Projects

If you have questions or need suggestions: donati[at]zib.de

1 Kramers theory

Consider a particle of mass m moving along a one-dimensional reaction coordinate $\Gamma \subset \mathbb{R}$, the equation of the Langevin dynamics is written as

$$\begin{cases} \dot{x} = v_t \\ m\dot{v} = -\nabla V(x_t) - \gamma m v_t + \sigma \eta_t \,, \end{cases} \tag{1}$$

where $V(x)$ is a potential energy function, x and v describe the position and the velocity of the particle, γ is a friction coefficient, η_t is a white noise process, $\sigma = \sqrt{2k_B T m \gamma}$, with Boltzmann constant k_B and temperature T. According to Kramers theory, the transition rate for a bimetastable system is estimated as

- Low friction regime ($\gamma < \omega_B$)

$$k = \beta \gamma E_b \exp\left[-\beta E_b\right] \,, \tag{2}$$

- Moderate friction regime ($\gamma > \omega_B$)

$$k = \frac{\gamma}{\omega_B}\left(\sqrt{\frac{1}{4} + \frac{\omega_B^2}{\gamma^2}} - \frac{1}{2}\right) \cdot \frac{\omega_A}{2\pi} \exp\left(-\beta E_b\right) \,, \tag{3}$$

- High friction regime ($\gamma \gg \omega_B$)

$$k = \frac{\omega_A \omega_B}{2\pi\gamma} \exp\left(-\beta E_b\right) \,, \tag{4}$$

where we introduced

$$\omega_A = \sqrt{\frac{1}{m}\left.\frac{d^2V}{dx^2}\right|_{x=x_A}} \,, \tag{5}$$

and

$$\omega_B = \sqrt{\frac{1}{m}\left.\frac{d^2V}{dx^2}\right|_{x=x_B}} \,. \tag{6}$$

1.1 The BBK integrator scheme

To solve the Langevin equation, i.e. to generate a trajectory (x_t, v_t), you can use the Brünger-Brooks-Karplus (BBK) Langevin integrator. Consider a time interval $[0,\, \tau]$, and a time-discretization into N_τ sub-intervals $[t_k, t_{k+1}]$ of equal length Δt such that

$$\begin{aligned} t_0 &= 0 \\ t_1 &= \Delta t \\ t_2 &= 2\Delta t \\ &\dots \\ t_{N_\tau} &= \tau = N_\tau \Delta t \end{aligned} \tag{7}$$

Assume to know the position $x_n := x_{t_n}$ and the velocity $v_n := v_{t_n}$ of the particle at time t_n, then the position $x_{n+1} := x_{t_{n+1}}$ and the velocity $v_{n+1} := v_{t_{n+1}}$ of the particle at time t_{n+1} is calculated

as

$$\begin{cases} v_{n+1/2} = (1 - 1/2\Delta t\,\gamma)v_n - \dfrac{1}{2m}\nabla V(x_n)\,\Delta t + \dfrac{\sigma}{2\sqrt{m}}\,\eta_n\,\sqrt{\Delta t} \\ x_{n+1} = x_n + v_{n+1/2}\,\Delta t \\ v_{n+1} = \dfrac{1}{(1 + 1/2\Delta t\,\gamma)}\left[v_{n+1/2} - \dfrac{1}{2m}\nabla V(x_{n+1})\,\Delta t + \dfrac{\sigma}{2\sqrt{m}}\,\eta_n\,\sqrt{\Delta t}\right], \end{cases} \tag{8}$$

where $v_{n+1/2}$ is the velocity of an intermediate step between t_n and t_{n+1}, and η_n is a random number drawn from the normal distribution $\mathcal{N}(0,1)$. Note the same random number η_n is used twice, to calculate $v_{n+1/2}$ and v_{n+1}.

1.2 Project

Consider the two-dimensional double-well potential energy function

$$V(x,y) = 6\cos\left(m\,\arctan(y,x)\right) + 60\left(\sqrt{x^2+y^2} - 1\right)^2, \tag{9}$$

with $m = 2$ and defined on the space $\Gamma = \Gamma_x \times \Gamma_y \subset \mathbb{R}^2$. The potential is illustrated in fig. 1.

1.2.1 Part 1

1. Implement the BBK integrator scheme to solve the Langevin dynamics.
2. Start a large number of simulations from the left well and calculate the Mean First Passage Time (MFPT) $\langle\tau_{fp}\rangle$, i.e. the average time it takes for a particle to cross the barrier at $x = 0$
3. Calculate the transition rate as
$$k = \frac{1}{\langle\tau_{fp}\rangle}\,. \tag{10}$$
4. Repeat the experiment for different values of friction (consider values between 0 and 100 ps^{-1}).
5. Implement eqs. 2, 3, 4 and compare the results with the numerical experiment.

1.2.2 Part 2

The two-dimensional potential energy function defined in eq. 34 can be projected onto relevant coordinates $u(x,y)$. Consider the relevant coordinate

$$u(x,y) = x\,, \tag{11}$$

and the free energy profile

$$F_x(x) = -\frac{1}{\beta}\log\int_\Gamma \exp(-\beta V(x,y))\delta(u(x,y) - x)\,dydx \tag{12}$$

$$= -\frac{1}{\beta}\log\int_{\Gamma_y} \exp(-\beta V(x,y))\,dy\,, \tag{13}$$

where $\delta(u(x,y) - x)$ is the delta function, equal to 0 if $u(x,y) \neq x$, equal to 1 if $u(x,y) = x$.

1. Recalculate eqs. 2, 3, 4 with the free energy profile defined in eq. 13.
2. Does the use of free energy profile improve the result?

1.2.3 Part 3

A better relevant coordinate is the circumference defined by the function

$$r(x, y) = \sqrt{x^2 + y^2} = 1\,. \tag{14}$$

1. Find an expression for the free energy profile

$$F_r(r) \quad = \quad -\frac{1}{\beta} \log \int_\Gamma \exp(-\beta V(x, y)) \delta(r(x, y) - r)\, dydx\,. \tag{15}$$

2. Recalculate eqs. 2, 3, 4 with the free energy profile defined in eq. 15.
3. Does the use of free energy profile improve the result?

Listing 1: Parameters for a Langevin simulation.

```
kB      = 0.008314463                   # kJ mol-1 K
T       = 300                           # K
mass    = 1                             # amu mol-1
gamma   = 1                             # ps-1
sigma   = np.sqrt(2 * kB * T * mass * gamma)   # nm ps-1/2
beta    = 1 / kB / T                    # kJ-1 mol
```

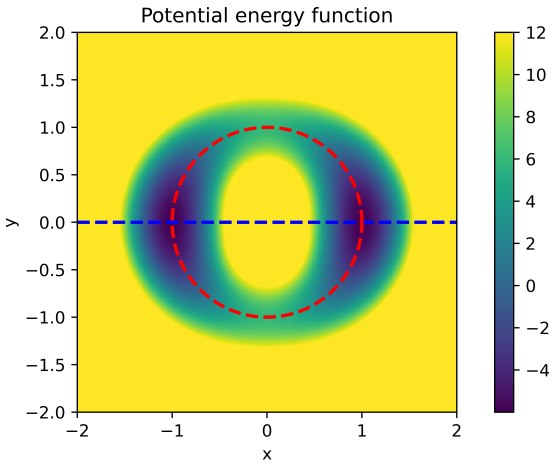

Figure 1: Potential energy function

2 SqRA of the infinitesimal generator

Consider the stochastic differential equation for the overdamped Langevin dynamics for a one-dimensional particle:

$$dx_t = -\frac{\nabla V(x_t)}{m\gamma}dt + \sqrt{2D}\,\eta_t \tag{16}$$

where $D = {}^{k_B T}/_{m\gamma}$ is the diffusion constant. Assume a Voronoi tessellation of the state space Γ into k disjoint subsets, then, the associated SqRA rate matrix is written as

$$Q_{ij} = \begin{cases} D\dfrac{\mathcal{S}_{ij}}{h_{ij}\,\mathcal{V}_i}\sqrt{\dfrac{\pi(x_j)}{\pi(x_i)}} & \text{if } i \sim j \\ -\displaystyle\sum_{\substack{j=1 \\ j\neq i}}^{k} Q_{ij} & \text{if } i = j \\ 0 & \text{else} \end{cases}, \tag{17}$$

where Q_{ij} represents the transition rate from the cell Γ_i to the cell Γ_j, $\mathcal{V}_i$ is the volume of the cell Γ_i, $\mathcal{S}_{ij}$ is the area of the intersecting surface between Γ_i and Γ_j, h_{ij} is the distance between the centers of the cells, π is the Boltzmann distribution:

$$\pi(x) = \frac{1}{Z}\exp\left(-\beta V(x)\right), \tag{18}$$

with $\beta = {}^1/_{k_B T}$ and Z is a normalization constant.

2.1 Project

Consider the two-dimensional double-well potential energy function defined on the space $\Gamma = \Gamma_x \times \Gamma_y \subset \mathbb{R}^2$

$$V(x,y) = 6\cos\left(m\,\arctan(y,x)\right) + 60\left(\sqrt{x^2+y^2}-1\right)^2, \tag{19}$$

with $m = 2$.

2.1.1 Part 1

Build the 2D-SqRA.

1. Discretize the space $[-2,2]\times[2,2]$ using a regular grid.
2. Build an adjacency matrix for a two-dimensional space.
3. Calculate the Boltzmann weight of each cell of the grid using eq. 18.
4. Build the rate matrix $\mathbf{Q}$ using eq. 17.
5. Given the matrix $\mathbf{Q}$, calculate and plot the first 5 eigenvectors and eigenvalues.

2.1.2 Part 2

The two-dimensional potential energy function defined in eq. 34 can be projected onto relevant coordinates $u(x,y)$. Consider the relevant coordinate

$$u(x,y) = x\,, \tag{20}$$

and the free energy profile

$$F_x(x) = -\frac{1}{\beta}\log\int_\Gamma \exp(-\beta V(x,y))\delta(u(x,y)-x)\,dydx \tag{21}$$

$$= -\frac{1}{\beta}\log\int_{\Gamma_y}\exp(-\beta V(x,y))\,dy\,, \tag{22}$$

where $\delta(u(x,y)-x)$ is the delta function, equal to 0 if $u(x,y)\neq x$, equal to 1 if $u(x,y)=x$.

1. Build the 1D-SqRA along the x-coordinate using the free energy profile obtained in eq. 22.
2. Compare the eigenvalues and eigenvectors of the two-dimensional model.

2.1.3 Part 3

A better relevant coordinate is the circumference defined by the function

$$r(x, y) = \sqrt{x^2 + y^2} = 1\,. \tag{23}$$

1. Find an expression for the free energy profile

$$F_r(r) \quad = \quad -\frac{1}{\beta} \log \int_\Gamma \exp(-\beta V(x, y)) \delta(r(x, y) - r)\, dydx\,. \tag{24}$$

2. Build the 1D-SqRA along the r(x,y)-coordinate using the free energy profile obtained in eq. 24.
3. Compare the eigenvalues and eigenvectors of the two-dimensional model.

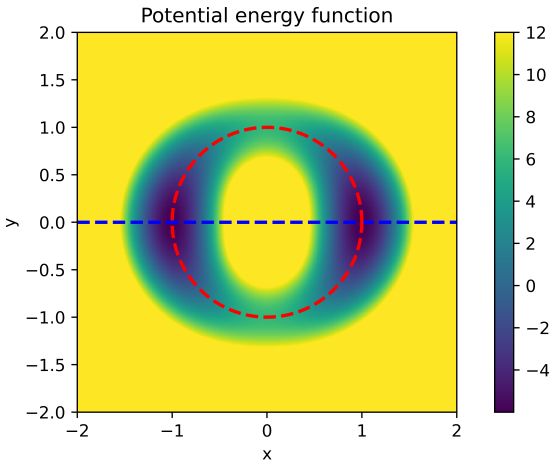

Figure 2: Potential energy function

3 PCCA+

In this project we will consider a non-reversible Markov process. One possibility is given by the Lotka-Volterra equations which model the evolution of the population of preditors and their prey:

$$\frac{dx}{dt} = \alpha x - \beta xy \tag{25}$$

$$\frac{dy}{dt} = \delta xy - \gamma y\,, \tag{26}$$

where x represents the number of prey, y represents the number of predators, α is the natural growth rate of the prey in the absence of predators, β is the rate at which predators destroy the prey, δ is the growth rate of predators per prey consumed, γ is the natural death rate of predators in the absence of prey.

The evolution is given by a deterministic ordinary differential equation. Thus, short-time trajectories from different starting points (different populations) can be generated.

Based on these trajectories and a fine discretization of the 2-dimensional state space (only positive numbers), a finite-dimensional matrix K^τ can be derived which represents the Koopman operator of this process.

3.1 Project

Here are the steps to be performed:

- The ordinary differential equation will be implemented. Trajectories will be generated. You will have to think about, how to transform the two dimensional starting and end points – (initial number of preditors, initial number of preys) and (final number of preditors, final number of preys) – into a set of micro states.
- Using the trajectories you will generate the matrix K^τ.
- The eigenvalue and eigenvector analysis of this matrix will lead to complex numbers. Plot these complex numbers in the 2-dimensional unit circle.
- Instead of the eigenvalue decomposition, compute the real Schur-decomposition of this matrix.
- Select 3 real Schur-vectors which span an invariant subspace of K^τ. Plot the simplex.
- Find a linear mapping from this simplex to the standard simplex.

4 ISOKANN

ISOKANN is an iterative algorithm in which a linear transformation S is applied to the Koopman operator $\mathcal{K}_\tau$ applied to an arbitrary function f_k:

$$f_{k+1} \quad = \quad S\mathcal{K}_\tau f_k \,, \tag{27}$$

After several iterations, the arbitrary function converges to a membership functions χ

$$\lim_{k\to\infty} f_{k+1} \quad = \quad \chi \,. \tag{28}$$

For a two-metastable system, S is defined by the shift-scale function

$$S\mathcal{K}_\tau f_k = \frac{\mathcal{K}_\tau f_k - \min\left(\mathcal{K}_\tau f_k\right)}{\max\left(\mathcal{K}_\tau f_k\right) - \min\left(\mathcal{K}_\tau f_k\right)} \,, \tag{29}$$

that guarantees that $f_k : \Gamma \to [0,1]$, and χ is one of the two membership functions χ_0 or χ_1. Note that, given χ_0, then $\chi_1 = 1 - \chi_0$, then the two membership functions span an invariant subspace of the Koopman operator.

4.1 Approximating the Koopman operator

In general, we do not know an analytical expression of the Koopman operator, nor a discretisation of it as a matrix. However, given a function f_t, it is possible to approximate the action of the operator on f_t as an expectation:

$$\begin{aligned} f_{t+\tau}(x) \quad &= \quad \mathcal{K}_\tau f_t(x) && (30)\\ &= \quad \mathbb{E}\left[f_t(x_{t+\tau})|x_t = x\right] && (31)\\ &\approx \quad \frac{1}{N}\sum_{n=1}^{N} f_t(x_{t+\tau,n}|x_t = x) \,, && (32) \end{aligned}$$

where $x_{t+\tau,n}$ are the final states of N trajectories starting at $x_t = x$, solutions of the equation of motion associated to the Koopman operator.

Then, within the ISOKANN algorithm, eq. 27 is rewritten as

$$f_{k+1}(x_0) \quad = \quad S\left(\frac{1}{N}\sum_{n=1}^{N} f_k(x_{\tau,n}|x_0 = x)\right) \,, \tag{33}$$

4.2 The choice of arbitrary function

The choice of the arbitrary function is crucial to the success of the algorithm. For simple problems with low dimensionality, a linear function may be sufficient and converges quickly to membership functions. For more complex problems, the use of interpolating functions is recommended. Here we suggest some possible solutions.

Radial Basis Functions Radial Basis Functions (RBFs) are a type of function used in various areas of numerical analysis and machine learning, particularly for interpolation, approximation, and classification tasks. The defining characteristic of RBFs is that their value depends only on the distance from a central point, or "center," rather than the specific coordinates of the input.

Feed-Forward Neural Network A Feed-Forward Neural Network (FNN) is the most basic type of artificial neural network. In an FNN, information moves in one direction-from input nodes, through hidden layers (if any), to output nodes. Each layer consists of a set of neurons, and each neuron in one layer is connected to every neuron in the next layer, but there are no cycles or loops in the network.

Radial Basis Function Networks Radial Basis Function Networks (RBFNs) are a type of artificial neural network that uses radial basis functions (RBFs) as activation functions. A Radial Basis Function (RBF) is a real-valued function whose value depends only on the distance from a center point, often referred to as the "radial distance." They are particularly well-suited for tasks such as interpolation, function approximation, and classification.

Graph Neural Network A Graph Neural Network (GNN) is a type of neural network designed to operate on graph-structured data. In a graph, data points (nodes) are connected by edges, which can represent relationships, interactions, or dependencies between the nodes. GNNs are designed to leverage this relational information to make predictions about nodes, edges, or the entire graph. GNNs work by iteratively updating node representations through message passing and aggregation mechanisms. During each iteration (or layer), each node receives information from its neighbors (nodes it's connected to), aggregates this information, and updates its own state.

4.3 Project

Consider the two-dimensional double-well potential energy function

$$V(x,y) = 6\cos\left(m\ \arctan(y,x)\right) + 60\left(\sqrt{x^2+y^2}-1\right)^2, \tag{34}$$

with $m = 2$ and defined on the space $\Gamma = \Gamma_x \times \Gamma_y \subset \mathbb{R}^2$.

4.3.1 Part 1

1. Implement the BBK integrator scheme to solve the Langevin dynamics.
2. Generate a first long trajectory.
3. Select a set of M random points (x_0, y_0) from the long trajectory.
4. From each starting point (x_0, y_0) generate N short trajectories of length τ. Store the final points (x_τ, y_τ).
5. Implement the ISOKANN algorithm. The algorithm should contains the following steps:
   (a) For each starting point (x_0, y_0), apply the arbitrary function f_k to the corresponding set of final points (x_τ, y_τ).
   (b) For each starting point (x_0, y_0), calculate the average and apply the transformation S as in eq. 33.
   (c) Given the points $f_{k+1}(x)$ obtained from eq. 33, train an interpolating function $\hat{f}_{k+1}$.
   (d) Restart the algorithm using $\hat{f}_{k+1}$ as arbitrary function.

4.3.2 Part 2

After several iterations, the function $\hat{f}_{k+1}$ will converge to a membership function $\chi(x)$.

1. Solve the linear regression problem

$$\min_{a_1,a_2} \|\mathcal{K}_\tau\chi(x) - a_1\chi(x) - a_2\|, \tag{35}$$

   where $\mathcal{K}_\tau\chi(x)$ can be approximated as in eq. 32.

2. Calculate the exit rate

$$\epsilon_1 = -\frac{1}{\tau}\log(a_1)\left(1 + \frac{a_2}{a_1 - 1}\right). \tag{36}$$

4.3.3 Part 3

Test the ISOKANN algorithm using different interpolating functions.