Using coupling method to detect underlying dynamics

Yao Li

Department of Mathematics and Statistics, University of Massachusetts Amherst

December 1st, 2022

BIRS workshop: Topics in Multiple Time Scale Dynamics

Overview: data-driven computing

General idea: simulation data + X

- Fokker-Planck solver: meshed version (with J. Zhai and M. Dobson), DNN version (with J. Zhai, M. Dobson, C. Meredith)
- Sensitivity analysis (with J. Zhai and M. Dobson)
- QSD and sensitivity (with Y. Yuan)
- DNN surrogate of neurons (with L. Tao, Z. Xiao et al)

This work (with S. Wang and M. Tao)

- Coupling time data from Monte Carlo
- Detect dynamics via the time scale of coupling rate

Heuristics: Random perturbation of dynamical systems

- Deterministic dynamical system $X_{n+1} = f(X_n)$ or $X' = f(X)$
- Add a small noise with magnitude ϵ
- Invariant probability measure π_ϵ
- Assume $\|\mu P^t - \pi_\epsilon\| \sim exp(-r(\epsilon)t)$
- Very important: $r(\epsilon)$ vs. ϵ
- The scale of $r(\epsilon)$ vs. ϵ reveals the dynamics

Difficulty

How to give a sharp bound of $r(\epsilon)$ efficiently??

Our answer: Coupling Method

A Markov process (Φ_n^1, Φ_n^2) on the state space $X \times X$ is said to be a Markov coupling if

1. Two marginal distributions are Markov processes Φ_n with initial distribution μ and ν, respectively
2. If $\Phi_n^1 = \Phi_n^2$, then $\Phi_m^1 = \Phi_m^2$ for all $m \geq n$.

$\tau_{Cp} = \inf_{n \geq 0}\{\Phi_n^1 = \Phi_n^2\}$ is the *coupling time*.

Coupling Lemma

Coupling Lemma

$$\|\mu P^n - \nu P^n\|_{TV} \leq 2\mathbb{P}[\tau_{Cp} > n] .$$

Optimal coupling (Pitman 1970s)

There exists a coupling (Φ_n^1, Φ_n^2) (may not be Markov) such that

$$\|\mu P^n - \nu P^n\|_{TV} = 2\mathbb{P}[\tau_{Cp} > n] .$$

The existence of "honest" optimal coupling remains open.

Upper and lower bound

Lower bound

Estimate r_l such that

$$\mathbb{P}[\tau_{C_p} > t] \approx e^{-r_l t}\,.$$

$r_l < r$ is a lower bound of geometric ergodicity rate.

Upper bound

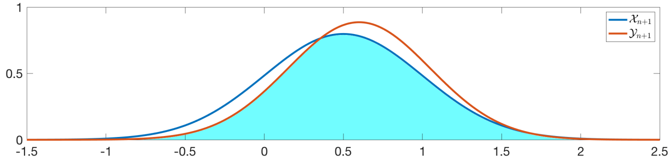
Construct disjoint sets (A_t, B_t). Run coupling $(\mathcal{X}_t, \mathcal{Y}_t)$ with $\mathcal{X}_0 \in A_0$ and $\mathcal{Y}_0 \in B_0$.

$$\xi_C = \min\left\{\inf_t\{\mathcal{X}_t \notin A_t\}, \inf_t\{\mathcal{Y}_t \notin B_t\}\right\}, \quad \mathbb{P}[\xi_C > t] \approx e^{-r_u t}$$

$r < r_u$ is an upper bound of geometric ergodicity rate.

Implementation: How to couple numerically?

- Numerical simulation comes with error
- Risk of "near miss" each other!
- Solution: Switch to maximal coupling when X_t and Y_t are sufficiently close to each other.
- Maximal coupling: compare probability density functions. Couple the "overlapping" part
- Robust against small perturbations



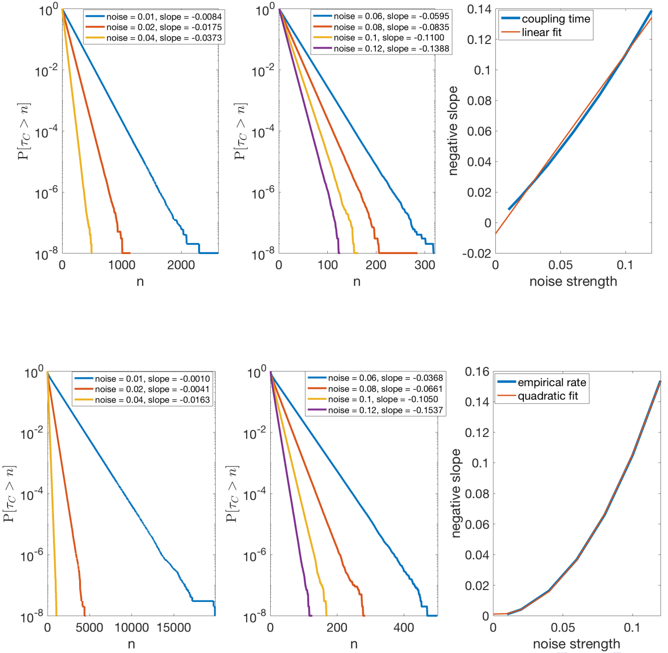
Warm-up example: 1D mappings

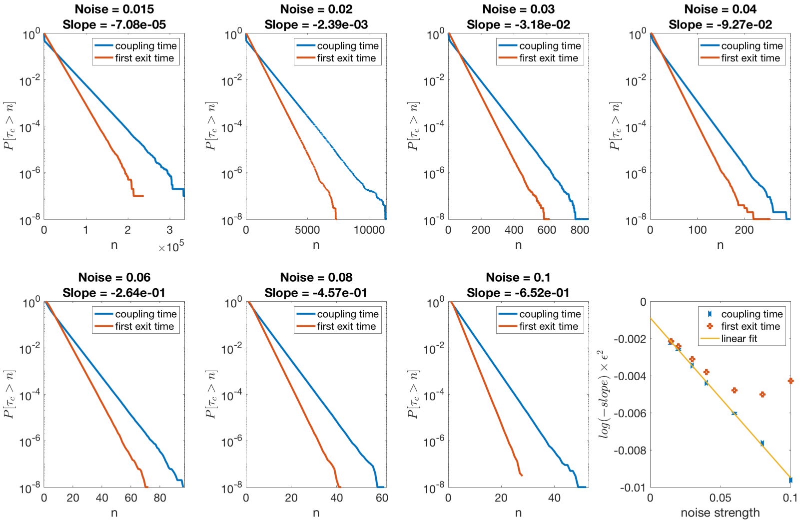
Consider $X_{n+1} = f(X_n)$ for

- $f_1(x) = 2x + a\sin(2\pi x) \quad (mod\, 1)$, $a < 1/(2\pi)$: mixing
- $f_2(x) = x + \sqrt{2} \quad (mod\, 1)$: ergodic but not mixing
- $f_3(x) = 3.2x(1 - x)$: stable periodic orbit

Result

- Add small noise $\epsilon\zeta_n$, $\zeta_n \sim N(0, 1)$
- Coupled trajectories (X_n, Y_n)
- Change to maximal coupling when $|X_n - Y_n| \le \epsilon$
- Exponential tail of coupling time distribution: $r(\epsilon)$
- Relation $r(\epsilon)$ vs. ϵ for three mappings: linear, quadratic, exponential


noise = 0.01, slope = -0.0084
noise = 0.02, slope = -0.0175
noise = 0.04, slope = -0.0373
$P[\tau_C > n]$
n
noise = 0.06, slope = -0.0595
noise = 0.08, slope = -0.0835
noise = 0.1, slope = -0.1100
noise = 0.12, slope = -0.1388
coupling time
linear fit
negative slope
noise strength
noise = 0.01, slope = -0.0010
noise = 0.02, slope = -0.0041
noise = 0.04, slope = -0.0163
noise = 0.06, slope = -0.0368
noise = 0.08, slope = -0.0661
noise = 0.1, slope = -0.1050
noise = 0.12, slope = -0.1537
empirical rate
quadratic fit


Noise = 0.015
Slope = -7.08e-05
Noise = 0.02
Slope = -2.39e-03
Noise = 0.03
Slope = -3.18e-02
Noise = 0.04
Slope = -9.27e-02
Noise = 0.06
Slope = -2.64e-01
Noise = 0.08
Slope = -4.57e-01
Noise = 0.1
Slope = -6.52e-01
coupling time
first exit time
linear fit
$P[\tau_c > n]$
n
$\times 10^5$
$log(-slope) \times \epsilon^2$
noise strength

Stochastic differential equations

1. Stochastic differential equation

$$\mathrm{d}X_t = f(X_t)\mathrm{d}t + \sigma \mathrm{d}W_t$$

2. Continuous time Markov process
3. Time-dependent transition kernel

$$P^t(x, A) = \mathbb{P}[\Phi_t \in A \mid \Phi_0 = x]$$

4. Euler-Maruyama scheme

$$X_{n+1} = X_n + hf(X_n) + \sqrt{h}\sigma Z_n$$

for $Z_n \sim N(0, Id_d)$

Couple stochastic differential equations

Let $(X_t^{(1)}, X_t^{(2)})$ be a coupling of two trajectories of SDEs

- Independent. $X_t^{(1)}$ and $X_t^{(2)}$ are independent until coupling.
- Synchronous. Use the "same noise".
- Reflection. Use "mirroring" random terms. Assume σ is a constant matrix.

$$\mathrm{d}X_t^{(1)} = f(X_t^{(1)})\mathrm{d}t + \sigma \mathrm{d}W_t$$
$$\mathrm{d}X_t^{(2)} = f(X_t^{(2)})\mathrm{d}t + \sigma(I - 2\mathbf{e}_t\mathbf{e}_t^T)\mathrm{d}W_t$$

with

$$\mathbf{e}_t := \frac{\sigma^{-1}(X_t^{(1)} - X_t^{(2)})}{\|\sigma^{-1}(X_t^{(1)} - X_t^{(2)})\|}$$

(Ref: Lindvall, Chen)

How to couple SDEs numerically

- Strategy: mixed coupling
- Step 1: Run reflection/independent/synchronous coupling until $\|X_t^{(1)} - X_t^{(2)}| \le \epsilon$
- Step 2: Compare density function with time step h. Run maximal coupling
- If success, $X_t^{(1)}$ and $X_t^{(2)}$ are coupled. Otherwise return to Step 1.

Theorem (Joint with S. Wang, 2020)

Let τ_{Cp} and $\bar{\tau}_{Cp}$ be coupling times of $(X_t^{(1)}, X_t^{(2)})$ and their numerical approximation, respectively. Let h be the time step of the Euler-Maruyama method. Under certain regularity condition

$$\lim_{h \to 0} |\mathbb{P}_{(x,y)}[\tau_{Cp} > t] - \mathbb{P}_{(x,y)}[\bar{\tau}_{Cp} > t]\| \to 0$$

for any $t > 0$.

Geometric ergodicity

Lower bound

- Run mixed coupling repeatedly
- Plot $\mathbb{P}[\tau_{Cp} > t]$ vs. t in log-linear plot
- If $\mathbb{P}[\tau_{Cp} > t]$ has slope $-r$ in log-linear plot for large t, by coupling lemma, X_t is geometrically ergodic with rate $\geq r$

Upper bound

- Divide the phase space into partition A and B
- General rule: Make transition between A and B as difficult as possible
- Simulate the minimum of first passage times

$$\xi_C = \min\left\{\inf_t\{\mathcal{X}_t \notin A_t\}, \inf_t\{\mathcal{Y}_t \notin B_t\}\right\}$$

repeatedly

Discussion

Mixed coupling strategy 1

- Run independent coupling using "identical noise"
- Change to maximal coupling when trajectories are close to each other
- Coupling rate depends on Lyapunov exponent (dissipative case) or Li-Yorke chaos (chaotic case)
- Connect random dynamical system and stochastic process

Mixed coupling strategy 2

- Run reflection coupling using "symmetric noise"
- Change to maximal coupling when trajectories are close to each other
- Reflection coupling is optimal for many SDEs
- Gives good estimate of convergence in applications

Application: Can you hear the shape of a landscape?

Can you hear the shape of a drum?

By Mark Kac in 1966.

- Difficult to know the shape of high dimensional landscapes
- Idea: "hit" the landscape by injecting noise to the gradient flow
- Applications: deep learning, molecular dynamics, Bayesian inference

Application: Can you hear the shape of a landscape?

- Coupling time distribution can be used to detect high dimensional landscape $V(x)$
- Consider the stochastic gradient flow

$$\mathrm{d}X_t = -\nabla V(x)\mathrm{d}t + \epsilon \mathrm{d}W_t$$

- Run mixed coupling for a few different ϵ. Estimate

$$r(\epsilon) = \lim_{t\to\infty} -\frac{1}{t}\log \mathbb{P}[\tau_{C_p}(\epsilon) > t]$$

- $r(\epsilon)$ is related to the landscape of $V(x)$

Application: Can you hear the shape of a landscape?

One global minimum x^*

- $r(\epsilon)$ does not change with ϵ
- $r(\epsilon)$ approximates the least eigenvalue of Hessian matrix at x^*

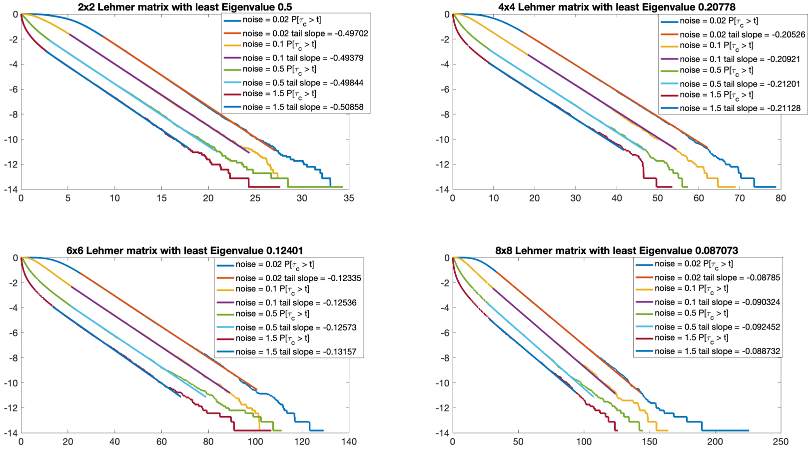
Many local minima

- $r(\epsilon)$ decay exponentially fast with smaller ϵ
- linear extrapolation of $-\epsilon^2 \log r(\epsilon)$ approximates maximal barrier height

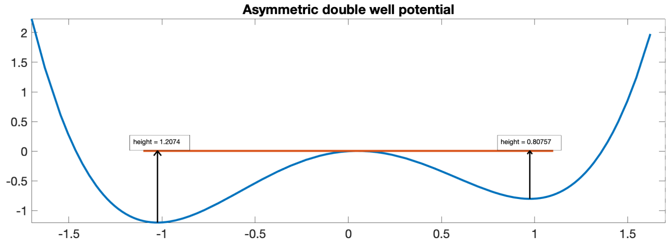
Application: Can you hear the shape of a landscape?

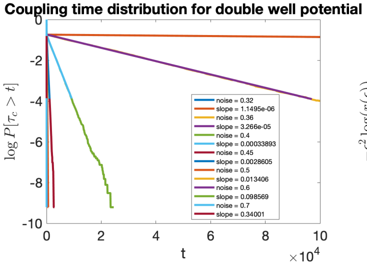
- Joint work with S. Wang and M. Tao.
- The scaling of $r(\epsilon)$ vs. ϵ depending on the dynamics
- One local minimum: linearized local dynamics when $\epsilon \ll 1$
- Many local minima: at least one trajectory needs to cross the barrier to couple
- Idea of proof: random sum of random variables

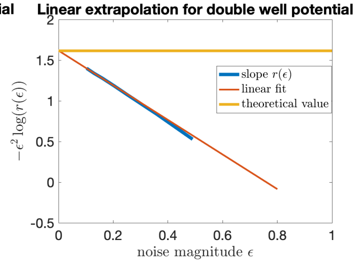
One global minimum



Double well potential

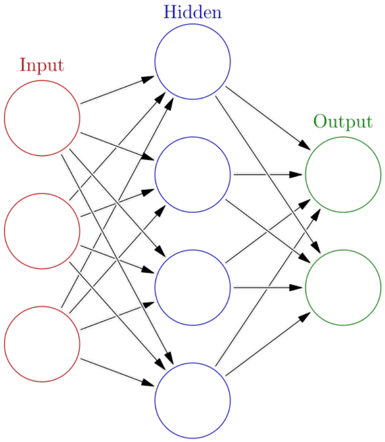




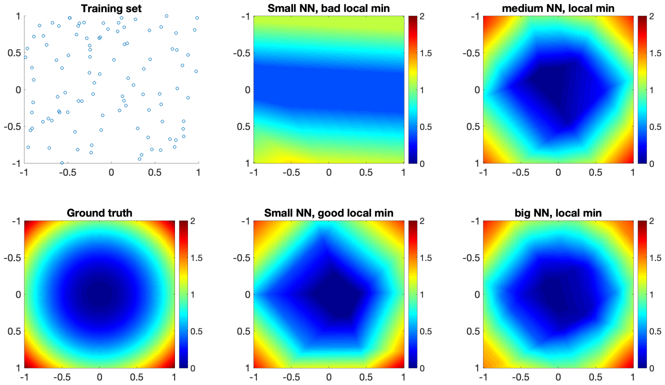


Application to deep learning

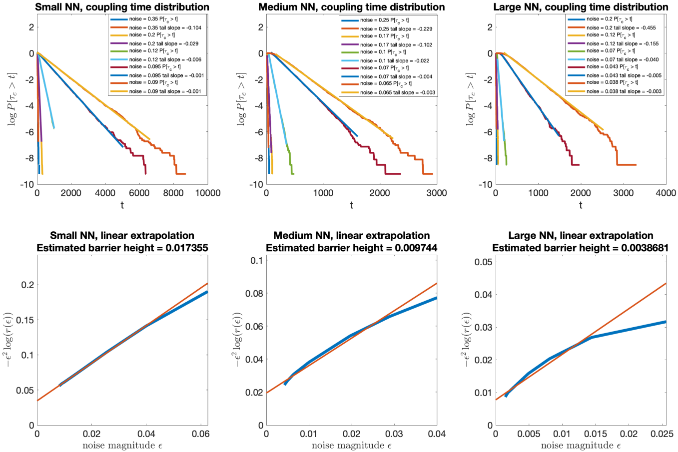
- Use three artificial neural networks (ANN) to learn a quadratic function
- Two hidden layers. Small ANN: $(4, 3)$, medium ANN: $(10, 10)$, large ANN: $(20, 20)$
- Small ANN has bad local minimum with higher barrier
- Consistent with other literatures



Learning result

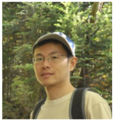


Coupling time distribution and linear extrapolation

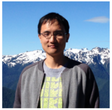


Acknowledgement

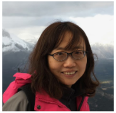
Collaborators



Molei Tao: Georgia Institute of Technology



Jiayu Zhai: Shanghai Tech University



Shirou Wang: Jilin University

Funding: NSF
DMS-1813246
DMS-2108628

Thank you