Information Theory, Predictability and Disequilibrium.

Lecture 11: An information theoretic approach to statistical disequilibrium

1 Introduction

In the last two lectures we have seen that the relative entropy between a prediction density and an equilibrium density is a natural measure of the predictability of a statistical system. Furthermore in Lecture 5 we have seen that in a continuous Markov random process governed by the Fokker Planck equation (FPE), the relative entropy between an arbitrary density and the equilibrium density declines monotonically and can therefore be used as a measure of the disequilibrium of such systems. This close relationship between disequilibrium, relative entropy and predictability suggests that non-equilibrium statistical physics is a subject worthy of very close attention. Unfortunately it is also a subject which is far from being in a final state as a complete physical theory. Many different approaches have been attempted with varying degrees of elegance and success. This situation is in stark contrast to the power, elegance and completeness of equilibrium statistical mechanics. A unifying principle in the latter area derives from information theory due to the maximum entropy principle identified by Jaynes. This suggests that a similar phisiophical approach may be productive for the non-equilibrium case. In this lecture we present a recent such effort by the instructor and co-workers.

2 Liouville equation and Gibbs densities

As we saw in Lecture 5, in a deterministic dynamical system obeying the autonomous equations

$$\frac{\partial x_i}{\partial t} = A_i(\mathbf{x}) \tag{1}$$

the probability density evolves according to the PDE

$$\partial_t p = -\partial_i \left[A_i(\mathbf{x}) p \right]$$

where the summation convention is assumed. If the system also satisfies the so-called Liouville property

$$\partial_i A_i = 0$$

then this becomes

$$\partial_t p = -A_i \partial_i \left[p \right]$$

Important examples of such systems include those referred to as Hamiltonian. Molecular dynamics is often assumed to be Hamiltonian. In this lecture we confine ourselves to this case for pedagogical reasons although the formalism to be described may be generalized to an autonomous dynamical system. It is convenient in such systems to introduce a so-called Poisson bracket

$$\{C, B\} \equiv \partial_i C J_{ij} \partial_j B$$

where the matrix J is antisymmetric which implies that the bracket is as well. The dynamical system is assumed to be given by

$$\frac{\partial x_i}{\partial t} = \{x_i, H\}$$

where $H(x)$ is called the Hamiltonian and characterises the dynamical system. Any quantity within such a system which has a vanishing Poisson bracket with H is easily shown to be time invariant. Energy is an example in molecular dynamics. It is also easily verified that such a system satisfies the Liouville property and furthermore that the density evolves according to

$$\begin{aligned} \frac{\partial p}{\partial t} + Lp &= 0 \\ Lg &\equiv \{g, H\} \end{aligned} \tag{2}$$

Now in the case that the density p is steady i.e. the system is in statistical equilibrium then we deduce immediately that this p_{eq} is an invariant of the dynamical system. Numerical and experimental experience shows that it has the form

$$p_{eq} = C \exp(-\beta_i I_i)$$

where I_i are low order algebraic invariants for the system. In molecular dynamics energy is one of the I_i and the corresponding β_i is called the inverse temperature. A density of this form is called a Gibbs density. The form above may also be derived by constraining the average of the I_i to fixed values and seeking the maximum entropy density as in Lecture 6. This principle due to Jaynes is a reasonable one for a system in equilibrium since one would expect the means of invariants to remain fixed when a statistical system is in equilibrium. As we saw in Lecture 5 the fixed mean values of these invariants are related to the inverse temperatures β_i by the so-called Legendre transformation.

3 Non-equilibrium densities

One might be tempted given the discussion of the previous section to simply integrate the Liouville equation from an arbitrary initial density. Indeed this is precisely what often occurs with the FPE discussed earlier. Unfortunately in many systems of practical interest (most fluids and molecular dynamical systems) this is impractical since they involve a large number of degrees of freedom which translates into a high dimensional PDE. This is commonly referred to as the curse of dimensionality. A popular approach to this difficulty is coarse-graining or closure. Here we identify the slow degrees of freedom of the system and attempt by various devices to treat the fast degrees of freedom as being in (statistical) equilibrium and forcing the slow modes in some way. This is actually one of the initial motivations for the study of FPEs with a diffusion term. Often the way in which the fast modes influence the slow modes is derived in an unpleasantly ad hoc manner although rigorous limiting approaches do exist.

The approach to be described here assumes a particular density family for the slow modes and also assumes the remaining fast modes satisfy a Gibbs density form as they are close to equilibrium. The slow mode density family must be selected from experimental investigation. It also must be sufficiently simple to allow analytical moment calculation. Fortunately in many important practical instances a Gaussian form is (approximately) justifiable which ensures tractable calculations. It is to be emphasized though that this (and the fast mode Gibbs assumption) are always approximations. This is in contrast to the equilibrium case where Gibbs densities can be shown to be extremely accurate. Presumably as time scale seperation between fast and

slow modes becomes large and densities closely approach Gaussian form then these approximatins become asymptotically exact. Such a result has yet to be rigorously established however.

The density families to be used are exactly those discussed in Lecture 6 and as we saw there may be regarded as differential manifolds with a Riemannian metric tensor given by the Fisher information matrix and coordinates given by the parameters of the family under consideration. In the case of a Gaussian family the parameters are the means and covariances. Note also that under certain circumstances, the equilibrium Gibbs densities may be included within the manifold. This perspective on the slow variable densities will play an important role below.

4 The Liouville equation constraint

4.1 Introduction

Consider a "path" through the slow manifold i.e. a time series of densities constructed in the manner outlined above. There is clearly no guarantee that it will satisfy the Liouville equation (2). Indeed some careful consideration leads one to the general conclusion that the path will often not since the fast modes are assumed to be equilibriated and this cannot be so in general for a short enough time interval. Note however that the static point consisting of the Gibbs density will satisfy the equation given the discussion in section 2.

Despite this lack of precise "Liouvillean evolution" one might hope that a path could be found which in some sense minimizes the deviation from such evolution. This is the approach we shall follow but how do we measure deviation from Liouvillean evolution? Intuitively this measure is one between a path density and a Liouvillean evolved density. As we have seen repeatedly in this course the obvious candidate for such a "density distance" is provided by the relative entropy.

4.2 Mathematical formulation

We refer to the approximating densities as trial densities and write them as $\hat{p}$. Consider various temporal evolutions over a short interval Δt which is however assumed sufficiently long that fast modes decorrelate. The evolution

according to the Liouville equation (2) will be

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

Now in general[1] this evolved density will lie outside the manifold described by trial densities. The evolved trial density must therefore be the different density

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

where the operator $T = \frac{\partial}{\partial t}$ is anti-Hermitian like the Liouville operator L. The information lost IL in assuming $\hat{p}(t+\Delta t)$ when in fact the density is $\overline{p}(t+\Delta t)$ is simply the relative entropy $D(*||*)$ of the second density with respect to the first. We have now the following

$$\begin{aligned}
IL &= D\left(e^{-\Delta t L}\hat{p}||e^{\Delta t T}\hat{p}\right) \\
&= \int e^{-\Delta t L}\hat{p}\left(e^{-\Delta t L}\hat{l} - e^{\Delta t T}\hat{l}\right) \\
&= \left\langle e^{\Delta t L}\left(e^{-\Delta t L} - e^{\Delta t T}\right)\hat{l}\right\rangle_{\hat{p}} \\
&= \left\langle \left(I - e^{\Delta t L}e^{\Delta t T}\right)\hat{l}\right\rangle_{\hat{p}} \qquad (3) \\
&= \left\langle \left(I - e^{\Delta t (T+L)}\right)\hat{l}\right\rangle \qquad (4)
\end{aligned}$$

with $\hat{l} \equiv \log\hat{p}$. On the second line we are using the fact that an arbitrary function of p also obeys the Liouville equation (2); on the third line we are using the anti-Hermitean property for L; and on the last line we are assuming that $[L,T] = 0$ and the expectation refers to the trial density at the start of the propagation interval. Define now the following useful random variable R which we call the Liouville residual

$$R(p) \equiv (T+L)\log p \qquad (5)$$

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, using the

[1] If the trial distribution gives an invariant measure for the system this will not be the case.

anti-hermitian property of L and the definition of R, to satisfy the following evolution equation (see Appendix A [2])

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

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

$$\langle R \rangle = 0 \tag{6}$$

and secondly (setting $F = R$) that

$$\langle (T + L) R \rangle = - \left\langle R^2 \right\rangle \tag{7}$$

Returning now to equation (4) we expand the exponential operator as a Taylor series. The terms in Δt of order zero and one vanish due to cancellation and equation (6) while the order two term remains and using (7) we derive the remarkably simply second order approximation

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

Thus the information loss per timestep 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 Δt which is consistent with the relative entropy geometrically being a distance squared as discussed in Lecture 6. In order to make further progress beyond this general equation we now specify the trial density manifold $\mathcal{T}$. 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 as discussed in the previous section

$$\hat{p}(t) = \exp \left[\lambda(t)^t A - G(\beta, \lambda) - \beta E \right] \tag{8}$$

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 λ and the expectation values 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

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

where the overdot denotes a time derivative and hence that

$$\begin{aligned} IL &= \frac{(\Delta t)^2}{2} \left(\dot{\lambda}^t g \dot{\lambda} - 2\dot{\lambda}^t \langle LA \rangle + \phi \right) + O((\Delta t)^3) \\ \phi &\equiv \lambda_i \langle LA_i LA_j \rangle \lambda_j \\ g_{ij} &\equiv \langle (A_i - a_i)(A_j - a_j) \rangle \end{aligned} \tag{9}$$

The matrix/tensor g here is the Fisher information matrix which as already noted is a Riemannian metric tensor for the manifold of trial densities. We have also used the following identity derived in Appendix A of [2]:

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

5 Optimal paths

Viewing equation (9) we see it has the form of a Lagrangian in classical mechanics. Indeed if only the quadratic term in $\dot{\lambda}$ is retained then this Lagrangian is that appropriate for calculating geodesics on the trial density manifold from the metric tensor. The other two terms may be seen in a Lagrangian for electromagnetism with the vector $\langle LA \rangle$ playing the role of a vector magnetic potential while ϕ plays the role of an electric scalar potential. As is well known, interesting dynamical paths are obtained by integrating the Lagrangian with time to produce an action:

$$S = \frac{\Delta t}{2} \sum \left(\dot{\lambda}^t g \dot{\lambda} - 2\dot{\lambda}^t \langle LA \rangle + \phi \right) \Delta t \simeq \frac{\Delta t}{2} \int_{t_1}^{t_2} \left(\dot{\lambda}^t g \dot{\lambda} - 2\dot{\lambda}^t \langle LA \rangle + \phi \right) dt$$

Fixing two endpoints for a path and minimizing this action produces a classical dynamical path. In the case that the potentials are ignored this is precisely the geodesic for the manifold. This manifold is the general one for the choice of trial density family and the metric tensor does not depend

on the dynamics of the problem under consideration. The introduction of the vector and scalar potentials which both involve the Liouville operator L introduce the dynamics for slow variables.

It appears therefore that we have a variational framework analogous to classical dynamics for non-equilibrium statistical systems. The treatment of endpoints however is a little different to classical dynamics (see below). Note further that all terms within the Lagrangian here are specified and arise purely from the interaction between fast and slow modes within the dynamical system. The only ad hoc aspect to our approach concerns the selection of the trial manifold and the specification of slow variables.

How should the action S, which measures deviation from a Liouvillean evolution, be used to determine the most consistent path? Such a path could be called the thermodynamical trajectory. In a practical situation one could ensure that the initial density belonged to the trial manifold. Thus the initial value of $\lambda(0)$ would be set and the endpoint values $\lambda(t)$ then determined by a variational principle. There are several choices as to which principle should be used:

1. One could find the semi-infinite path $\hat{\lambda}(t)$ which minimizes

$$S_\infty\left(\lambda(0)\right) \equiv \int_0^\infty \mathscr{L}(\lambda, \dot{\lambda}) dt \tag{10}$$

   subject to the constraint that $\lambda(0)$ is prescribed.

2. One could find the path minimizing

$$S\left[\lambda(t)\right] \equiv \int_0^t \mathscr{L}(\lambda, \dot{\lambda}) dt$$

   again with the same constraint. Denote the action of this extreme path by $s(\lambda)$ one could then seek the endpoint $\hat{\lambda}(t)$ which minimizes $s(\lambda)$.

3. Rather than considering only the extreme path action in evaluating $s(\lambda)$ one could sum over all possible paths leading to the endpoint from the fixed initial point and weight them with an appropriate factor such as

$$\exp\left(-\frac{\Delta t}{2} S\left[\lambda\right]\right)$$

   This then forms a path integral of a standard (Wiener) type which can be evaluated by many techniques popular in mathematical physics. The

generalized function $s(\lambda)$ can then be minimized as before to obtain $\hat{\lambda}(t)$.

Note that method 2 can be obtained from method 3 by allowing Δt to become arbitrarily large. All three methods have been used by the author and co-worker Turkington (see [2], [4] and [3]). In general the second two methods give better results when compared to direct numerical simulation of a variety of turbulent systems. The first method gives a good account of the asymptotic behaviour of a system as it approaches equilibrium while the other methods also enable a better simulation of finite time (spin up) effects. We briefly sketch the solution to method 1.

Associated with the Lagrangian is the Hamiltonian[2] obtained in the usual manner of Hamiltonian mechanics:

$$\begin{aligned} \mathscr{H}(p,\lambda) &= (p+M)^t\, g^{-1}\,(p+M) - \phi \\ p &= g\dot{\lambda} - M \\ M &\equiv \langle LA \rangle \end{aligned}$$

where p is the conjugate momentum to λ. A fundamental theorem of Hamiltonian mechanics says that the action from equation (10) satisfies the stationary Hamilton Jacobi equation

$$\mathscr{H}\left(\nabla S_\infty, \lambda\right) = 0$$

The required thermodynamical path can now be obtained from the equation

$$\begin{aligned} \frac{d\hat{\lambda}}{dt} &= g^{-1}(\nabla S_\infty + M) \\ \hat{\lambda}(0) &= \lambda(0) \end{aligned}$$

Such thermodynamical relaxation equations are widely seen in the non-equilibrium literature. Solutions to Method 2 above can be found using similar Hamilton-Jacobi machinery.

[2]Note that this is the slow variable Hamiltonian not the Hamiltonian applicable to the entire dynamical system.

References

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

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

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

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