lagrangians proposed here from those of Onsager-Machlup form where the extremal endpoints always form an extremal path. This behavior is also a reflection of the macrostate ambiguity mentioned above since at a given time there is a continuum of differing manifold points corresponding to extremals with differing temporal endpoints. In section 4 we compare our derived Lagrangian with one of Onsager-Machlup form using a transformation due to Roncadelli [26] and clarify exactly how they differ. We also consider the situation at asymptotically large times and show that then an Onsager-Machlup Lagrangian does actually become appropriate for defining a time propagator of consistency distributions. Such asymptotic times are presumably when the system is close to equilibrium which is consistent with Onsager-Machlup theory being relevant to near equilibrium relaxation.

In section 5 we show that the Lagrangian derived is the same as that for a non-relativistic particle moving in an external magnetic field as well as an external potential. The particle moves in a manifold specified by a metric tensor given by the Fisher information matrix \( g \) of the exponential family assumed.

In section 6 we consider the Schrödinger equation associated with the proposed path integral. In section 7 we note the similarity and differences to the Wick rotated electromagnetic path integral of quantum statistical mechanics. In section 8 we consider the associated (time) transfer operator and show using compact operator theory that there exists a unique consistency distribution associated with equilibrium. Section 9 contains a discussion.

2 Path Liouville discrepancy

Since we intend invoking a generalized Boltzmann principle in the next section we derive here an information theoretic based measure the discrepancy of a time sequence of trial densities from Liouvillean evolution. The reader is also referred to the earlier work BT where this idea was first introduced for more detail and a somewhat different approach.

Suppose we are dealing with a Hamiltonian dynamical system with the symplectic evolution equation for a general variable given by:

\[
\frac{dF}{dt} = \{F, H\} + \frac{\partial F}{\partial t}
\]

where \( H \) is the system Hamiltonian and the Poisson bracket is given by

\[
\{A, B\} = (\nabla A)^T J \nabla B
\]
with the gradient taken with respect to the dynamical variables and the matrix $J$ is antisymmetric which ensures the bracket is antisymmetric with respect to its two arguments. A (smooth) probability density $p$ on this dynamical system satisfies the Liouville equation

$$\frac{\partial p}{\partial t} + L p = 0 \quad (4)$$

with the operator $L$ anti-Hermitian with respect to the usual Hilbert space inner product.

Consider the anti-Hermitian differential operators

$$L \equiv \frac{\partial H}{\partial x_i} J_{ij} \frac{\partial}{\partial x_j} \quad T \equiv \frac{\partial}{\partial t}$$

where $x_i$ are the basic (fine grained) dynamical system variables. We assume that these operators commute i.e. that the gradient of $H$ and $J$ do not depend explicitly on $t$. Denote now a trial density by $\hat{p}$ and consider various temporal evolutions over a short interval $\Delta t$ which is however assumed sufficiently long that unresolved degrees of freedom decorrelate. The evolution according to the Liouville equation (4) will be

$$p(t + \Delta t) \equiv e^{-\Delta t L} \hat{p}(t)$$

Now in general this evolved density will lie outside the manifold described by trial densities. The evolved trial density must therefore be the different density

$$\hat{p}(t + \Delta t) = e^{\Delta t T} \hat{p}(t)$$

The information lost $IL$ in assuming $\hat{p}(t + \Delta t)$ when in fact the density is $\bar{p}(t + \Delta t)$ is simply the relative entropy $D(\hat{\cdot} \| \bar{\cdot})$ of the second density with the first. We have now the following

$$IL = D(e^{-\Delta t L} \hat{p} || e^{\Delta t T} \hat{p})$$

$$= \int \hat{p} \left( e^{-\Delta t L} \bar{p} - e^{\Delta t T} \bar{p} \right)$$

$$= \left\langle e^{\Delta t L} \left( e^{-\Delta t L} - e^{\Delta t T} \right) \hat{\bar{1}} \right\rangle_{\hat{p}}$$

$$= \left\langle \left( I - e^{\Delta t L} e^{\Delta t T} \right) \hat{\bar{1}} \right\rangle_{\hat{p}} \quad (5)$$

$$= \left\langle \left( I - e^{\Delta t (T + L)} \right) \hat{\bar{1}} \right\rangle \quad (6)$$

with $\hat{\bar{1}} \equiv \log \bar{p}$. On the second line we are using the fact that an arbitrary function of $p$ also obeys the Liouville equation (4); on the third line we are using the anti-Hermitian property for $L$; and on the last line we are using $[L, T] = 0$ and the expectation refers to the trial density at the start of

---

* If the trial distribution gives an invariant measure for the system this will not be the case.
the propagation interval. Define now the following useful random variable $R$ which we call the Liouville residual

$$ R(p) \equiv (T + L) \log p \quad (7) $$

Note that for a probability evolving according to the Liouville equation, $R$ vanishes but will not in general for a $\hat{p}$ constrained to lie within the trial density manifold. A general random variable $F$ can be shown (see Appendix) to satisfy the following evolution equation

$$ \frac{\partial \langle F \rangle}{\partial t} - \langle LF \rangle = \langle TF + FR \rangle $$

from which we deduce (setting $F = 1$) firstly that

$$ \langle R \rangle = 0 \quad (8) $$

and secondly (setting $F = R$) that

$$ \langle (T + L) R \rangle = -\langle R^2 \rangle \quad (9) $$

Returning now to equation (6) we expand the exponential operator as a Taylor series. The terms in $\Delta t$ of order zero and one vanish due to cancellation and equation (8) while the order two term remains and using (9) we derive the remarkably simple second order approximation

$$ IL = \frac{(\Delta t)^2}{2} \langle R^2 \rangle + O((\Delta t)^3) $$

Thus the information loss to lowest order is simply proportional to the variance of the Liouville residual $R$. It is worth observing that this loss is quadratic in the time interval $\Delta t$ which is consistent with the relative entropy geometrically being a distance squared (see [1]).

In order to make further progress beyond this general equation we now specify the trial density manifold $T$. We identify a subset of functions $A$ (assumed a vector) from the dynamical system which we label as the resolved (or coarse grained) variables. In general these will be functions of the slow variables for the dynamical system. Secondly we assume that equilibrium densities are of a Gibbs type and for simplicity we assume that the only invariant involved here is the energy. The general trial density is then deduced by minimizing the relative entropy with respect to the Gibbs density under the assumption that the resolved variable expectations are known. They therefore take the form

$$ \hat{p}(t) = \exp \left[ \lambda(t)^t A - G(\beta, \lambda) - \beta E \right] \quad (10) $$

where $E$ is the energy of the system which we are assuming is one of the resolved variables and satisfies $LE = 0$. Note also that $G$ normalizes the distribution and the partition function $Z = \exp G$. In addition there is a one to one relationship between the co-ordinates of the manifold $\lambda$ and the expectation values $\hat{a}$ of the chosen $A$. Either can serve as co-ordinates for the
trial distribution manifold and are related by a Legendre transform (see, for example, [1]). With this specification it is easy to calculate $R$ as

$$R = \dot{\lambda}^i (A - a) + \lambda^i L A$$

where the overdot denotes a time derivative and hence that

$$IL = \frac{(\Delta t)^2}{2} \left( \dot{\lambda}^i g_{i\dot{j}} - 2 \dot{\lambda}^i (LA) + \phi \right) + O((\Delta t)^3) \tag{11}$$

$$\phi \equiv \lambda_i \langle (LA_i LA_j) \rangle \lambda_j$$

$$g_{ij} \equiv \langle (A_i - a_i) (A_j - a_j) \rangle$$

The matrix/tensor $g$ here is the Fisher information matrix which plays a central role as a Riemannian metric tensor in the field of information geometry (see [1]). We have also used the following identity derived in Appendix A:

$$\langle LA_i \rangle = -\lambda_j \langle (A_i - a_i) LA_j \rangle$$

There is an interesting decomposition of the information loss $IL$ which relates both to reversible thermodynamics and to the basic information geometry we are considering. The entropy $S$ along a general trajectory may easily be computed as

$$S = -\langle \log \hat{p} \rangle = -\lambda^t a + G + \beta u$$

$$u \equiv \langle E \rangle$$

Taking the time derivative we obtain (see Appendix A)

$$\dot{S} = -\lambda^t g \dot{\lambda} + \beta \dot{u} \tag{12}$$

Suppose we now define a particular trajectory in our trial distribution manifold which satisfies the following first order differential equation:

$$\frac{d\lambda}{dt} = g^{-1} \langle LA \rangle \tag{13}$$

where $\lambda$ is used to distinguish this particular trajectory from a general trajectory which we write simply as $\lambda$. Obviously a specification of co-ordinates for a given time will then specify the particular trajectory given equation (13). Combining equations (12) and (13) we obtain for this particular trajectory that (see Appendix A):

$$\dot{S} = \beta \dot{u}$$

which is the usual expression for reversible entropy change in an open system with varying mean energy. We therefore identify the particular trajectory above as a reversible trajectory. The information loss along this reversible trajectory can be computed simply by substituting (13) into (11) giving to second order accuracy

$$IL_{rev} = \frac{(\Delta t)^2}{2} \left( \phi - \langle LA \rangle^t g^{-1} \langle LA \rangle \right) \tag{14}$$
Finally we can compute $IL_{irr}$ the relative entropy between a reversible and a general irreversible trajectory within our manifold. Since both lie within the manifold their relative entropy can be calculated to second order accuracy by the following well known relation in information geometry between relative entropy and the Fisher metric (see [1]):

$$D(\tilde{\rho}(\lambda) || \tilde{\rho}(\lambda + \epsilon v)) = \frac{\epsilon^2}{2} v^t g v + O(\epsilon^3)$$

thus to second order accuracy we obtain, using the defining relation for a reversible trajectory

$$IL_{irr} = \frac{(\Delta t)^2}{2} \left( \dot{\lambda} - g^{-1} (LA) \right)^t g \left( \dot{\lambda} - g^{-1} (LA) \right)$$  \hspace{1cm} (15)$$

It is now trivial to verify the following interesting relation between various information losses which is accurate to second order:

$$IL = IL_{rev} + IL_{irr}$$  \hspace{1cm} (16)$$

The non-negativity of relative entropy now shows that over the timestep $\Delta t$ the information loss to second order can be minimized to $IL_{rev}$ by choosing the reversible trajectory. The endpoint of the reversible trajectory can thus be viewed as a projection from the fully Liouvillian evolved initial trial distribution back into the trial manifold. $IL_{irr}$ represents the information loss in not choosing this infinitesimally optimal reversible trajectory while $IL_{rev}$ represents the minimum possible information loss for all trajectories. The full situation is depicted schematically in Figure 1. It should be clear however that if one chooses a large number of timesteps that the reversible trajectory will no longer in general minimize information loss since $IL_{rev}$ clearly depends on the trajectory chosen and there are usually irreversible trajectories which result in small values of this quantity at a given time than that occurring on the reversible trajectory. The relation (16) has been discussed at length in information theoretic contexts (see [1] and [7] Chapter 11) where it is referred to as the relative entropy Pythagorean relation since this functional is best viewed as a distance squared. Note that the decomposition above was first discussed in BT in a somewhat different context. Here we have emphasized the information theoretic perspective for reasons that will become apparent when we turn to the path integral formalism in the next section.

The relevant dynamical object of interest is, of course, a long time path in the trial distribution manifold. The total informational discrepancy of interest is then simply proportional to the sum of each $IL$ along the time interval. Mathematically it is convenient to pass partially to the infinitesimal time limit in which case this becomes the time integral of a Lagrangian i.e. the action

$$S = \Delta t \int_0^T \mathcal{L} dt$$  \hspace{1cm} (17)$$

\text{7} Strictly this identification as a projection is precise only in the limit as $\Delta t \to 0$
Fig. 1 Information loss decomposition to second order accuracy. Liouville evolution takes a distribution $A$ in the trial manifold $T$ to the distribution $E$ which lies outside $T$. The “nearest” distribution in $T$ to $E$ is $B$ in the sense that it has minimum relative entropy $D(E||*)$. Thus $B$ may be considered a projection of $E$ into $T$. A general distribution $C$ in the trial manifold differs from $B$ by the relative entropy of $D(B||C)$ and the projective nature of $B$ ensures that $D(E||C) = D(E||B) + D(B||C)$ or $IL = IL_{rev} + IL_{irr}$. This relation is known in information geometry as a Pythagorean relation since relative entropy for small displacements within the manifold $T$ can be regarded as a squared distance.

$$\mathcal{L} \equiv \frac{1}{2}\left(\dot{\lambda}^2 g \lambda - 2\dot{\lambda} M + \phi\right)$$
$$M \equiv \langle LA \rangle$$

Notice that the timestep $\Delta t$ enters into the final result as a consequence of the information loss (relative entropy) being geometrically a distance squared.

Finally it is worth observing that a somewhat more general formulation than above has been proposed and tested in BT and KT. There the two parts of the information loss $IL_{rev}$ and $IL_{irr}$ are weighted differently. This was in recognition of the fact that the formalism being considered is an idealisation in two important respects:

Firstly in reality the fast and slow time scales are never cleanly separated. Secondly there is arbitrariness in how resolved variables $A$ are selected from functions of the system slow variables. In the two concrete dynamical systems examined to date in KT and BT it has been found convenient to chose the weighting somewhat differently than the unit ratio in (16). For the truncated Burgers-Hopf turbulence system investigated in KT the optimal weighting for agreement with direct numerical simulations of the full system was found
by increasing the weight of $IL_{rev}$ to around 1.3. In that case however the set of resolved variables was simply the slow, small wave number spectral modes. Since in direct simulations of the full system, slow mode variance variation is apparent, such a set of resolved variables may well be too restrictive and the set should be extended to include quadratic functions of the slow modes.

Notice that if we ignore $IL_{rev}$ altogether in the decomposition then it is easily seen that the reversible trajectory results from minimization. These issues will be examined in more depth in future publications by considering the convergence issue of larger sets of resolved variables $A$ and also by analyzing a range of different dynamical systems.

3 Path integral formulation

In the previous section we have identified in a rather natural way an arbitrary differentiable path in the manifold $T$ of trial densities with a non-negative information loss. Thus from this calculation there exists an obvious way of weighting paths with respect to optimality. We are however not interested in path optimality directly. Instead we are interested in best describing the statistical system at a particular time.

One reasonable approach to this issue is to use the path information loss functional to identify the best approximating trial density at the time of interest i.e. find an optimal co-ordinate in $T$. Now clearly any particular point in this manifold at the time of interest may be reached by an infinite number of different paths. Thus the weight attached to this point should clearly be some function of the information loss of all such paths. One possible approach here might be to simply take the unique optimal path defined by the endpoints via the Euler-Lagrange equations corresponding with our Lagrangian and then weight the endpoint with that of the identified extremal. Considering then a second optimization on such weights we obtain an optimal set of coordinates in $T$ at time $t$. This seems a rather obvious approach for identifying a trajectory of macrostates for the system however there is an inconsistency caused by the two stage optimization. The optimal set of extremal endpoints may well not be an extremal path itself. Indeed if we consider the extremal path to a later time $t' > t$ the point on that path at time $t$ may well differ from the already identified optimal point for time $t$. If a statistical system has this property then there is clearly an ambiguity in identification of the best approximating macrostate at a particular time.

Suppose now we consider a system for which the Lagrangian may be written in OM form i.e. in the form of the path weights from equation (1). Clearly the extremal path is unique given only the initial conditions and is given by the solution of (2). Evidently then the optimal endpoints and extremal paths correspond exactly and there is no macrostate ambiguity. However it is clear that in general the Lagrangian identified in the previous section is unlikely to be of such a form since the extremal path would have an information loss which was independent of the path length. Of course for the special case that the initial state was the equilibrium state this would be fine since then such a trial density would be a Gibbs measure and so give zero information loss since it is a solution of the Liouville equation. Aside
from this special case however, it seems there is an inconsistency between the Lagrangians (1) and (17) and this is due to the positive information loss incurred by even an extremal path under general circumstances. As we shall see below, such an inconsistency can imply that extremal and optimal paths do not in general coincide and furthermore it is a property of very simple Lagrangians of clear relevance to the equilibration process.

Given such an ambiguity of the macrostate we consider instead a positive distribution on $T$ at each time $t$ as a more appropriate macrostate representation. Further with regard to defining this distribution at each time, we consider the general set of paths leading to a particular point rather than only the extremal since other paths are intuitively relevant as well. Motivated by the general utility and fundamental nature of path integral approaches across many applications in mathematical physics, we adopt a path Boltzmann principle and assign a positive Wiener path measure by

$$W(path) = C \exp \{-\Delta t S(path)\}$$

Thus the slow timescale $\Delta t$ can be considered intuitively (but obviously not rigorously) as an inverse Planck constant. The integral then over all suitable paths of this measure is given by the usual Wiener path integral and is termed the amplitude $K(\lambda(t_2), \lambda(t_1))$ to go from one point in $T$ at $t_1$ to another at a later time $t_2$. Clearly as usual in path integral theory in the classical limit $\Delta t \to \infty$ such an integral is completely dominated by the extremal path meaning we return effectively to the initial optimization scenario discussed above with the extremal action determining endpoint weight since the exponential function is monotonic.

The consistency distribution at time $t_2$ may then be defined, as in most path integral approaches, as the integral of this amplitude multiplied by the consistency distribution at $t_1$. There remains then the issue of identifying the appropriate consistency distribution at the initial time. Now obviously we can, as a practical matter, specify the initial probability density exactly from the manifold of trial distributions. Given this knowledge the obvious choice for an initial consistency distribution is simply a Dirac delta function centered on the manifold point chosen. One may evidently consider other choices for the initial density which do not lie within the trial distribution manifold. We defer consideration of that case to a later publication.

Given a positive consistency distribution at a particular time the obvious approach to identifying the best approximating trial density is via maximization. Other approaches may also be possible since, as was pointed out in section 1, this optimal trial density is clearly only an approximation to the true marginal density and moreover the very approximate nature is important to the equilibration process due to the plateau effect discussed there. One might expect therefore that the full nature of the consistency distribution may contain further information which could be used to obtain a better estimate of the desired moments of slow system random variables. We again defer discussion of this interesting possibility to a later publication and for

\[\footnote{Indeed the optimal non-extremal path gives a natural plateau effect as we shall see below and this seems highly relevant to practical problems as discussed in the introduction. The extremal path does not give such a plateau.}\]
the remainder of this publication consider the maximum of the consistency distribution as identifying the best approximating thermodynamics for the system.

3.1 A simple pedagogical example with macrostate ambiguity and plateau behaviour.

In order to gain some concrete insight into the formalism proposed we now consider the simplest relevant case namely that for exponentially damped relaxation to equilibrium. Analysis in KT indicates that a straightforward generalization of this system is relevant to the near equilibrium relaxation of the TBH system. As we shall see below this very simple system exhibits precisely the macrostate ambiguity discussed above as well as the plateau behaviour discussed in the introduction. The Lagrangian here is given by

\[ 2\mathcal{L} = \ddot{u}^2 + k^2 u^2 \]

which has the Euler Lagrange equation

\[ \ddot{u} = k^2 u \]

The solution of these equations with fixed endpoints is

\[ u(t) = Ae^{kt} + Be^{-kt} \]

\[ B = \frac{1}{2} \frac{u(0)e^{kt} - u(T)}{\sinh(kT)} \quad A = u(0) - B \]  

Note the importance of not just the damped solution but also the exponential growing one. The classical action can now be computed with a little algebra

\[ S_{cl}(k, T) = \frac{k}{2} \left[ \coth(kT) (u(0)^2 + u(T)^2) - 2u(0)u(T)\text{csch}(kT) \right] \]

which is a very standard result (see e.g. [13] equation (10.44)). Suppose we fix \( u(0) \) then this action is minimized by a \( u_m(T) \) satisfying

\[ u_m(T) = u(0)\text{sech}(kT) \]  

If we set \( k = i\omega \) the system above becomes a standard harmonic oscillator for which the Feynman path integral is well known [13] to be simply

\[ K_F(u(0), u(T)) = C \exp \left( \frac{i}{\hbar} S_{cl}(i\omega, T) \right) \]

which implies that the Wiener path integral for this problem is

\[ K(u(0), u(T) = C \exp \left( -\Delta t S_{cl}(k, T) \right) \]

which is a Gaussian density whose peak is obviously given by equation (22). Thus in this very simple case the optimal trajectory does not depend on
the slow timescale $\Delta t$ since it is obtained by simply minimizing the “classical” action. For more complicated Lagrangians appropriate to conditions far from equilibrium however this will not necessarily be the case as is also not the general situation in quantum mechanics.

Suppose now we set $u(T) = u_m(T)$ then it is easy to see from (21) and (22) that for $t < T$ we have $u(t) \neq u_m(t)$. Furthermore if one restarts the system at $u_m(t)$ then the future thermodynamical trajectory differs markedly from the original. This is illustrated in top panel of Figure 2 for $k = u(0) = 1$. Note in both cases the initial plateau in the equilibration before exponential decay occurs. This behaviour is qualitatively the same as seen in DNS simulations of the truncated Burgers turbulence system analyzed in [19]. This situation suggests intuitively that the quasi-equilibrium distribution at time $t > 0$ can only be an approximation to the actual marginal distribution for that time. This can be seen concretely by computing the consistency distribution which is proportional to $\exp(-\Delta t S_{cl})$. The results are shown in the bottom panel of Figure 2 for $\Delta t = 1$ where it is clear that at the restart time there is a rather broad distribution.

Another interesting aspect of the solutions is that the optimal trajectory $u_m(t)$ does not correspond with the extremal trajectory $u(t)$ that produces the endpoint optimal action. This is demonstrated in Figure 3 for an endpoint close to equilibrium relative to the initial conditions. Thus the trajectory with the minimal total information loss does not correspond with the sequence $u_m(t)$. Interestingly the latter appears “more realistic” in that it exhibits the spinup character universally noted in DNS solutions and discussed earlier.

4 Transformation to an Onsager Machlup like path integral

The Lagrangian specified by equation (18) is not of the same generic OM form of (1). There exists however an interesting transformation that illuminates the relationship between the two which was originally suggested in the quantum context by Roncadelli [26]. The transformation can also be viewed as a gauge transformation in the sense of electromagnetism as we shall see in the next section. Suppose we add to the Lagrangian a term $df \frac{dt}{dt} - \dot{\lambda} \nabla f - \frac{\partial f}{\partial t} = 0$ where $f(\lambda, t)$ is to be determined. The extra terms allow us to “complete the square” in the Lagrangian as follows:

$$\frac{1}{2} \left( \dot{\lambda} - Q(f, \lambda) \right) g \left( \dot{\lambda} - Q(f, \lambda) \right) + F = \frac{1}{2} \dot{\lambda}^t g \dot{\lambda} - \dot{\lambda}^t (M + \nabla f) + \frac{1}{2} \phi - \frac{\partial f}{\partial t} + \frac{df}{dt}$$  \hspace{1cm} (23)

Equating terms gives the three equations:

$$Q = g^{-1}(\nabla f + M)$$  \hspace{1cm} (24)

$$\frac{1}{2} Q^t g Q = \frac{1}{2} \phi - \frac{\partial f}{\partial t}$$  \hspace{1cm} (25)

$$F = \frac{df}{dt}$$  \hspace{1cm} (26)
Fig. 2 Top panel (a): The most likely or thermodynamical trajectories $u_m$ are shown for the cases where the start is at either $t = 0$ or at $t = 1.5$. In both cases the start values are given by the original thermodynamical trajectory (see text). Note the initial plateau periods in both cases before exponential decay to equilibrium occurs. Bottom panel (b): The weights at various times for the case $\eta = 1$. Note the (Gaussian) spread at $t = 1.5$ where the restart occurs.

Now the momenta $p$ corresponding to $\lambda$ and the Hamiltonian $H$ are easily computed to be

\[
p = g\dot{\lambda} - M
\]

\[
H(\dot{\lambda}, \lambda) = \frac{1}{2} \dot{\lambda}^T g \dot{\lambda} - \frac{1}{2} \phi
\]  \hspace{1cm} (27)

\[
H(p, \lambda) = \frac{1}{2} (p + M)^T g^{-1} (p + M) - \frac{1}{2} \phi
\]  \hspace{1cm} (28)
Substitution of (24) into (25) and comparison with (28) shows that

\[ \mathcal{H}(\nabla f, \lambda) + \frac{\partial f}{\partial t} = 0 \]  

in other words the Hamilton-Jacobi (HJ) equation for this Lagrangian.

The non-negative action \( S \), which represents the information loss of a particular trajectory, may now be written instructively as proportional to the terms

\[ S \propto \int_0^T \left( \dot{\lambda} - Q(f(\lambda,t),\lambda) \right) g \left( \dot{\lambda} - Q(f(\lambda,t),\lambda) \right) dt + f(\lambda(T), T) - f(\lambda_0, 0) \]

Such an equation holds for many choices for the gauge function \( f \) providing they satisfy the HJ equation. A useful choice for interpreting this reformulation occurs if we specify it at the endpoint:

\[ f(\lambda, T) = 0 \]

The values of \( f \) between \( t = 0 \) and \( t = T \) which gives \( Q \), may be obtained by integrating the HJ equation back in time from the endpoint to the start point. We denote this particular solution by \( f(\lambda, T, t) \). With such a choice and a specified \( \lambda_0 \), the last two terms for the action become independent of \( \lambda(T) \). It is clear now since \( g \) is non-negative definite, that the action at \( t = T \) will be minimized providing that

\[ \dot{\lambda} = Q(f(\lambda, T, t), \lambda) = g^{-1}(\nabla f(\lambda, T, t) + M) \]

\[ \lambda(0) = \lambda_0 \]

which obviously uniquely specifies a trajectory between \( t = 0 \) and \( t = T \). Note that this particular trajectory is the “classical” path discussed in the previous section not the locus of optimum endpoints. The minimum action
(or information loss) is clearly proportional to \(-f(\lambda_0, T, 0)\) i.e. the gauge function at the initial time.

It is clear however that the resulting Lagrangian with this choice of gauge \(f\) is not of the normal OM form since \(Q\) depends on the endpoint \(T\) chosen. As a consequence it is easily seen that unlike the traditional OM case the locus of extremal endpoints is not in itself an extremal. Indeed it is precisely this disconnect between extremal paths and their endpoint locii which was the motivation above for the introduction of a consistency distribution.

Another choice of gauge \(f\) does however reveal the proper connection with OM theory. Suppose instead we fix it as \(f'(\lambda, t = 0) = 0\) then integrate instead forward in time. In this case \(Q\) will not depend on the endpoint time \(T\) however the price paid of course is that the gauge endpoint term remains in the action as a function of both \(\lambda\) and \(T\):

\[
S \propto \int_0^T \left( \dot{\lambda} - Q(f'(\lambda, t), \lambda) \right) g \left( \dot{\lambda} - Q(f'(\lambda, t), \lambda) \right) dt + f'(\lambda(T), T)
\]

which again reflects the difference from an OM form for the Lagrangian.

Consider now however the asymptotic limit as \(t \to \infty\). Solutions of the HJ equation in this limit have been much studied recently in the mathematical literature (see [17] in particular and also [12, 23, 2] and [27]) and we have the result that there is convergence of \(f'\) to a solution \(f_s\) of the stationary HJ equation

\[
\mathcal{H}(\nabla f_s, \lambda) = 0
\]

providing that such stationary solutions actually exist and also that \(\mathcal{H}(p, \lambda)\) is strictly convex in \(p\) for fixed \(\lambda\). The second condition is easily shown to hold using equation (28) when \(g\) (and hence \(g^{-1}\)) is positive definite. Consider now a sufficiently large \(t_a\) such that \(f' \approx g\) then the action between this time and the endpoint \(T\) is clearly approximately given by

\[
S(t_a, T) \propto \int_{t_a}^T \left( \dot{\lambda} - Q(f_s(\lambda), \lambda) \right) g \left( \dot{\lambda} - Q(f_s(\lambda), \lambda) \right) dt
\]

which is of OM form. The propagator of consistency distributions from this large \(t_a\) forward in time is then simply

\[
W(t_a, T) = C \exp (-\Delta t S(t_a, T))
\]

which is now of a (generalized) OM path integral form. Thus the present formulation approaches an OM form for sufficiently large time\(^9\). This behaviour is consistent with the direct numerical simulations described in [19] where relaxation to equilibrium takes some time to establish due to the plateau effect discussed in the introduction and is then rather simple as in OM theory.

\(^9\) providing there is a solution to the stationary HJ equation and that \(g\) is positive definite. Note that this can be an issue as asymptotically HJ solutions can blow up in finite time (i.e. develop shocks) depending on the specific system Hamiltonian.
It is worth observing that the extremal (thermodynamical) path for this asymptotic OM Lagrangian is precisely that discussed in BT as a stationary closure. As observed there it is of the same generic form as proposed by Öttinger (see [25]) which is relevant to a large number of practical thermodynamical systems.

5 Connection to the motion of a charged particle in an external electromagnetic field

The original form of the Lagrangian (18) is familiar from classical mechanics. Indeed if we set $g_{ij} = m \delta_{ij}$ then the Hamiltonian from equation (27) is identical with that of a non-relativistic particle moving in an external fixed electromagnetic field. Here $-M$ and $-\phi$, which generate the reversible and irreversible flows, are proportional to the magnetic vector potential $\mathbf{A}$ and the scalar potential $\phi$ respectively (see [20] p.421).

The more general case for the Fisher metric tensor $g$ is also interesting. Here the Euler-Lagrange equations corresponding to the Lagrangian (18) take after a straightforward calculation the following forced geodesic form

$$\ddot{\lambda}_i + \dot{\lambda}_k \Gamma^k_{il} = g^{li} \left[ \dot{\lambda}_k \left( \frac{\partial M_i}{\partial \lambda_k} - \frac{\partial M_k}{\partial \lambda_i} \right) + \frac{\partial \phi}{\partial \lambda_i} \right]$$

(30)

where $\Gamma$ is the Christoffel symbol corresponding to the Riemannian metric tensor $g$ (and the summation convention is assumed). Such equations are similar in form to the geodesic equations for a particle subject to an external electromagnetic field within a general space-time manifold (see [30] pp.41 and 69) which read

$$\frac{du^a}{d\tau} + u^c u^d \tilde{\Gamma}^a_{cd} = \frac{q}{m} g^{ab} F_{bc} u^c$$

(31)

$$u^a \equiv \frac{dx^a}{d\tau}$$

$$F_{ab} \equiv \nabla_a M_b - \nabla_b M_a = \partial_a M_b - \partial_b M_a$$

Where the tensor indices here are on space-time; the Christoffel symbol $\tilde{\Gamma}$ is appropriate for the usual Lorentzian (as opposed to Riemannian) space-time manifold; $\tau$ is the proper time for the charged particle and finally the electromagnetic potential 4-vector $M_a$ is the combined 3-vector potential and the scalar potential. Note that the electromagnetic field $F$ can be defined from the potential using an arbitrary derivative operator not just the covariant derivative corresponding to the metric since it is the exterior derivative of the potential. If we assume that the particle is moving non-relativistically then we have in a suitable co-ordinates that

$$u^0 = \frac{dx^0}{d\tau} \approx \text{constant} >> u^i \quad i = 1, 2, 3$$

(32)

10 We use $M$ to denote the magnetic vector potential to avoid confusion with the resolved variable set $A$

11 This can include both an electric potential and other potentials such as gravitation
Finally if we assume that the space-time is static then we can choose an appropriate co-ordinate frame\(^{12}\) in which the metric tensor is Riemannian with respect to the spatial co-ordinates; the cross terms \(g_{0i}\) vanish and further
\[
g_{00} = V(x^1, x^2, x^3)
\]
(see [30] p119). With respect to the spatial indices, the left hand side of the geodesic equations are now the same as our Riemannian version [30] with the exception of terms deriving from cross spatial-temporal Christoffel symbols \(\tilde{\Gamma}^i_{00} = \frac{\partial V}{\partial x^i}\). Using the nonrelativistic approximation [32] it is clear this term can be moved to the right hand side and included in the gradient of the scalar potential. Finally if we assume that the external electromagnetic field is static as well as the space-time then equations (30) and (31) are easily seen to be of the same form.

6 Corresponding Schrödinger equations

There is considerable discussion in the literature as to the exact relationship between the Onsager-Machlup path integral discussed above and a corresponding Euclidean Schrödinger equation for the transition probability. The interested reader is also referred to the book [5] where the connection with the issue of quantum operator ordering in Hamiltonians is explained. In general, the relationship depends on the precise nature of the temporal limiting process adopted in defining the path integral. Differing temporal discretisation\(^{13}\) of quantities within the Lagrangian lead to Schrödinger equations with different drift and potential terms. This ambiguity could be seen as somewhat academic since it depends on taking the limit \(\Delta t \to 0\) which violates the spirit of working on a slow timescale (further discussion of such a viewpoint can be found in [21] in the context of general stochastic processes). Nevertheless, a unique fully covariant correspondence has been given in [14] (see also [10]). Graham and co-workers show how this can be achieved concretely by an appropriately chosen discretisation procedure motivated by Wilson’s renormalization group (see [9] and [10]). We follow the Graham formalism below.

We begin for pedagogical reasons with consideration of the simple case at the beginning of the last section with \(g_{ij} = m\delta_{ij}\) namely a charged particle in a flat space with an externally prescribed electromagnetic field. The Feynman path integral of this system is very well known and important ([13] p79) and the wave function \(\psi(\lambda) \equiv K(\lambda, \lambda_{\text{fixed}})\) satisfies a Schrödinger equation discussed at length in standard texts such as [20]. Formally the derivation of this equation from the path integral proceeds identically in our case with the identification
\[
\Delta t \longleftrightarrow -\frac{i}{\hbar}
\]

\(^{12}\) Set by the static space-time Killing vector
\(^{13}\) Or Fourier/phase decompositions.
and so we obtain the Schrödinger equation
\[
\frac{1}{\Delta t} \frac{\partial \psi}{\partial t} = \frac{1}{2m} \left( \frac{1}{\Delta t} \nabla - M \right)^t \left( \frac{1}{\Delta t} \nabla - M \right) \psi - \frac{1}{2} \phi \psi
\]

Noteworthy here is that this formal derivation assumes that in the path integral Lagrangian, the electromagnetic fields are evaluated at the midpoint of the time interval used to define the time derivatives. If other choices are made then a different equation results (see [5]). In the quantum case these alternate equations do not exhibit gauge invariance so are ruled out.

[14] derived the following path integral Lagrangian
\[
L_g(\lambda) = \frac{1}{2} Q^{-1}_{ij} \left( \dot{\lambda}_i - \omega_i \right) \left( \dot{\lambda}_j - \omega_j \right) + \frac{1}{2} \sqrt{Q} \frac{\partial}{\partial \lambda_k} \left( \frac{\omega_k}{\sqrt{Q}} \right) - V + \frac{1}{12} R
\]

\[
\sqrt{|Q|} \equiv \sqrt{\det(Q)}
\]

\[
\omega_i \equiv K_i - \frac{1}{2} \sqrt{|Q|} \frac{\partial}{\partial \lambda_k} \left( \frac{Q_{ik}}{\sqrt{|Q|}} \right)
\]

\[
R = \text{Riemann scalar of } Q^{-1}
\]

from the Schrödinger equation
\[
\left( \frac{1}{2} \frac{\partial}{\partial \lambda_i} \frac{\partial}{\partial \lambda_j} Q_{ij} - \frac{\partial}{\partial \lambda_k} K_k + V \right) \psi = \psi_t \quad (33)
\]

Comparison of this Lagrangian with one derived earlier gives the following identifications
\[
Q = g^{-1}
\]

\[
K_i = g^{-1}_{ij} M_j + \frac{1}{2 \sqrt{|g|}} \frac{\partial}{\partial \lambda_k} \left( \sqrt{|g|} g^{-1}_{ik} \right)
\]

\[
V = \frac{1}{2 \sqrt{|g|}} \frac{\partial}{\partial \lambda_k} \left( \sqrt{|g|} g^{-1}_{ki} M_i \right) + \frac{1}{12} R - \phi + \frac{1}{2} g^{-1}_{ij} M_i M_j
\]

which when substituted in (33) gives the appropriate equation for the current application. In order to make this precise identification of a Schrödinger equation the prescription of [9] for the limit \( \Delta t \to 0 \) must be assumed. To reiterate, these equations really are only approximate asymptotic relations given that on physical grounds \( \Delta t \) must be bounded below by the fast time scale of the dynamical system under consideration.

7 Relationship to quantum statistical mechanics.

The most familiar application of path integrals to statistical mechanics is that which gives the density matrix for an equilibrium ensemble of quantum states i.e. describes a mixed quantum state (see [13] Chapter 10). The path integral then has a Lagrangian which is the Wick rotation of the classical
Lagrangian with imaginary time associated with inverse temperature. For the case of a particle moving in an electromagnetic vector potential and a scalar potential as discussed in section 5 the effect of the Wick rotation is to reverse the sign of the scalar potential and make the vector (magnetic) potential term purely imaginary. Without the magnetic potential the resulting path integral is Wiener and consequently able to be practically evaluated (see e.g. [4] for application to Bose condensates). The present path integral is entirely analogous except the magnetic term in the Lagrangian (which is associated with reversible trajectories) is real rather than imaginary. In addition the scalar potential $\phi$ is exactly the same i.e. there is no reversal of sign as in the classical analog of Section 5. This is important since from equations (11) and (40) it is easily deduced that $\phi \geq 0$ and the potential achieves the lower bound when $\lambda = 0$ i.e. when the trial density is a Gibbs density. Such lower bounded potentials are of course common in many different dynamical contexts. The very simple example discussed above in Section 3 is obviously the density matrix for an ensemble of particles in a harmonic potential. A major practical advantage of the present path integral is clearly that it is always Wiener and thus likely amenable to the numerical methods widely used in quantum statistical mechanics when there is no magnetic potential.

8 Existence of a unique equilibrium consistency distribution

Consider a time transfer operator $K$ of consistency distributions for one timestep $\Delta t$. If we choose to time discretise on the backward timestep then we may write

$$\psi(t + \Delta t, \lambda) = K\phi(t, \lambda) = \int Q(\lambda, \kappa)\psi(t, \kappa) d\kappa$$

where all the functions in the exponent are of $\kappa$ the backward variable rather than $\lambda$. We also choose the function $N$ to preserve the consistency distribution $L_1$ norm so

$$N(\kappa) = [2\pi]^{-m/2} \sqrt{|g|}$$

where $m$ is the coarse grained dimension. This choice for normalization also has the attractive property that it “preserves volumes” in $\kappa$ space i.e. $\sqrt{|g|}d\kappa$ is the natural volume element for the metric tensor $g$ (see e.g. Appendix B [30]).

The operator $K$ as chosen above turns out to be compact with the addition of some sufficiency conditions. An operator is compact if the image of any bounded set is totally bounded (see [4]). The Kolmogorov-Riesz theorem on totally bounded sets of $L_1$ spaces asserts (see [16]) that as well as the
image of $K$ being bounded we also require that for every $\epsilon > 0$ there exists a $R(\epsilon)$ such that for all $K\psi$

$$\int_{|\lambda|>R(\epsilon)} |K\psi| \, d\lambda < \epsilon$$

(35)

and secondly that for every $\epsilon > 0$ there exist some $\rho > 0$ such that for all $K\psi$ and $\gamma$ with $|\gamma| < \rho$

$$\int d\lambda |K\psi(\lambda + \gamma) - K\psi(\lambda)| < \epsilon$$

(36)

We have the following theorem the proof of which is rather technical and may be found in Appendix B:

**Theorem:** If the transfer operator $K$ defined by (34) satisfies the conditions

1. $IL_{rev}(\kappa) \to \infty$ as $|\kappa| \to \infty$.
2. In any bounded region $|\kappa| \leq M$ $g$ is bounded below and $g^{-1}M$ is bounded above by the usual $R^n$ norm.

then it is compact.

Condition 1. here is the most significant. Such a property holds for the practical cases examined to date by the author. It corresponds with the quantum case of an infinite confining potential which is widely relevant. Note that this potential is not the scalar potential $\phi$ rather it is $IL_{rev}$ given by equation (14). As was observed in section 2 the absence of this term means the reversible trajectory is optimal and the Lagrangian reduces to Onsager-Machlup form. Thus in some sense this term is fundamentally responsible for irreversibility.

Consider now the cone $C$ of non-negative functions belonging to the $L_1$ Banach space of real functions. Suppose $\psi \in C$ and that

$$\int Q(\lambda_0, \kappa)\psi(\kappa) \, d\kappa = 0$$

for fixed $\lambda_0$. It follows that $\|Q(\lambda_0)\psi\|_1 = 0$ which implies (see 22 Chapter 2) that $Q(\lambda_0, \kappa)\psi(\kappa)$ considered as a function of $\kappa$ vanishes almost everywhere in the Lebesgue measure. But since $Q$ is strictly positive everywhere this must imply that $\psi$ also vanishes almost everywhere in the Lebesgue measure i.e. it is part of the zero equivalence class of $L_1$ functions. Thus the only functions belonging to $C$ mapped by the operator $K$ to the boundary of the cone are those that are zero in the sense of the $L_1$ space. Re-expressed: The compact operator $K$ is strongly positive in that all members of $C$ apart from the zero function class are mapped by $K$ into its interior.

Thus all conditions for the Krein-Rutman Theorem (a generalization of the better known Perron-Frobenius theorem to Banach spaces) are met (see Theorem 1.2 [11]) which implies that $K$ has a unique eigenvector belonging to $C$ with a positive eigenvalue. Any other eigenvalue cannot be positive and must have an eigenvector outside $C$. This unique eigenvector can clearly be identified with a unique equilibrium consistency distribution.

14 Up to a scalar multiple and the addition of a function vanishing almost everywhere with respect to the Lebesgue measure.
9 Discussion and future work

In the present work we have argued that macrostates at a particular time are best described by a consistency distribution which is analogous to a quantum wave function. The transfer operator which controls the time evolution of such distributions can be written as a path integral. From a practical viewpoint at any fixed time one is usually interested in an estimate of the moments of the resolved or slow variables of the system. As discussed above one approach to this issue is to identify the maximum value of the consistency distribution and then use the moments of the corresponding quasi-equilibrium density as the required estimate. A possible difficulty with this may occur if the mentioned maximum is not unique. Another approach may be to weight the moments of all such densities with the (normalized) consistency distribution. Which particular approach is a better one awaits a more extensive investigation of the consistency distributions and direct numerical simulations of several relevant dynamical systems.

We saw in Section 5 that for asymptotically large times the path integral approaches one of Onsager-Machlup form. The unique optimizing path in that case was one previously discussed in BT as a stationary formulation and is one widely useful in many practical thermodynamical contexts as documented by Öttinger (see [25]). BT also discussed a non-stationary formulation for thermodynamical paths which involved a backward time integration of Lagrangians like those discussed above. The relationship between this and the maximum of the consistency distributions introduced here remains to be completely clarified. A straightforward computation by the author reveals that they coincide for simple Lagrangians of the type discussed in section 3.1. A general demonstration however awaits further work.

A key practical advantage of the present formulation lies in the fact that extensively tested numerical methods from quantum statistical mechanics exist for the efficient evaluation of the proposed path integrals. An excellent review of this field from the viewpoint of quantum chemistry may be found, for example, in the article by Ceperley [4].

Another issue requiring further investigation concerns the choice of resolved variables. These are functions of the slow variables of the original dynamical system but the key question is which ones to select. Intuitively one expects the marginal distributions for random variables averaged over the time interval $\Delta t$ to be general functions of the slow variables of the original system. Practical experience however shows that only rather simple such functions are apparent when direct numerical simulations are examined. Thus, for example, the author has examined the TBH system discussed in section 1 and discovered that to a very good approximation the square of slow variables suffices in addition to linear functions. Clearly however an important topic to examine is the convergence of results from the present formalism as higher order slow variable functions are included among the resolved variables.

The slow variable averaging interval $\Delta t$ used in the proposed generalized Boltzmann principle above also deserves further investigation using the methods just discussed. It will be interesting to document the sensitivity of
moment estimates to variations in this parameter. It seems reasonable to infer that such a time interval could be determined independently to be the temporal decorrelation interval of unresolved variables within the system.

Finally the equilibrium behaviour of the consistency distribution deserves close attention. We have seen in the last section that there exists a unique distribution left invariant by a reasonable choice of time transfer operator. The convergence toward such an invariant distribution however evidently requires more attention. In the case that a system is modelled using a stochastic process such convergence is well understood and controlled by the monotonic decline of the relative entropy of the transient and equilibrium densities. It would be interesting to locate a similar principle for the present formalism. Furthermore the symmetries and maxima of the invariant consistency distribution are evidently of interest.

Appendix A: Some useful relations

Define the expectation bracket

$$\langle F \rangle \equiv \int F \hat{p}$$

for a general function of the state variables and time $F$. We have now

$$\frac{\partial \langle F \rangle}{\partial t} - \langle L(F) \rangle = \left\langle \frac{\partial F}{\partial t} \right\rangle + \int (F \hat{p}_t - L(F) \hat{p})$$

$$= \left\langle \frac{\partial F}{\partial t} \right\rangle + \int (F \hat{\lambda}_t + L \hat{p})$$

$$= \left\langle \frac{\partial F}{\partial t} + FR \hat{p} \right\rangle$$

$$= \left\langle \frac{\partial F}{\partial t} + FR \hat{\lambda} \right\rangle$$

(37)

where we are using the anti-Hermitian nature of $L$ on the second line. Setting $F = 1$ we obtain immediately that

$$\langle R \rangle = 0$$

(38)

For an exponential family $\hat{p}$ it follows from the definition (7) that

Now it is easily derived from the definition of $R$ and the form of the exponential family of distributions that

$$R = \dot{\lambda} U + \lambda L A$$

(39)

which when combined with (38) yields

$$\lambda \langle LA \rangle = 0$$

(40)
The anti-Hermitian nature of $L$ also allows us to deduce the following two useful relations (using the summation convention and vector/matrix indices for clarity):

$$M_i = \langle L(A_i) \rangle = \int L(A_i) \exp (\lambda_j A_j - G) \exp (-\beta E)$$

$$= \int L(A_i \exp (\lambda_j A_j - G)) \exp (-\beta E) - \int A_i L(\exp (\lambda_j A_j - G)) \exp (-\beta E)$$

$$= - \int A_i \exp (\lambda_j A_j - G) L(\exp (-\beta E)) - \lambda_j \int A_i L(A_j) \hat{p}$$

$$= - \lambda_j \langle A_i L(A_j) \rangle = - \lambda_j \langle (A_i - a_i)L(A_j) \rangle \equiv - h_{ij} \lambda_j$$

(41)

where we are using the fact that $L$ annihilates $E$ and (40) for the second last step. Combining (40) and (41) we obtain

$$\lambda^* h \lambda = 0$$

In a completely analogous way to (41) we deduce that

$$\langle L^2(A_i) \rangle = - \lambda_j \langle LA_j LA_i \rangle \equiv - k_{ij} \lambda_j.$$ 

and more generally

$$\langle L^n A_j \rangle = - \lambda_i \langle L A_i L^{n-1} A_j \rangle$$

(42)

It is easily shown also that

$$\frac{\partial M_i}{\partial \lambda_j} = \langle L(A_i) (A_j - a_j) \rangle = h_{ji}$$

Appendix B: Section 8 Theorem proof.

We first establish that the operator $K$ is bounded with respect to the $L_1$ norm. Consider the effect of $K$ on a distribution $\phi$ with $L_1$ norm unity:

$$OP \equiv \|K\phi\|_1 = \int \left| \int Q \psi dk \right| d\lambda$$

$$\leq \int \left| \int |Q\psi| dk \right| d\lambda$$

$$= \int N \exp [-IL_{irr} - IL_{rev}] |\psi| dkd\lambda$$

$$\leq \int N \exp [-IL_{irr}] |\psi| dkd\lambda$$

$$= \int |\psi| dk = 1$$

where on line 4 we have used the fundamental fact derived in section 2 that $IL_{rev} \geq 0$ while the last line follows after switching variables of integration and using the normalization condition which also holds for $\exp [-IL_{irr}]$. 
We further establish that the image of $K$ is totally bounded which means establishing the additional two properties (35) and (36).

From condition 1. of the Theorem we deduce that there exists an $|\kappa_0|$ such that
\[
\exp(-IL_{rev}) < \frac{\epsilon}{2} \quad \text{if} \quad |\kappa| > |\kappa_0|
\]

Consider now the bounded region $|\kappa| \leq |\kappa_0|$. From condition 2. of the Theorem; the region boundedness and the fact that $\exp(-IL_{irr})$ is Gaussian in $\lambda$, if follows that there exists an $R(\epsilon) > 0$ such that for all $|\kappa| \leq |\kappa_0|
\[
\int_{|\lambda| > R(\epsilon)} N \exp(-IL_{irr}) \, d\lambda < \frac{\epsilon}{2}
\]

Thus
\[
\int_{|\lambda| > R(\epsilon)} |K\psi| \, d\lambda = \int_{|\lambda| > R(\epsilon)} d\lambda \int d\kappa N \exp(-IL_{irr} - IL_{rev}) \psi
\leq \int_{|\lambda| > R(\epsilon)} d\lambda \int d\kappa N \exp(-IL_{irr} - IL_{rev}) |\psi|
\leq \int_{|\lambda| > R(\epsilon)} d\lambda \int_{|\kappa| \leq |\kappa_0|} d\kappa N \exp(-IL_{irr}) |\psi|
\quad + \int_{|\kappa| > |\kappa_0|} d\kappa N \exp(-IL_{irr}) |\psi|
\leq \frac{\epsilon}{2} + \frac{\epsilon}{2}
\]
which establishes (35).

To establish the other required property consider an arbitrary $\rho_t > 0$ and all $|\gamma| < \rho_t$. The triangle inequality plus (35) implies that
\[
\int_{|\lambda| > R(\epsilon) + \rho_t} |K\psi(\lambda + \gamma) - K\psi(\lambda)| \, d\lambda \leq \int_{|\lambda| > R(\epsilon) + \rho_t} |K\psi(\lambda + \gamma)| \, d\lambda + \int_{|\lambda| > R(\epsilon) + \rho_t} |K\psi(\lambda)| \, d\lambda
\leq \frac{\epsilon}{4} + \frac{\epsilon}{4}
\]
(43)

Set $S(\epsilon, \rho) = R(\epsilon) + \rho$ and $V_{\epsilon\rho}$ the volume of the region $Z : |\lambda| \leq S(\epsilon, \rho_t)$. Let $\kappa_0$ be such that $|\kappa| > |\kappa_0|$ implies that
\[
\exp(-IL_{rev}) < \frac{\epsilon}{SV_{\epsilon\rho}}
\]
(44)

We have
\[
\int_{|\lambda| \leq S(\epsilon, \rho U)} |K\psi(\lambda + \gamma) - K\psi(\lambda)| \, d\lambda \\
= \int_{|\lambda| \leq S(\epsilon, \rho U)} \int_{|\kappa| > |\kappa_0|} \exp(-IL_{red}) N |\exp(-IL_{irr}(\lambda + \gamma)) - \exp(-IL_{irr}(\lambda))| |\psi| \, d\kappa d\lambda \\
+ \int_{|\lambda| \leq S(\epsilon, \rho U)} |f_{\kappa_0}(\lambda + \gamma) - f_{\kappa_0}(\lambda)| \, d\lambda
\]

(45)

with

\[
f_{\kappa_0}(\lambda) \equiv \int_{|\kappa| \leq |\kappa_0|} Q(\lambda, \kappa) \psi(\lambda) d\kappa
\]

an integral transform defined on a bounded domain. The first integral on the RHS of (45) is easily shown using the triangle inequality; the inequality (44) and the non-negativity of the \(IL\) terms to be less than \(\frac{\epsilon}{2}\).

The function \(f_{\kappa_0}\) can be shown by standard arguments to be continuous since the integral transform is defined on a bounded domain and the function \(Q\) is continuous with respect to the first argument. By the Heine-Cantor theorem it is therefore uniformly continuous on the bounded region \(Z\). It follows that there exists a \(\rho_U\) such that for all \(\gamma : |\gamma| \leq \rho_U\) and all \(\lambda \in Z\)

\[
|f_{\kappa_0}(\lambda + \gamma) - f_{\kappa_0}(\lambda)| < \frac{\epsilon}{4V_{\rho U}}
\]

(46)

and so for such \(\gamma\) the second integral from (45) is also less than \(\frac{\epsilon}{2}\). Compare now \(\rho_U\) and \(\rho_U\). If \(\rho_U \geq \rho_U\) then we can replace \(\rho_U\) with \(\rho_U\) in the last integral inequality discussed and obtain the required inequality (46) by combining the three inequalities derived from (43) and (45). Conversely if \(\rho_U < \rho_U\) then \(V_{\rho U} < V_{\rho_U}\). Thus inequalities (44) and (46) still hold if we use \(\rho_U\) in place of \(\rho_U\). Furthermore the newly defined bounded \(Z\) is a subset of the old \(Z\) whence the uniform continuity just discussed holds with the same \(\rho_U\) and hence we are done.

Acknowledgements
Comprehensive discussions with Bruce Turkington on matters related to the present contribution are very gratefully acknowledged. Useful discussions on related matters over many years with Andy Majda are also acknowledged.

References
1. Amari, S., Nagaoka, H.: Methods of Information Geometry. Translations of Mathematical Monographs, AMS, Oxford University Press (2000)
2. Barles, G., Souganidis, P.E.: On the large time behavior of solutions of Hamilton–Jacobi equations. SIAM J Math Anal 31(4), 925–939 (2000)
3. Battetzzati, M.: Onsager principle for nonlinear mechanical systems modeled by stochastic dissipative equations. Arch. Mech. 64(2), 177–206 (2012)
4. Ceperley, D.M.: Path integrals in the theory of condensed helium. Rev Mod Phys 67(2), 279 (1995)
5. Chaichian, M., Demichev, A.: Path integrals in physics. Vol. 1: Stochastic processes and quantum mechanics. IOP, London (2001)
6. Conway, J.B.: A course in functional analysis. Springer (1990)
7. Cover, T., Thomas, J.: Elements of information theory, 2nd edn. Wiley-Interscience, New York (2006)
8. Darve, E., Solomon, J., Kia, A.: Computing generalized Langevin equations and generalized Fokker-Planck equations. Proc. Nat. Acad. Sci. 106(27), 10,884–10,889 (2009)
9. Deiningerhaus, U., Graham, R.: Nonlinear point transformations and covariant interpretation of path integrals. Z. Phys. B Con. Mat. 34(2), 211–219 (1979)
10. Dekker, H.: On the path integral for diffusion in curved spaces. Physica A 103(3), 586–596 (1980)
11. Du, Y.: Order structure and topological methods in nonlinear partial differential equations: Vol. 1: Maximum principles and applications. World Scientific Publishing Company (2006)
12. Fathi, A.: Sur la convergence du semi-groupe de Lax-Oleinik. CR Acad Sci I-Math 327(3), 267–270 (1998)
13. Feynman, R.P., Hibbs, A.R., Styer, D.F.: Quantum mechanics and path integrals. McGraw-Hill New York (1965)
14. Graham, R.: Path integral formulation of general diffusion processes. Z. Phys. B Con. Mat. 26(3), 281–290 (1977)
15. Haken, H.: Generalized Onsager-Machlup function and classes of path integral solutions of the Fokker-Planck equation and the Master equation. Z. Phys. B Con. Mat. 24(3), 321–326 (1976)
16. Hanche-Olsen, H., Holden, H.: The Kolmogorov-Riesz Compactness Theorem. Expositiones Mathematicae 28, 385–394 (2010)
17. Ishii, H.: Asymptotic solutions for large time of Hamilton-Jacobi equations in Euclidean n space. In: Ann I H Poincare-An, vol. 25, pp. 231–266. Elsevier (2008)
18. Jaynes, E.T.: Information theory and statistical mechanics. Phys. Rev. 106, 620 (1957)
19. Kleeman, R., Turkington, B.E.: A nonequilibrium statistical model of spectrally truncated Burgers-Hopf dynamics. Comm. Pure Appl. Math. (2013). In press
20. Landau, L.D., Lifshitz, E.M.: Quantum mechanics non-relativistic theory. Pergamon, London (1965)
21. Lavenda, B.H.: On the validity of the Onsager-Machlup postulate for nonlinear stochastic processes. Found. Phys. 9(5-6), 405–420 (1979)
22. Lieb, E.H., Loss, M.: Analysis. American Mathematical Society, Providence R.I. (2001)
23. Namah, G., Roquejoffre, J.M.: Remarks on the long time behaviour of the solutions of Hamilton-Jacobi equations. Commun Part Diff Eq 24(5-6), 883–893 (1999)
24. Onsager, L., Machlup, S.: Fluctuations and irreversible processes. Phys. Rev. 91(6), 1505–1512 (1953)
25. Òttinger, H.C.: Beyond equilibrium thermodynamics. Wiley-Interscience (2005)
26. Roncadelli, M.: New path integral representation of the quantum mechanical propagator. J. Phys. A-Math. Gen. 25(16), L997 (1992)
27. Roquejoffre, J.M.: Convergence to steady states or periodic solutions in a class of Hamilton–Jacobi equations. J Math Pure Appl 80(1), 85–104 (2001)
28. Taniguchi, T., Cohen, E.G.D.: Nonequilibrium theory for nonequilibrium steady states and fluctuation theorems. J. Stat. Phys. 126(1), 1–41 (2007)
29. Turkington, B.: An optimization principle for deriving nonequilibrium statistical models of hamiltonian dynamics. J. Stat. Phys 152, 569–597 (2013)
30. Wald, R.M.: General Relativity. University of Chicago Press, Chicago and London (1984)
31. Zubarev, D.N.: Nonequilibrium Statistical Thermodynamics. Plenum Press, New York (1974)
32. Zwanzig, R.: Nonequilibrium statistical mechanics. Oxford University Press, USA (2001)
