Background on nonequilibrium stat mech, chaotic dynamics and effective theories

Tom Jackson

July 20, 2022

Abstract

This is a (disorganized) summary writeup on with pointers to literature I've compiled for my own use.

Chapter 1

Coarse-graining — non-RG

1 Koopman

1.1 Definition

Koopman-von Neumann formalism proposed as classical analogue of quantum time evolution operator:

$$\langle x|e^{iHt/\hbar}|y\rangle \sim \mathcal{K}_t(x, y).$$

Acts *linearly* on Hilbert space of observables $a(x)$, not system state x directly.

$$\mathcal{K}_t(x, y) = \delta(y - f_t(x)); \qquad [\mathcal{K}_t a](x) = a(f_t(x)),$$

where $f_t(x)$ denotes the forward time evolution starting from state x. The operators $\mathcal{K}_t$ form a one-parameter semigroup in t.

It's the adjoint (not inverse) of Perron-Frobenius operator $\mathcal{L}$, in terms of evolving probability densities:

$$\int_M dx\, [\mathcal{K}_t a](x)\rho(x) = \int_M dy\, a(y)[\mathcal{L}_t\rho](x).$$

Right eigenvectors of PF are left eigenvectors of Koopman, and vice-versa. Perron-Frobenius, by virtue of acting on ρ, is taking about a bundle of trajectories. Koopman talking about one single trajectory. Note that the "state space" M is very arbitrary and could be, e.g., a fractal attractor.

If we knew eigendecomposition of Koopman, we could compute time evolution of any observable. Other statements about its spectrum:

- The constant function on M is always an eigenfunction with eigenvalue 1. If this is the only eigenfunction with this eigenvalue, the system is ergodic (time averages are M-averages).

1.2 Local vs. global

Stuff by Page and Kerswell on convergence.

1.3 Mode decompositions

Linearity of $\mathcal{K}$ means learning from data is a convex optimization problem, which we might hope is tractable.

Obstacles to tractable approximations: $\mathcal{K}$ may not have eigenvalues (other than 1) and is not compact.

Could try projecting $\mathcal{K}$ onto a finite-dimensional subspace; if an eigenvector (or invariant subspace) of full $\mathcal{K}$ lies in the subspace, it'll be an invariant of the projection too. Don't know a good subspace, though, and also need some regularization (since even point spectrum may be dense.)

1.3.1 Ulam's method

Impose a disjoint partitioning of state space, and compute Markov transition matrix between partitions induced by dynamics; this approximates PF. Basis functions are indicators on partitions; regularization by small phase-space diffusion.

Problem is required number of partitions grows exponentially with phase space dimension.

1.3.2 (E)DMD

Choose a basis ("dictionary") of observables ψ_i with Gram matrix $G = \langle \psi_i, \psi_j \rangle$; formally Koopman matrix is $K_{ij} = \langle \psi_i, \mathcal{K}\psi_j \rangle$; so formally Koopman projection (in least-squares sense) is $G^+ K$, where we took the Moore-Penrose pseudoinverse.

We can't evaluate $\langle \psi_i, \mathcal{K}\psi_j \rangle$. Instead we proceed by a data-driven approximation: we assume we have access to dynamical snapshots of the observables $\psi_i(x_n)$, along with their time-stepped counterparts $\psi_i(Tx_n)$, for a large number of system states. This defines an empirical inner product $(f, g) = \frac{1}{N} \sum_n f^\dagger(x_n) g(x_n)$, allowing the projection to be evaluated.

Can show this converges to the real inner product as $N \to \infty$ if the system is ergodic, or if the states $\{x_n\}$ are sampled from the dynamical invariant measure. Spectrum of projected Koopman does *not* converge to that of $\mathcal{K}$, but claim weak convergence if all $\{\psi_i\}$ are orthonormal.

Frequently regularize by doing PCA on G.

Original DMD chose ψ_i to be state variables. Can extend this by augmenting basis with powers or other nonlinear functions of state variables. Can also use time-delayed copies of observables (cf. Takens embedding).

"One may also use an implicit basis determined by a kernel function (28), in which case the (possibly infinite-dimensional) space of observables is the corresponding reproducing kernel Hilbert space."

kernel DMD; Hankel DMD

1.3.3 Neural methods: autoencoder

1.3.4 Neural methods: manifold learning

For ergodic systems specifically; claim work even with chaotic spectrum? Need observational timeseries long enough to probe ergodic timescale.

2 Mori-Zwanzig

2.1 Relation to Koop

Ref: Mezic appendix in "geometry and learning."

Have space of observables $\{a_i\}$ and orthogonal projection P onto it; Q is the complement.

Chapter 2

Renormalization group

1 RG background

1.1 Conventional overview

1.1.1 Terminology

Couplings in $\mathcal{L}$ as "coordinates on theory space." Scattering amplitudes; partition function; free energy.

Feynman diagrams as a weak-coupling perturbation series around a noninteracting theory. Expansion in number of loops equivalent to expanding action in powers of $\hbar$ (argument?)

1.1.2 Gell-Man-Low version

Take scalar ϕ^4 theory for simplicity. Loop integrals diverge, so require regularization parameterized by Λ. Then have scattering amplitude $\Gamma^{(4)} = \Gamma^{(4)}(\{\mathbf{k}_{\text{ext}}\}; \lambda_0, \Lambda)$. In particular, couplings gain *scale dependence* on the magnitude of the external momenta with which we probe them.

1. "Bare" couplings $\{\lambda_0\}$ are unobservable: we instead can only measure the infinite sum of diagrams making up each proper vertex $\Gamma^{(n)}$ at external reference momenta $\mathbf{k}_P$; for simplicity refer scale of these momenta to a scale μ.
   This observational data provides *normalization conditions* defining physical couplings $\lambda_P(\mathbf{k}_P)$.
2. Regularization is unphysical and artificial, so physical observables can't depend on Λ. This isn't a problem, because as argued in the previous point the couplings λ_0 are physically unobservable as well.
   So the preceding notation is misleading: we need to make λ_0 dependent on Λ in such a way that overall Λ-dependence of physical quantities vanishes. Do this in the context of the diagrammatic expansion by introducing *counterterms* for λ_0.
3. Combining these, a sufficient set of normalization conditions fixes all counterterms at a given order in perturbation theory, up to finite parts: this is the *cancellation of divergences* in old-style terminology.

$$\lambda_P(\mu) = \lambda_0 - 3c\lambda_0^2 \log\left(\frac{\Lambda^2}{\mu^2}\right) + \dots, \tag{1.1}$$

which can be expressed as the RG flow equation

$$\mu \frac{d}{d\mu} \lambda_P(\mu) = 6c\lambda_P^2(\mu) + \ldots .$$

(1.2)

BHPZ version is just that counterterms can be constructed diagram-by-diagram.

Ambiguity in finite part of counterterms describes our freedom to choose a regularization scheme, i.e. to choose Λ.

Bring in power-counting. A "nonrenormalizable" theory requires an infinite number of conditions to fix counterterms: conditions in 1-1 correspondence with naively divergent terms in initial $\mathcal{L}$.

1.1.3 Wilsonian version

1.1.4 Exact RG

[?]

Coarse-grain fields by convolution with a smoothing filter of scale Λ: $\phi_\Lambda(k) = K_\Lambda(k)\phi(k)$. Gaussian part of the theory just picks up $G_\Lambda = K_\Lambda^2 G$. Non-Gaussian part evolves according to a functional Fokker-Planck:

$$\frac{\partial}{\partial \Lambda} e^{-V_\Lambda} = \frac{1}{2} \frac{\partial G_\Lambda}{\Lambda} \frac{\delta^2}{\delta^2 \phi_\Lambda^2} e^{-V_\Lambda};$$

V_Λ is the smoothed effective potential (so transforming 1PI Γ?). Can also accommodate Gaussian part in a single equation.

Argues for a stochastic approach to RG, since we throw away information both in truncating the exact RG and in coarse-graining field configurations; ideally these would be coupled, so that we (progressively?) truncate coupling constants "governing" small-scale degrees of freedom.

2 RG: new directions

2.1 Jona-Lasinio

2.2 Langevin simulations

For a generic action $S[\phi]$, Langevin dynamics are

$$\partial_t \phi_i = \frac{\delta S}{\delta \phi_i} + \eta_i(t).$$

Associated Fokker-Planck is

$$\partial_t P(\phi, t) = \frac{\delta}{\delta \phi_i} \left[-\frac{\delta S}{\delta \phi_i} P + \frac{\delta P}{\delta \phi_i} \right],$$

with a unique stationary solution given by the Boltzmann weight $P \propto \exp S$.

For numerics, we need to discretize the Langevin equation with as large a timestep as possible. Interpret a noninfinitesimal step as leading to an exact solution for an action with modified couplings.

2.3 Caticha: Changes of variables

[?]

2.4 Goldenfeld-Oono

2.5 Mitter: RG as Bellman

[?]
Linear parabolic equation

$$dp(x,t) = \widehat{L}^\dagger p + \frac{1}{\varepsilon} V(x,t)p,$$

where $\widehat{L}^\dagger$ is the formal adjoint of the diffusion operator

$$\widehat{L} = \sum_i f_i(x)\partial_i + \frac{\varepsilon}{2}\partial_i^2.$$

Initial condition $p(t = 0) \propto \exp[-S_0(x)/\varepsilon]$. $\varepsilon > 0$ is some singular perturbing parameter.

Make the transformation $S(x,t) = -\varepsilon \log p(x,t)$, then S satisfies a Hamilton-Jacobi-Bellman equation

$$\partial_t S - \frac{\varepsilon}{2}\Delta S + H[\nabla S] = 0;$$

where $H[p] = p'f(x) + \frac{1}{2}|p|^2 - V$. Formal $\varepsilon \to 0$ limit is a Hamilton-Jacobi equation $\partial_t S + H[\nabla S] = 0$. Recall the definition of that; canonical momenta $p_i = \partial_i S$, where "Hamilton's principal function" is the action (up to a constant) $S(x,t) = \int_\gamma d\tau \mathcal{L}$, where γ is the extremizing trajectory of the Euler-Lagrange equations going from the initial conditions to a configuration x at time t.

Claim the $\varepsilon \to 0$ limit of S is the value function J of a deterministic optimal control problem:

$$J(t,u,x_0) = S_0(x_0) + \frac{1}{2}\int^t |u(s)|^2 \, ds;$$

S is the minimal (under what?) J satisfying $\frac{dx}{ds} = f[x(s)] + u(s)$ when $u(s)$ is taken over trajectories connecting the initial condition to x at time t.

When $\varepsilon \to 0$, claim S has the same interpretation, but for a stochastic control problem.

3   Disorganized speculation

RG flow analogous to dynamics of particles in a filter. Same notion of solving for a fixed point "incrementally" via a dynamical system.

Have exact RG, in a full functional space, and its natural truncation in terms of relevant/irrelevant ops. Can we use that?

Understand how exact RG is a Fokker-Planck. Understand how (doubled?) Fokker-Planck gives us Schrodinger. [so is ERGE saying something about ground-state wavefunctional??]

Ref in Mitter 1989 of Hamilton-Jacobi-Bellman arising from RG.

Are particles doing something deeper than providing a finite-D truncation of a function space (that of distributions, via sampling?) Is it Nelson's stochastic QM? Can we do better with another representation of distributions (as diffeomorphisms) on function spaces? Could we use (wavelet-style) normalizing flows?

MacKay remark in MCMC with "sampling as a lossy channel."

3.0.1 sidebar

Can we use info-theoretic ideas to regularize the function space? E.g. discretization according to KL divergence, or our vague notion of "Bayesian distinguishability" via a finite number of samples.

"KS, other nonparametric tests have poor small-sample efficiency." But we want this to be a feature. Wikipedia: K-S test, Anderson-Darling test, Donsker's Theorem, Minimum-distance estimation. Other metrics: total variation, Wasserstein/earth-mover, Jensen-Shannon, Lévy-Prokhorov; KL divergence vs KS distance.

"A distribution-free multivariate Kolmogorov?Smirnov goodness of fit test has been proposed by Justel, Peña and Zamar (1997).[22] The test uses a statistic which is built using Rosenblatt's transformation, and an algorithm is developed to compute it in the bivariate case. An approximate test that can be easily computed in any dimension is also presented." https://academic.oup.com/mnras/article/451/3/2610/1186451 https://www.sciencedirect.com/science/article/pii/S0047259X030007

https://stats.stackexchange.com/questions/67341/what-are-some-bayesian-alternatives-to-the-kolmogorov-smirnov-test https://arxiv.org/abs/0910.5060

3.1 versus filtering

Fokker-Planck connected to Langevin dynamics, Feynman-Kac path integral (before we get to Nelson stuff). Goldenfeld book chapter on RG for nonlinear diffusion, but where's our nonlinearity?

Other refs on optimal transport.

What's the notion of "RG scale" in filtering?

3.2 How do we use it?

How can we bring this to problems we want to solve? 1) expensive forward model dynamics 2) particles expensive as a consequence; number of particles much less than number of variables 3) crappy covariance estimates for ensemble KF as a result.

What if we make further constraining assumptions on forward model, e.g. that dynamics given my a least-action or other optimum principle?

"multi-object tracking" = efficiently approx general distribution as mixture of gaussians? Although that would be similar to applying a RBF estimator to a bunch of particles.

References