Statistical-dynamical models that minimize the rate of information loss

Bruce Turkington
University of Massachusetts Amherst

Simon Thalabard
(former UMass postdoc; now IMPA, Brazil)

Alan Turing Institute, May 2018

Statistical closure via dynamical optimization

- Coarse-grain the full deterministic dynamics using a parametric statistical model on phase space – that is, nonequilibrium ensembles associated with resolved variables
- Quantify the lack-of-fit of paths in the model parameter space using the incremental Kullback-Leibler divergence between exact and model p.d.f.s
- Minimize the lack-of-fit over paths to derive the reduced equations governing the mean resolved variables

  In theory, Hamilton-Jacobi theory supplies the closed equations.

  In practice, optimal control algorithms find the best-fit path.

B.T., "An optimization principle for deriving nonequilibrium statistical models of Hamiltonian dynamics." *J. Stat. Phys.* (2013).

R. Kleeman, "A path integral formalism for nonequilibrium Hamiltonian statistical systems." *J. Stat. Phys.* (2014).

General framework

Given deterministic dynamics governs the microstate z that evolves in a high-dimensional phase space Γ.

$$\frac{dz}{dt} = F(z) \qquad \text{where} \qquad \text{div } F = 0\,, \quad F \cdot \text{grad } H = 0\,.$$

Conserved quantities for $F(z)$:

- Phase volume dz on Γ.
- Energy function H, plus any other conserved functions.

Canonical example: $z = (q_1, p_1, \ldots, q_n, p_n) \in R^{2n}$:

$$\frac{dq_j}{dt} = \frac{\partial H}{\partial p_j}\,, \quad \frac{dp_j}{dt} = -\frac{\partial H}{\partial q_j}\,.$$

Propagation of an *ensemble* of microscopic trajectories is described exactly by p.d.f. $\rho(z,t)$ on Γ, which solves the *Liouville equation:*

$$\left(\frac{\partial}{\partial t} + L \right) \ln \rho(z,t) = 0\,, \qquad \text{where} \qquad L \doteq F \cdot \text{ grad}$$

Statistical-dynamical model

Replace $\rho(z,t)$ by a *nonequilibrium statistical model:*

$$\rho_{neq}(z;\xi(t)) \text{ parameterized by } \xi = (\xi^1, \ldots, \xi^m), \quad m \ll n\,.$$

Canonical choice:

$$\rho_{neq}(z;\xi) \;=\; \exp\Big(\xi^k A_k(z) \;-\; \beta H(z) \;-\; \phi(\xi,\beta)\Big)$$

- Vector of resolved variables, $A = (A_1, \ldots, A_m)\,, \;\; m \ll n \;\;.$
- Inverse temperature $\beta = \beta(\xi)$ determined by the energy constraint $\langle H \rangle_{neq} = E$
- Potential $\phi(\xi,\beta)$ normalizes p.d.f.
- Equilibrium state included: $\rho_{eq}(z) = \rho_{neq}(z;0)$.

Justifications for canonical trial densities

Jaynes' "MaxEnt" : $\rho = \rho_{neq}(z;\xi)$ solves

$$s(a,E) = \max_{\rho} \langle -\ln \rho \rangle \quad \text{over} \quad \langle A_k \rangle = a_k\,, \quad \langle H \rangle = E\,.$$

Parameters $\xi^k = -\frac{\partial s}{\partial a_k}$ are conjugate to means $a_k = \langle A_k \rangle_{neq}$.

Zubarev's "Nonequilibrium Statistical Ensemble" :

ρ_{neq} is quasi-equilibrium associated with "slow" variables A_k; generally, include memory in the resolved variables.

Statisticians' "Exponential Family" :

A_k and H are minimal sufficient statistics for inference of parameters ξ^k and β within the model ρ_{neq}.

Quantifying the lack-of-fit of the model

Along any feasible path $\xi(t)$ in statistical parameter space, evaluate the *Liouville residual* of the model:

$$R \doteq \left(\frac{\partial}{\partial t} + L \right) \ln \rho_{neq}(z\,;\xi(t))$$

For canonical trial densities, ρ_{neq}:

$$R(\,z\,;\,\xi(t),\dot{\xi}(t)\,) \,=\, \dot{\xi}^k(t) Q_H[\,A_k(z) - a_k(t)\,] + \xi^k(t) L A_k(z)\,,$$

where $Q_H = I - P_H$ is projection orthogonal to $H(z) - E$.

$$\langle\, R \,\rangle_{neq} = 0\,, \qquad \langle\, R(H - E)\,\rangle_{neq} = 0\,.$$

Define the model *lack-of-fit* to be

$$\mathcal{L}(\xi, \dot{\xi}) \,=\, \frac{1}{2} \langle\, R^2 \,\rangle_{neq}$$

Two interpretations of the Liouville residual R

1. *Moment* of Liouville equation against any observable $B(z)$:

$$\frac{d}{dt}\langle B\rangle_{neq} = \langle LB\rangle_{neq} + \langle BR\rangle_{neq} \,.$$

2. *Kullback-Leibler divergence* between exact and model p.d.f.s:

$$D_{KL}(\rho_{exact} \,\|\, \rho_{model}) = \int_\Gamma \ln \frac{\rho_{exact}(z)}{\rho_{model}(z)} \, \rho_{exact}(z)\, dz$$

Incremental information loss along the path $\xi(t)$ in time Δt:

$$D_{KL}(\, e^{-\Delta t\, L}\rho_{neq}(\xi(t)) \,\|\, \rho_{neq}(\xi(t+\Delta t)\,) \,) = (\Delta t)^2\, \mathcal{L}(\xi, \dot{\xi}) + O(\Delta t)^3$$

Path optimization

To model relaxation to equilibrium from initial state solve

$$\min_{\xi(t)} \int_0^\infty \mathcal{L}(\xi, \dot{\xi})\, dt \qquad \text{subject to} \quad \xi(0) = \xi_0\,.$$

This optimization resembles a classical least action principle, but

- its lack-of-fit "Lagrangian" $\mathcal{L}$ is positive-definite
- its "action" has units of entropy production (information loss rate)
- its extremals tend to equilibrium as $t \to +\infty$

An optimal path $\xi(t)$ solves the Euler-Lagrange equations with initial and terminal conditions: $\xi(0) = \xi_0\,, \;\; \xi(+\infty) = 0.$

Optimal control algorithms can be implemented to solve these two-point boundary value problems.

Closure via Hamilton-Jacobi theory

Conjugate "momenta" and lack-of-fit "Hamiltonian"

$$
\begin{aligned}
\pi_k &= \frac{\partial \mathcal{L}}{\partial \dot{\xi}^k} = \langle A_k R \rangle_{neq} = \frac{d}{dt}\langle A_k \rangle_{neq} - \langle L A_k \rangle_{neq} \\
\mathcal{H}(\xi, \pi) &= \max_{\dot{\xi}} \, \pi_k \dot{\xi}^k - \mathcal{L}(\xi, \dot{\xi}) = \ldots
\end{aligned}
$$

The *value function* for the optimization problem

$$
v(\xi_0) = \min_{\xi(0)=\xi_0} \int_0^\infty \mathcal{L}(\xi, \dot{\xi}) \, dt
$$

solves the stationary Hamilton-Jacobi equations:

$$
\mathcal{H}\left(\xi, -\frac{\partial v}{\partial \xi}\right) = 0 \,, \qquad \pi_k = -\frac{\partial v}{\partial \xi^k} \,.
$$

Eliminating π produces the *optimal closure.*

Closed reduced equations

$$\frac{d}{dt}\langle A_k \rangle_{neq} = \langle LA_k \rangle_{neq} - \frac{\partial v}{\partial \xi^k}$$

Thermodynamic structure of optimal closure

"General Equations of NonEquilibrium Reversible Irreversible Coupling"
= "GENERIC", due to Grmela and Öttinger (1984, 1997)
"Metriplectic dynamics", concurrently due to Morrison (1984).

$$\frac{da_k}{dt} = J_{k\ell}(a)\frac{\partial h}{\partial a_\ell} - \frac{\partial v}{\partial \xi_k}$$

- Conjugate variables: $a_k = \langle A_k \rangle_{neq}$ and $\xi_k = -\frac{\partial s}{\partial a_k}$
- *Reversible part* has generalized symplectic structure with Poisson matrix $J_{k\ell} = \langle \{A_k, A_\ell\} \rangle_{neq}$, mean energy $h(s, a) = \langle H \rangle_{neq}$
- *Irreversible part* is gradient of dissipation potential $v(\xi)$

Entropy production: $s = -\langle \ln \rho_{neq} \rangle_{neq}$ satisfies

$$\frac{ds}{dt} = \xi^k \frac{\partial v}{\partial \xi^k} = \begin{cases} \approx 2v & \text{near equilibrium} \\ \geq v & \text{convex potential} \end{cases}$$

Relation to linear irreversible thermodynamics:

When the optimal closure is linearized around equilibrium, ξ_k and π_k are thermodynamic "force" and "flux" pair :

$$\pi_k = -\sum_{\ell} M_{k,\ell}\xi_\ell \qquad \text{and} \qquad v(\xi) = \frac{1}{2}\sum_{k,\ell} M_{k,\ell}\xi_k\xi_\ell$$

Hamilton-Jacobi equation reduces to matrix Riccati equation:

$$JC^{-1}M \,-\, MC^{-1}J \,+\, MC^{-1}M \,=\, D$$

$$C = \langle AA^T \rangle_{eq}\,, \quad J = \langle (LA)A^T \rangle_{eq}, \quad D = \langle (Q_A LA)\,(Q_A LA)^T \rangle_{eq}\,,$$

Optimal closure for energy spectrum in TBH

S. Thalabard and B.T., "Optimal response to nonequilibrium disturbances under truncated Burgers-Hopf dynamics." J. Phys. A: Math. Theor. (2017).

Inviscid Burgers-Hopf equation

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} = 0\,, \qquad u(x+2\pi, t) = u(x)\,.$$

under Galerkin projection, $u(x,t) = \sum_{k=-n}^{n} z_k(t)\, e^{ikx}$,

$$\frac{dz_k}{dt} = -ik \sum_{p+q=k} z_p z_q \qquad \text{with} \quad z_{-k} = z_k^*, \;\; z_0 = 0\,.$$

Statistical equilibrium:

$$\rho_{eq}(z) = \exp(-\beta H - \phi(\beta))\,, \quad H = \frac{1}{2}\sum_{k=-n}^{n} |z_k|^2$$

Equipartition energy spectrum: $\langle |z_k|^2 \rangle = \frac{1}{\beta}\,, \qquad E = \langle H \rangle = \frac{n}{\beta}$

Thermalization of the energy spectrum

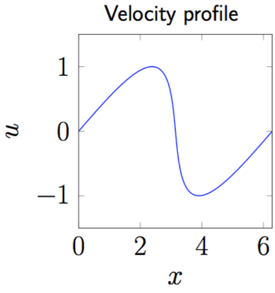
Dynamical behavior:

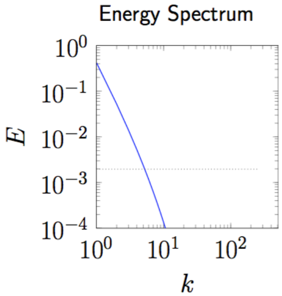
- Shock formation excites high modes, leading to a near-equipartition spectrum in the high modes.
- The equipartition spectrum gradually invades the low modes as the system thermalizes.

Goal: Model the relaxation of the energy spectrum toward statistical equilibrium using the resolved variables $A_k = |z_k|^2$, for $k = 1, \ldots, m \leq n$.

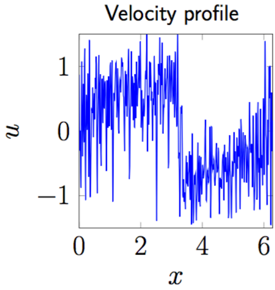
Derive closed reduced equations for the energy spectrum $\langle A_k \rangle(t)$, which tends to equilibrium from an initial perturbation.

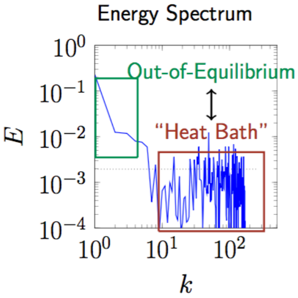
Early Times : Pre-shock





Late Times : Self-Thermalization





Nonequilibrium statistical model: $\rho_{neq}(z;\xi)$ is Gaussian with

$$\langle z_k \rangle_{neq} = 0, \quad \langle z_k z_\ell^* \rangle_{neq} = \xi_k^{-1} \delta_{k\ell} \qquad \left[\ E_k = \xi_k^{-1}\ \right]$$

- Quasi-equilibrium ensembles are simplest choice
- Memory not included in these trial p.d.f.s

Lack-of-fit Lagrangian:

$$\mathcal{L}(\xi,\dot{\xi}) = \frac{1}{2}\langle R^2 \rangle_{neq} = \frac{1}{2}\sum_k \left(\frac{\dot{\xi}^k}{\xi^k} \right)^2 + \frac{1}{6} \sum_{p+q=k} \frac{\left(p\xi^p + q\xi^q - k\xi^k\right)^2}{\xi^p \xi^q \xi^k}$$

Envelope closure (non-stationary version) :

- Initially disturbed state coincides with a trial density
- Feasible paths have fixed initial state and free terminal state
- Predicted evolution is envelope of computed optimal paths

Single mode relaxation – the analytically solvable case

Only one low mode k out of equilibrium, $\xi^{\ell} = \beta$ for $\ell \neq k$.

$$\mathcal{L}(\xi^k, \dot{\xi}^k) = \frac{1}{2}\left(\frac{\dot{\xi}^k}{\xi^k}\right)^2 + \frac{Ek^2(\beta - \xi^k)^2}{\beta \xi^k}, \qquad E = \frac{n}{\beta} \quad \text{is total energy.}$$

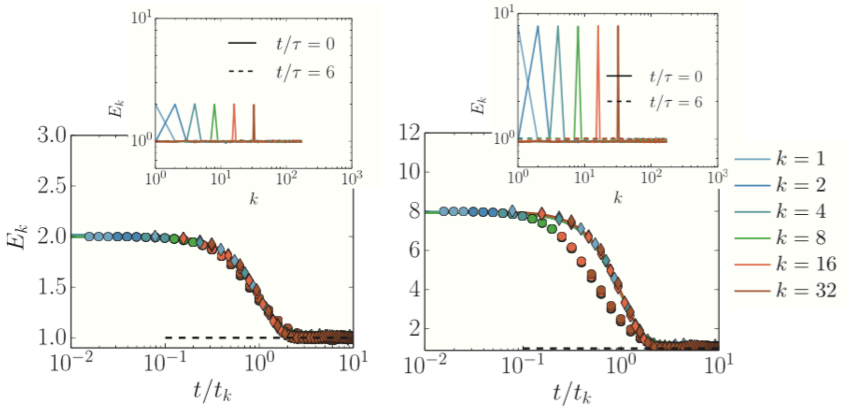
Reduced equation for relaxation of mode k :

$$\dot{\xi}^k = \frac{\sqrt{2}}{\tau_k}\left(\frac{\xi^k}{\beta}\right)^{\frac{1}{2}}\left(\beta - \xi^k\right), \qquad \tau_k = k^{-1}E^{-\frac{1}{2}} \quad \text{is time scale .}$$

Optimal closure predicts time scaling and temporal profile of relaxation *with no parameters tuned to data.*

Comparison to ensemble DNS with $n = 512$

Initial energy in mode k doubled (left) or times 8 (right)



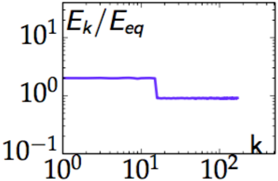
Quasi-normal approximation breaks down for very large perturbations.

DNS :

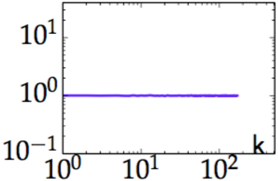
Perturbation of the lowest modes :
k<16

Resolution : 512
20, 000 samples

$t/\tau = 0$: $E_k^{pert} = 2E_{eq}$



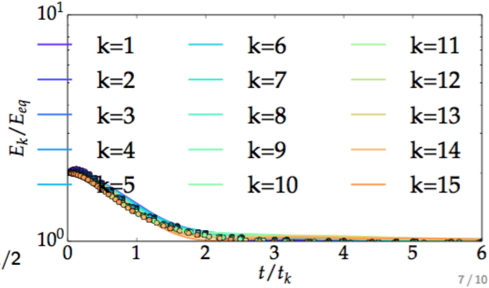
$t/\tau = 6$: $E_k \simeq E_{eq}$



Quasi-Normal Optimal Closure :

● : closure
— : dns

$$t_k = k^{-1}E^{-1/2}$$



Conclusions and prospects

- Optimal closure is a coarse-graining that derives a macroscopic evolution from an underlying microscopic dynamics using an intrinsic best-fit criterion without tuning to data
- The equations governing the reduced model have a natural *thermodynamic structure*
- Near-equilibrium optimal closure produces *transport coefficients*, and *linear response kernels* for forced systems
- Beyond linear response its predictions are not analytically accessible, but are computable via *optimal control algorithms*
- Model fidelity depends on good choices of the resolved variables and trial p.d.f.s.
  - *Data-driven models* that include memory are a topic of current investigations.
  - What about much smarter trial densities?!