## Appendix A: Dynamical operators

The conformational dynamics of the molecule can be described by a Markov propagator $\mathcal{P}_\tau$ as

$$\rho_{t+\tau}(x) = (\mathcal{P}_\tau \rho_t)(x) \tag{A1}$$

with

$$(\mathcal{P}_\tau \rho_t)(x) \triangleq \int p(x_{t+\tau} = x | x_t = x') \rho_t(x') \mathrm{d}x' \tag{A2}$$

and

$$\mathcal{P}_\tau \pi = \pi \tag{A3}$$

where $\rho_t$ denotes the probability density function of $x_t$. According to the ergodicity and reversibility of $\mathcal{P}_\tau$, we can conclude that $\mathcal{P}_\tau$ is a compact and self-adjoint operator on the Hilbert space $\mathcal{L}^2_{\pi^{-1}} = \{f | \langle f, f \rangle_{\pi^{-1}} < \infty\}$ with inner product

$$\langle f, g \rangle_{\pi^{-1}} = \int f(x) g(x) \pi^{-1}(x) \mathrm{d}x \tag{A4}$$

and the probability density of $x_{t+\tau}$ can be expressed as

$$\rho_{t+\tau}(x) = \sum_{i=1}^{\infty} \lambda_i(\tau) \langle \rho_t, l_i \rangle_{\pi^{-1}} l_i(x) \tag{A5}$$

where $\lambda_i(\tau)$ is the $i$-th largest eigenvalue of $\mathcal{P}_\tau$, $l_i$ denotes the corresponding eigenfunction which satisfies $\langle l_i, l_j \rangle_{\pi^{-1}} = 1_{i=j}$. ($\lambda_1(\tau) \equiv 1 > \lambda_2(\tau)$ and $l_1 = \pi$ due to the uniqueness of the stationary distribution.)

Additionally, it is worth pointing out that the above conformational dynamics can also be described by using the weighted density

$$\rho_t^r = \pi^{-1} \rho_t \tag{A6}$$

and the transfer operator[1] $\mathcal{T}_\tau$ defined by

$$\begin{aligned} \rho^r_{t+\tau}(x) &= (\mathcal{T}_\tau \rho^r_t)(x) \\ &= \int \frac{\pi(x')}{\pi(x)} p(x_{t+\tau} = x | x_t = x') \\ &\quad \cdot \rho^r_t(x') \mathrm{d}x' \end{aligned} \tag{A7}$$

The transfer operator is a compact and self-adjoint operator on the Hilbert space $\mathcal{L}^2_\pi = \{f | \langle f, f \rangle_\pi < \infty\}$ with inner product

$$\langle f, g \rangle_\pi = \int f(x) g(x) \pi(x) \mathrm{d}x \tag{A8}$$

Table I: Comparison of different integral operators of the reversible Markov process $\{x_t\}$

| | Markov propagator | symmetrized propagator | Transfer operator |
|---|---|---|---|
| Notation | $\mathcal{P}_\tau$ | $\mathcal{S}_\tau$ | $\mathcal{T}_\tau$ |
| Integral kernel | $p\left(x_{t+\tau}=x \mid x_t=x'\right)$ | $\sqrt{\frac{\pi(x')}{\pi(x)}}p\left(x_{t+\tau}=x \mid x_t=x'\right)$ | $\frac{\pi(x')}{\pi(x)}p\left(x_{t+\tau}=x \mid x_t=x'\right)$ |
| Self-adjoint property | $\langle\mathcal{P}_\tau f,g\rangle_{\pi^{-1}}=\langle f,\mathcal{P}_\tau g\rangle_{\pi^{-1}}$ | $\langle\mathcal{S}_\tau f,g\rangle=\langle f,\mathcal{S}_\tau g\rangle$ | $\langle\mathcal{T}_\tau f,g\rangle_\pi=\langle f,\mathcal{T}_\tau g\rangle_\pi$ |
| Dynamical equation | $\rho_{t+\tau}=\mathcal{P}_\tau\rho_t$ with $\rho_t\left(x\right)=p\left(x_t=x\right)$ | $u_{t+\tau}=\mathcal{S}_\tau u$ with $u\left(x\right)=\rho_t\left(x\right)/\sqrt{\pi\left(x\right)}$ | $\rho^r_{t+\tau}=\mathcal{T}_\tau\rho^r_t$ with $\rho^r_t\left(x\right)=\rho_t\left(x\right)/\pi\left(x\right)$ |
| Eigenpairs | $\{(\lambda_i,l_i)\}_{i\geq1}$ with $\lambda_1=1$, $l_1=\pi$ | $\{(\lambda_i,\phi_i)\}_{i\geq1}$ with[a] $\phi_i\left(x\right)=l_i\left(x\right)/\sqrt{\pi\left(x\right)}$ | $\{(\lambda_i,r_i)\}_{i\geq1}$ with $r_i\left(x\right)=l_i\left(x\right)/\pi\left(x\right)$ |

[a] See proof in Appendix B

and its eigenpairs are $\{(\lambda_i\left(\tau\right),r_i)\}$ with $r_i=\pi^{-1}l_i$. Therefore, we can decompose the transfer equation of $\rho^r_t$ as

$$\rho^r_{t+\tau}\left(x\right)=\sum_{i=1}^{\infty}\lambda_i\left(\tau\right)\langle\rho^r_t,r_i\rangle_\pi r_i\left(x\right) \tag{A9}$$

which leads to the following approximation:

$$\rho^r_{t+\tau}\left(x\right)\approx\sum_{i=1}^{n}\lambda_i\left(\tau\right)\langle\rho^r_t,r_i\rangle_\pi r_i\left(x\right) \tag{A10}$$

Table I compares definitions and properties of the Markov propagator, transfer operator and symmetrized propagator investigated in this paper.

## Appendix B: Proofs of properties of symmetrized propagators

The purpose of this section is to show the properties of the symmetrized propagator $\mathcal{S}_\tau$ of $\{x_t\}$ stated in Section II under the following assumptions:

**Assumption 1.** $\{x_t\}$ *is a time-homogeneous, ergodic and reversible Markov process on space* $\mathbb{R}^d$.

**Assumption 2.** *The correlation density*

$$\begin{aligned} c_\tau\left(x', x\right) &\triangleq \lim_{t\to\infty} p\left(x_t = x', x_{t+\tau} = x\right) \\ &= \pi\left(x'\right) p\left(x_{t+\tau} = x | x_t = x'\right) \end{aligned} \tag{B1}$$

*is a continuous and positive function.*

**Assumption 3.** *The eigenvalues* $\{\lambda_i\}$ *of the Markov propagator* $\mathcal{P}_\tau$ *of* $\{x_t\}$ *is square summable, i.e.,* $\sum_i \lambda_i^2 < \infty$.

(Note that Assumption 3 is a sufficient condition for the compactness of $\mathcal{P}_\tau$, and is satisfied if the truncation approximation (6) is exact for a sufficiently large but finite $k$.)

Based on the above assumptions, we have the following lemma concerning the properties of $\mathcal{S}_\tau$ and $s_\tau$.

**Lemma 4.** *Let* $\mathcal{S}_\tau$ *be a symmetrized propagator of a Markov process* $\{x_t\}$ *which satisfies Assumptions 1–3, and let* $s_\tau$ *be the integral kernel of* $\mathcal{S}_\tau$*. Then*

1. $s_\tau\left(x', x\right) > 0$ and $s_\tau\left(x', x\right) = s_\tau\left(x, x'\right)$ for all $x', x \in \mathbb{R}^d$;
2. the eigenpair set of $\mathcal{S}_\tau$ is $\{(\lambda_i, \phi_i)\} = \left\{\left(\lambda_i, l_i/\sqrt{\pi_i}\right)\right\}$ with $(\lambda_i, l_i)$ being the $i$-th eigenpair of Markov propagator $\mathcal{P}_\tau$ of $\{x_t\}$;
3. $s_\tau$ is square integrable with $\|s_\tau\|_2 < \infty$.

*Proof.* □

*Part (1)*

*Proof.* According to the definition of $s_\tau$, we have

$$s_\tau\left(x',x\right)=\frac{c\left(x',x\right)}{\sqrt{\pi\left(x'\right)\pi\left(x\right)}} \tag{B2}$$

Then $s_\tau\left(x',x\right)>0$ and $s_\tau\left(x',x\right)=s_\tau\left(x,x'\right)$ hold for all $x',x$ due to Assumptions 1 and 2. □

*Part (2)*

*Proof.* In this part, we will prove the second conclusion of the theorem.

On the one hand, for an eigenpair $(\lambda,l)$ of $\mathcal{P}_\tau$, we can obtain that

$$\begin{aligned}\left(\mathcal{S}_\tau\phi\right)(x)&=\int\sqrt{\frac{\pi\left(x'\right)}{\pi\left(x\right)}}p\left(x_{t+\tau}=x|x_t=x'\right)\frac{l\left(x'\right)}{\sqrt{\pi\left(x'\right)}}\mathrm{d}x'\\&=\frac{1}{\sqrt{\pi\left(x\right)}}\int p\left(x_{t+\tau}=x|x_t=x'\right)l\left(x'\right)\mathrm{d}x'\\&=\frac{1}{\sqrt{\pi\left(x\right)}}\cdot\left(\mathcal{P}_\tau l\right)(x)\\&=\lambda\phi\left(x\right)\end{aligned} \tag{B3}$$

with $\phi\left(x\right)=l\left(x\right)/\sqrt{\pi\left(x\right)}$. Hence $(\lambda,\phi)$ is an eigenpair of $\mathcal{S}_\tau$.

On the other hand, if $(\lambda,\phi)$ is an eigenpair of $\mathcal{S}_\tau$ and $l\left(x\right)=\sqrt{\pi\left(x\right)}\phi\left(x\right)$, we have

$$\begin{aligned}\left(\mathcal{P}_\tau l\right)(x)&=\int p\left(x_{t+\tau}=x|x_t=x'\right)\sqrt{\pi\left(x'\right)}\phi\left(x'\right)\mathrm{d}x'\\&=\sqrt{\pi\left(x\right)}\cdot\left(\mathcal{S}_\tau\phi\right)(x)\\&=\lambda l\left(x\right)\end{aligned} \tag{B4}$$

Then $(\lambda,l)$ is an eigenpair of $\mathcal{P}_\tau$.

From the above results we can conclude that the eigenpair set of $\mathcal{S}_\tau$ is $\{(\lambda_i,\phi_i)\}$ with $\phi_i\left(x\right)=l_i\left(x\right)/\sqrt{\pi\left(x\right)}$. □

*Part (3)*

*Proof.* We now prove the third conclusion. Since the eigenpair set of $\mathcal{S}_\tau$ is $\{(\lambda_i,\phi_i)\}$ and $\{\lambda_i\}$ is square summable,

$$\left\|\mathcal{S}_\tau\right\|_{HS}^2=\sum_{i,j}\left\langle\phi_i,\mathcal{S}_\tau\phi_j\right\rangle^2=\sum_i\lambda_i^2<\infty \tag{B5}$$

where $\|\cdot\|_{HS}$ denotes the Hilbert-Schmidt norm. Therefore $\mathcal{S}_\tau$ is a Hilbert-Schmidt operator[2] on $\mathcal{L}^2$, and the integral kernel $s_\tau$ is a square integrable function according to the kernel theorem of Hilbert-Schmidt operators[3]. $\square$

## Appendix C: Proof of Theorem 1

According to the Hilbert-Schmidt theorem[4], an integral operator $\mathcal{S}_\tau$ can be decomposed as

$$\mathcal{S}_\tau u = \sum_{i=1}^{\infty} \mu_i \langle u, \phi_i \rangle \phi_i \tag{C1}$$

if $\mathcal{S}_\tau$ is compact and its kernel function $s_\tau$ is symmetric, where $|\mu_i|$ is monotonically non-increasing. Further, if $s_\tau$ is a positive function and the spectral radius of $\mathcal{S}_\tau$ is 1, we can conclude that $\mu_1 = 1$ and $\phi_1(x)$ is a positive function according to Krein-Rutman theorem[5].

Let

$$c_\tau(x', x) = \phi_1(x') s_\tau(x', x) \phi_1(x) \tag{C2}$$

It can be seen that $c_\tau(x', x)$ is a positive and symmetric function, and

$$\int c_\tau(x', x) \,\mathrm{d}x = \phi_1(x')^2 \tag{C3}$$

$$\iint c_\tau(x', x) \,\mathrm{d}x \mathrm{d}x' = \langle \phi_1, \phi_1 \rangle = 1 \tag{C4}$$

Therefore we can construct a Markov process $\{z_0, z_\tau, \ldots\}$ such that

$$p(z_t = x', z_{t+\tau} = x) = c_\tau(x', x) \tag{C5}$$

and the symmetrized propagator of the Markov process is $\mathcal{S}_\tau$.

## Appendix D: Universal approximation using Gaussian Markov transition models

In this section we show the following theorem which implies the universal approximation capability of the proposed GMTMs.

**Theorem 5.** *Let $s_\tau$ be the symmetrized transition density of a Markov process $\{x_t\}$ which satisfies Assumptions 1–3. Then, given any $\epsilon > 0$, there exists a GMTM with symmetrized transition density $\hat{s}$ such that*

$$\|\hat{s} - s_\tau\|_2 \leq \epsilon \tag{D1}$$

*Proof.* According to Lemma 4, we can conclude that the symmetrized transition density $s_\tau$ of $\{x_t\}$ is a square integrable and symmetric function under Assumptions 1–3. Then, by the universal approximation theorem of radial-basis-function networks[6], for any $\epsilon_0 > 0$, there exists a GMM

$$s^{(1)}\left(x', x\right) = \sum_{i=1}^{m} v_i \mathcal{N}\left(x' | \mu_i', \sigma^2 I\right) \mathcal{N}\left(x | \mu_i, \sigma^2 I\right) \tag{D2}$$

such that

$$\left\| s_\tau - s^{(1)} \right\|_2 \leq \epsilon_0 \tag{D3}$$

where $v_i \geq 0$, $\sigma > 0$.

Define

$$s^{(2)}\left(x', x\right) = \frac{1}{2}\left(s^{(1)}\left(x', x\right) + s^{(1)}\left(x, x'\right)\right) \tag{D4}$$

We can construct a GMTM with symmetrized propagator $\hat{\mathcal{S}}$ and symmetrized transition density $\hat{s}$ by

$$\begin{aligned} \hat{s}\left(x', x\right) &= \frac{1}{Z} s^{(2)}\left(x', x\right) \\ &= \frac{1}{Z} \chi\left(x'\right)^\intercal W \chi\left(x\right) \end{aligned} \tag{D5}$$

where

$$W = \begin{bmatrix} & & & \frac{v_1}{2} & & \\ & & & & \ddots & \\ & & & & & \frac{v_m}{2} \\ \frac{v_1}{2} & & & & & \\ & \ddots & & & & \\ & & \frac{v_m}{2} & & & \end{bmatrix} \tag{D6}$$

$$\begin{aligned} \chi\left(x\right) = \Big(&\mathcal{N}\left(x | \mu_1', \sigma^2 I\right), \ldots, \mathcal{N}\left(x | \mu_m', \sigma^2 I\right), \\ &\mathcal{N}\left(x | \mu_1, \sigma^2 I\right), \ldots, \mathcal{N}\left(x | \mu_m, \sigma^2 I\right)\Big)^\intercal \end{aligned} \tag{D7}$$

and the normalizing constant $Z$ enforces that the spectral radius of the symmetrized propagator defined by $\hat{s}$ is 1.

According to (D3) and Theorem 3 in Ref.[7], we have

$$\begin{aligned}\left\|s_\tau - s^{(2)}\right\|_2 &= \left\|\frac{1}{2}\left(s_\tau - s^{(1)}\right) + \frac{1}{2}\left(s_\tau - 2s^{(2)} + s^{(1)}\right)\right\|_2 \\ &\leq \frac{1}{2}\left\|s_\tau - s^{(1)}\right\|_2 + \frac{1}{2}\left\|s_\tau - 2s^{(2)} + s^{(1)}\right\|_2 \\ &= \left\|s_\tau - s^{(1)}\right\|_2 \\ &\leq \epsilon_0 \end{aligned} \tag{D8}$$

and

$$|Z-1| \leq \left\|s_\tau - s^{(2)}\right\|_2 \leq \epsilon_0 \tag{D9}$$

Therefore,

$$\begin{aligned}\|s_\tau - \hat{s}\|_2 &= \left\|s_\tau - \frac{1}{Z}s^{(2)}\right\|_2 \\ &= \left\|\frac{1}{Z}\left(s_\tau - s^{(2)}\right) + \left(1 - \frac{1}{Z}\right)s_\tau\right\|_2 \\ &\leq \frac{1}{Z}\epsilon_0 + \left|1 - \frac{1}{Z}\right|\|s_\tau\|_2 \\ &\leq \left(1 + \|s_\tau\|_2\right)\epsilon_0 + o\left(\epsilon_0\right) \end{aligned} \tag{D10}$$

as $\epsilon_0 \to 0$, which establishes the theorem. $\square$

## Appendix E: Expression of matrix $B$

For any $i, j$, the product $\mathcal{N}\left(x|\mu_i, \Sigma_i\right) \cdot \mathcal{N}\left(x|\mu_j, \Sigma_j\right)$ can be written as

$$\mathcal{N}\left(x|\mu_i, \Sigma_i\right) \cdot \mathcal{N}\left(x|\mu_j, \Sigma_j\right) = a\mathcal{N}\left(x|\mu_{(i,j)}, \Sigma_{(i,j)}\right) \tag{E1}$$

where

$$\Sigma_{(i,j)} = \left(\Sigma_i^{-1} + \Sigma_j^{-1}\right)^{-1} \tag{E2}$$

$$\mu_{(i,j)} = \Sigma_{(i,j)}\left(\Sigma_i^{-1}\mu_i + \Sigma_j^{-1}\mu_j\right)^{-1} \tag{E3}$$

and $a$ is a constant independent of $x$.

Since the matrix $B$ is defined by

$$B = \int \chi(x)\, \chi(x)^{\intercal}\, \mathrm{d}x \tag{E4}$$

with

$$\chi(x) = \left(\mathcal{N}(x|\mu_1, \Sigma_1), \ldots, \mathcal{N}(x|\mu_m, \Sigma_m)\right)^{\intercal} \tag{E5}$$

the $(i, j)$-th element of $B$ can be calculated by

$$\begin{aligned} B_{ij} &= \int \mathcal{N}(x|\mu_i, \Sigma_i) \cdot \mathcal{N}(x|\mu_j, \Sigma_j)\, \mathrm{d}x \\ &= \int a\mathcal{N}\left(x|\mu_{(i,j)}, \Sigma_{(i,j)}\right) \mathrm{d}x \\ &= a \end{aligned} \tag{E6}$$

with

$$a = \frac{\mathcal{N}(0|\mu_i, \Sigma_i) \cdot \mathcal{N}(0|\mu_j, \Sigma_j)}{\mathcal{N}\left(0|\mu_{(i,j)}, \Sigma_{(i,j)}\right)} \tag{E7}$$

## Appendix F: Calculation of $p(\{x_{k\tau}\}|\mathcal{I}, s_\tau)$ in Section IV

For convenience of notation and expression, here we let

$$\begin{aligned} f_{\mathcal{N}} = &\prod_{k=1}^{K} \mathcal{N}\left(x_{(k-1)\tau}|\mu_{I_k}, \Sigma_{I_k}\right) \\ &\cdot \prod_{k=1}^{K} \mathcal{N}\left(x_{k\tau}|\mu_{J_k}, \Sigma_{J_k}\right) \end{aligned} \tag{F1}$$

and define sufficient statistics $\bar{n}_{ij}, \bar{n}_i, \bar{v}_i, \bar{V}_i$ as follows:

$$
\begin{aligned}
\bar{n}_{ij} &\triangleq |\{k|(I_k, J_k) = (i,j)\}| \\
&= \sum_k 1_{(I_k,J_k)=(i,j)} \\
\bar{n}_i &\triangleq |\{k|I_k = i\}| + |\{k|J_k = i\}| \\
&= \sum_j \bar{n}_{ij} + \bar{n}_{ji} \\
\bar{v}_i &\triangleq \sum_{k\in\{k|I_k=i\}} x_{(k-1)\tau} + \sum_{k\in\{k|J_k=i\}} x_{k\tau} \\
\bar{V}_i &\triangleq \sum_{k\in\{k|I_k=i\}} x_{(k-1)\tau} x_{(k-1)\tau}^{\intercal} \\
&\quad + \sum_{k\in\{k|J_k=i\}} x_{k\tau} x_{k\tau}^{\intercal}
\end{aligned} \tag{F2}
$$

Then $f_{\mathcal{N}}$ can be calculated by

$$
\begin{aligned}
f_{\mathcal{N}} &= \prod_{k=1}^{K} \mathcal{N}\left(x_{(k-1)\tau}|\mu_{I_k}, \Sigma_{I_k}\right) \\
&\quad \cdot \prod_{k=1}^{K} \mathcal{N}\left(x_{k\tau}|\mu_{J_k}, \Sigma_{J_k}\right) \\
&= \prod_i \left( \prod_{x\in\{x_{(k-1)\tau}|I_k=i\}} \mathcal{N}\left(x|\mu_i, \Sigma_i\right) \right. \\
&\quad \left. \cdot \prod_{x\in\{x_{k\tau}|J_k=i\}} \mathcal{N}\left(x|\mu_i, \Sigma_i\right) \right) \\
&= \prod_i (2\pi)^{-\frac{\bar{n}_i d}{2}} |\Sigma_i|^{-\frac{\bar{n}_i}{2}} \\
&\quad \exp\left(-\frac{1}{2}\mathrm{tr}\left(\Sigma_i^{-1}\left(\bar{V}_i + \bar{n}_i\mu_i\mu_i^{\intercal} - 2\bar{v}_i\mu_i^{\intercal}\right)\right)\right)
\end{aligned} \tag{F3}
$$

in time $\mathcal{O}\left(md^2\right)$ if all the sufficient statistics are given, and $p\left(\{x_{k\tau}\}|\mathcal{I}, s_\tau\right)$ can be calculated by using (F3) and

$$
p\left(\{x_{k\tau}\}|\mathcal{I}, s_\tau\right) = Z^{-K}\rho_0\left(x_0\right)\frac{\phi_1\left(x_{K\tau}\right)}{\phi_1\left(x_0\right)} f_{\mathcal{N}} \tag{F4}
$$

## Appendix G: Implementation details of the Gibbs sampler for Bayesian estimation

### 1. Initialization

Note that each covariance matrix $\Sigma_n$ $(n = 1, \ldots m)$ in the GMTM can be expressed as

$$\Sigma_n^{-1} = \Psi_n^{\intercal} \Psi_n \tag{G1}$$

where $\Psi_n = [\psi_{n,ij}]$ is a lower-triangular matrix with positive diagonal elements. We can define the parameter vector $\theta$ by stacking all the following elements:

$$\begin{gathered}
\{\log W_{ij} - \log W_{mm} | i \geq j, (i,j) \neq (m,m)\} \\
\{\mu_n | n = 1, \ldots m\} \\
\{\log \psi_{n,ii} | n = 1, \ldots m, i = 1, \ldots, d\} \\
\{\psi_{n,ij} | n = 1, \ldots m, i > j\}
\end{gathered} \tag{G2}$$

It can be seen that a given $\theta$ can uniquely determine a GMTM and the feasible space of $\theta$ is the whole vector space, which means we can perform the Bayesian inference of the GMTM by sampling $\theta$ without any equality or inequality constraint.

Moreover, the prior distribution of $\theta$ is chosen based on the following independent uninformative priors

$$\begin{aligned}
p(W) &\propto 1 \\
p(\mu_n) &\propto 1 \\
p(\Psi_n) &\propto \prod_{i=1}^{d} \psi_{n,ii}^{-i}, \quad \text{for } n = 1, \ldots, m
\end{aligned} \tag{G3}$$

where $p(\Psi_n)$ represents the right-Haar prior[8] for covariance matrices. According to (G3), we have

$$p(\theta) \propto \prod_{i \geq j} W_{ij}(\theta) \cdot \prod_{n=1}^{m} \prod_{i=1}^{d} \psi_{n,ii}^{1-i}(\theta) \tag{G4}$$

### 2. Sampling of latent variables

From the Bayes' rule, we can obtain that

$$p\left(\mathcal{I}|\{x_{k\tau}\}, s_\tau(\theta)\right) \propto \prod_{k=1}^{K} W_{I_k J_k} \mathcal{N}\left(x_{(k-1)\tau}|\mu_{I_k}, \Sigma_{I_k}\right) \cdot \mathcal{N}\left(x_{k\tau}|\mu_{J_k}, \Sigma_{J_k}\right) \tag{G5}$$

This implies that $(I_k, J_k)$ for different $k$ are conditionally independent of each other given $\{x_{k\tau}\}$ and $\theta$, i.e.,

$$p\left(\mathcal{I}|\{x_{k\tau}\}, s_\tau(\theta)\right) = \prod_{k=1}^{K} p\left(I_k, J_k|x_{(k-1)\tau}, x_{k\tau}, s_\tau(\theta)\right) \tag{G6}$$

with

$$p\left(I_k, J_k|x_{(k-1)\tau}, x_{k\tau}, s_\tau(\theta)\right) \propto W_{I_k J_k} \mathcal{N}\left(x_{(k-1)\tau}|\mu_{I_k}, \Sigma_{I_k}\right) \cdot \mathcal{N}\left(x_{k\tau}|\mu_{J_k}, \Sigma_{J_k}\right) \tag{G7}$$

Then the full condition posterior distribution of $\{(I_k, J_k)\}$ can be exactly sampled by using (G7).

### 3. Sampling of $\theta$

Considering that the full conditional posterior distribution $p\left(\theta|\mathcal{I}, \{x_{k\tau}\}\right)$ of $\theta$ is an intractable one, we utilize the Metropolis-Hastings step to iteratively sample $\theta$, which generates a candidate sample $\theta' \sim \mathcal{N}\left(\theta'|\theta, \eta^2 I\right)$ by using the current sample of $\theta$ and accepts $\theta'$ as a new sample with probability

$$\begin{aligned} & p_{\text{acc}}\left(\theta, \theta'\right) \\ & = \min\left\{1, \frac{p\left(\theta'\right) p\left(\{x_{k\tau}\}|\mathcal{I}, s_\tau(\theta')\right)}{p(\theta)\, p\left(\{x_{k\tau}\}|\mathcal{I}, s_\tau(\theta)\right)}\right\} \end{aligned} \tag{G8}$$

Here $\eta > 0$ denotes the sampling step, $I$ represents the identity matrix, and the expression of $p\left(\{x_{k\tau}\}|\mathcal{I}, s_\tau(\theta)\right)$ is given by (F3) and (F4).

## 4. Detailed description of the Gibbs sampler

Summarizing all the above analysis and discussion, we now provide a step-by-step description of the Gibbs sampler for GMTMs in Algorithm 1. In applications of this paper, we set $M' = M = 10^4$, and the sampling step $\eta$ is determined by an adaptive algorithm proposed in Ref.[9] such that the average acceptance probability is close to 0.234, which is the optimal acceptance rate for random walk based Metropolis-Hastings sampling steps.

---

Algorithm 1: Gibbs sampler for GMTMs

---

Choose an initial value $\theta^{(0)}$ of the parameter vector $\theta$ arbitrarily.
**for** $l = 1$ to $M' + M$ **do**
    **for** $k = 1$ to $K$ **do**
        Draw $(I_k, J_k)$ according to (G7).
    **end for**
    Calculate the sufficient statistics $\bar{n}_{ij}, \bar{n}_i, \bar{v}_i, \bar{V}_i$ for $i, j = 1, \ldots, m$ by (F2).
    **for** $\ell_m = 1$ to $L_m$ **do**
        Draw $\theta' \sim \mathcal{N}\left(\theta' | \theta, \eta^2 I\right)$.
        Calculate the acceptance probability $p_{\mathrm{acc}}$ by using (G8), (G4), (F3) and (F4).
        Set $\theta := \theta'$ with probability $p_{\mathrm{acc}}$.
    **end for**
    Let $\theta^{(\ell)} = \theta$.
**end for**
**return** $\theta^{(M'+1)}, \ldots, \theta^{(M'+M)}$

---

## Appendix H: Implementation details of the EM algorithm for maximum likelihood estimation

### 1. E-step

From (25), (F3) and (F4), we can get

$$\begin{aligned}\log p\left(\mathcal{I},\{x_{k\tau}\}|s_\tau\right) &= \log p\left(\mathcal{I}|s_\tau\right)+\log p\left(\{x_{k\tau}\}|\mathcal{I},s_\tau\right)\\ &= \sum_{k=1}^{K}\log W_{I_kJ_k}\\ &\quad +\log\left(Z^{-K}\rho_0\left(x_0\right)\frac{\phi_1\left(x_{K\tau}\right)}{\phi_1\left(x_0\right)}f_{\mathcal{N}}\right)\\ &= \log\left(Z^{-K}\rho_0\left(x_0\right)\frac{\phi_1\left(x_{K\tau}\right)}{\phi_1\left(x_0\right)}\right)\\ &\quad +\sum_{i,j}\bar{n}_{ij}\log W_{ij}\\ &\quad -\frac{1}{2}\sum_i\bar{n}_i\left(d\log\left(2\pi\right)+\log|\Sigma_i|\right)\\ &\quad -\frac{1}{2}\sum_i\mathrm{tr}\left(\Sigma_i^{-1}\bar{V}_i\right)\\ &\quad -\frac{1}{2}\sum_i\bar{n}_i\mu_i^\intercal\Sigma_i^{-1}\mu_i\\ &\quad +\sum_i\mu_i^\intercal\Sigma_i^{-1}\bar{v}_i\end{aligned}\qquad\text{(H1)}$$

where $\bar{n}_{ij},\bar{n}_i,\bar{v}_i,\bar{V}_i$ are sufficient statistics of $\mathcal{I}$ and $\{x_{k\tau}\}$ defined in (F2). Furthermore, according to (G6) and (G7), we have

$$q^{(\ell)}\left(\mathcal{I}\right)=p\left(\mathcal{I}|\{x_{k\tau}\},s_\tau\left(\theta^{(\ell-1)}\right)\right)=\prod_{k=1}^{K}q^{(\ell)}\left(I_k,J_k\right)\qquad\text{(H2)}$$

where $q^{(\ell)}\left(I_k,J_k\right)$ is a probability distribution of $\left(I_k,J_k\right)$ and given by

$$\begin{aligned}q^{(\ell)}\left(I_k,J_k\right)\propto\ & W_{I_kJ_k}^{(\ell-1)}\mathcal{N}\left(x_{(k-1)\tau}|\mu_{I_k}^{(\ell-1)},\Sigma_{I_k}^{(\ell-1)}\right)\\ &\cdot\mathcal{N}\left(x_{k\tau}|\mu_{J_k}^{(\ell-1)},\Sigma_{J_k}^{(\ell-1)}\right)\end{aligned}\qquad\text{(H3)}$$

Combining (H1) and (H2) leads to

$$
\begin{aligned}
Q\left(\theta|\theta^{(\ell-1)}\right) = \log & \left(Z^{-K}\rho_0\left(x_0\right)\frac{\phi_1\left(x_{K\tau}\right)}{\phi_1\left(x_0\right)}\right) \\
& + \sum_{i,j}\mathbb{E}_{q^{(\ell)}}\left[\bar{n}_{ij}\right]\log W_{ij} \\
& - \frac{1}{2}\sum_i \mathbb{E}_{q^{(\ell)}}\left[\bar{n}_i\right]\left(d\log\left(2\pi\right)+\log\left|\Sigma_i\right|\right) \\
& - \frac{1}{2}\sum_i \mathrm{tr}\left(\Sigma_i^{-1}\mathbb{E}_{q^{(\ell)}}\left[\bar{V}_i\right]\right) \\
& - \frac{1}{2}\sum_i \mathbb{E}_{q^{(\ell)}}\left[\bar{n}_i\right]\mu_i^{\intercal}\Sigma_i^{-1}\mu_i \\
& + \sum_i \mu_i^{\intercal}\Sigma_i^{-1}\mathbb{E}_{q^{(\ell)}}\left[\bar{v}_i\right]
\end{aligned}
\tag{H4}
$$

with

$$
\begin{aligned}
\mathbb{E}_{q^{(\ell)}}\left[\bar{n}_{ij}\right] &= \sum_k q^{(\ell)}\left(i,j\right) \\
\mathbb{E}_{q^{(\ell)}}\left[\bar{n}_{i}\right] &= \sum_j \mathbb{E}_{q^{(\ell)}}\left[\bar{n}_{ij}\right]+\mathbb{E}_{q^{(\ell)}}\left[\bar{n}_{ji}\right] \\
\mathbb{E}_{q^{(\ell)}}\left[\bar{v}_{i}\right] &= \sum_{k,j} q^{(\ell)}\left(i,j\right)x_{(k-1)\tau} \\
&\quad + q^{(\ell)}\left(j,i\right)x_{k\tau} \\
\mathbb{E}_{q^{(\ell)}}\left[\bar{V}_{i}\right] &= \sum_{k,j} q^{(\ell)}\left(i,j\right)x_{(k-1)\tau}x_{(k-1)\tau}^{\intercal} \\
&\quad + q^{(\ell)}\left(j,i\right)x_{k\tau}x_{k\tau}^{\intercal}
\end{aligned}
\tag{H5}
$$

## 2. M-step

Note that $Q\left(\theta|\theta^{(\ell-1)}\right)$ is a highly nonlinear and nonconvex function with respect to $\theta$, so we select an evolutionary algorithm called "$(1+1)$ evolution strategy with $1/5$ success rule"[10] to search the optimal solution to $\max_\theta Q\left(\theta|\theta^{(\ell-1)}\right)$. The pseudo-code of the optimization algorithm is given by Algorithm 2.

## 3. Detailed description of the EM algorithm

Based on the above discussion, the complete EM algorithm for the ML estimation of GMTMs is shown in Algorithm 3. We set $M_{\mathrm{em}} = 500$ and $M_{\mathrm{es}} = 2000$ in applications

Algorithm 2: $(1+1)$ evolution strategy with $1/5$ success rule for solving $\max_\theta Q(\theta|\theta^{(\ell-1)})$

Calculate $\mathbb{E}_{q^{(\ell)}}[\bar{n}_{ij}]$, $\mathbb{E}_{q^{(\ell)}}[\bar{n}_i]$, $\mathbb{E}_{q^{(\ell)}}[\bar{v}_i]$ and $\mathbb{E}_{q^{(\ell)}}\left[\bar{V}_i\right]$ by (H5) for $i, j = 1, \ldots, m$.
Set $\theta^{(\ell)} := \theta^{(\ell-1)}$ and $\eta := 1$.
**for** $\ell_{\mathrm{es}} = 1$ to $M_{\mathrm{es}}$ **do**
    Draw $\theta' \sim \mathcal{N}\left(\theta'|\theta^{(\ell)}, \eta^2 I\right)$.
    **if** $Q\left(\theta'|\theta^{(\ell-1)}\right) \geq Q\left(\theta^{(\ell)}|\theta^{(\ell-1)}\right)$ **then**
        Set $\theta^{(\ell)} := \theta'$ and $\eta := 1.5\eta$.
    **else**
        Set $\eta := 1.5^{-\frac{1}{4}}\eta$.
    **end if**
**end for**
**return** $\theta^{(\ell)}$

of this paper.

Algorithm 3: EM algorithm for GMTMs

Choose an initial value $\theta^{(0)}$ of the parameter vector $\theta$ arbitrarily.
**for** $\ell = 1$ to $M_{\mathrm{em}}$ **do**
    Calculate $q^{(\ell)}(i, j)$ by (H3) for $i, j = 1, \ldots, m$.
    Solve $\theta^{(\ell)} = \arg\max_\theta Q\left(\theta|\theta^{(\ell-1)}\right)$ by Algorithm 2.
**end for**
**return** $\theta^{(M_{\mathrm{em}})}$

## Appendix I: Symmetric diffusion densities and spectral components of Markov state models

It is interesting to point out that the widely used MSM can also be interpreted as a *Markov transition model* with piecewise basis functions. Consider an MSM of the reversible process $\{x_t\}$ with discrete bins $\{A_1, \ldots, A_m\}$ and transition matrix $P = [P_{ij}] = [\Pr(x_{t+\tau} \in A_j | x_t \in A_i)]$, and suppose that we are given the stationary density $\pi(x)$ of $\{x_t\}$. The transition density can be expressed as

$$\begin{aligned} p(x_{t+\tau} = x | x_t = x') = & \sum_{i,j} 1_{A_i}(x') 1_{A_j}(x') \\ & \cdot \Pr(x_{t+\tau} \in A_j | x_t \in A_i) \\ & \cdot p(x_{t+\tau} = x | x_{t+\tau} \in A_j) \\ = & \sum_{i,j} 1_{A_i}(x') 1_{A_j}(x) \\ & \cdot P_{ij} \frac{\pi(x)}{\pi_j} \end{aligned} \tag{I1}$$

where $\pi(x)$ denotes the stationary probability density of $\{x_t\}$, $\pi_i = \int_{A_i} \pi(x) \, \mathrm{d}x$ represents the stationary probability of $A_i$ and $1_{A_i}(x)$ denotes the indicator function of $A_i$. Then the symmetrized transition density $s_\tau(x', x)$ of this MSM can be written in the form of a Markov transition model as

$$\begin{aligned} s_\tau(x', x) = & \sum_{i,j} \frac{P_{ij}}{\pi_j} \cdot \left(1_{A_i}(x') \sqrt{\pi(x')}\right) \\ & \cdot \left(1_{A_j}(x) \sqrt{\pi(x)}\right) \\ = & \chi(x)^\intercal W \chi(x) \end{aligned} \tag{I2}$$

with

$$\begin{aligned} W &= [W_{ij}] = P\Pi^{-1} \\ \chi(x) &= [\chi_i(x)] = [1_{A_i}(x) \sqrt{\pi(x)}] \end{aligned} \tag{I3}$$

and

$$\Pi = \mathrm{diag}(\pi_1, \ldots, \pi_m) \tag{I4}$$

Due to the reversibility of $\{x_t\}$, $W$ is a symmetric matrix with

$$\begin{aligned} W_{ij} &= \frac{P_{ij}}{\pi_j} \\ &= \frac{\Pr(x_t \in A_i, x_{t+\tau} \in A_j)}{\pi_i \pi_j} \\ &= \frac{\Pr(x_t \in A_j, x_{t+\tau} \in A_i)}{\pi_j \pi_i} \\ &= W_{ji} \end{aligned} \tag{I5}$$

Therefore, the spectral component extraction process developed in Section IIIB can be utilized to get the spectral components of the MSM.

Note the inner product matrix of $\chi(x)$ is

$$B = \int \chi(x)\,\chi(x)^\intercal \,\mathrm{d}x = \Pi \tag{I6}$$

Then eigenpairs $\{(\lambda_i, b_i)\}$ of $BW$ are solutions to problem

$$\begin{aligned} b^\intercal BW &= \lambda b^\intercal \\ \Rightarrow (b^\intercal \Pi)\,P &= \lambda\,(b^\intercal \Pi) \end{aligned} \tag{I7}$$

This implies that the eigenvalues of the MSM are exactly equal to that of the transition matrix $P$ and the eigenfunctions are

$$\begin{aligned} r_i(x) = \frac{l_i(x)}{\pi(x)} &= \frac{\phi_i(x)}{\sqrt{\pi(x)}} \\ &= \frac{b_i^\intercal \chi(x)}{\sqrt{\pi(x)}} \\ &= \frac{\tilde{l}_i^\intercal \Pi^{-1} \chi(x)}{\sqrt{\pi(x)}} \\ &= \frac{\tilde{r}_i^\intercal \chi(x)}{\sqrt{\pi(x)}} \\ &= \sum_j \tilde{r}_{ij} 1_{A_j}(x) \end{aligned} \tag{I8}$$

where $\tilde{l}_i$ and $\tilde{r}_i$ denote the $i$-th left and right eigenvectors of $P$ with $\tilde{r}_i = \tilde{l}_i^\intercal \Pi^{-1}$ and $\tilde{r}_i^\intercal \Pi \tilde{r}_i = 1$, and $\tilde{r}_{ij}$ denotes the $j$-th element of $\tilde{r}_i$. This conclusion is consistent with that in literature[11,12] on MSMs.

From the above, it can be seen that MSMs approximate right eigenfunctions $\{r_i\}$ by step functions and are unable to provide the estimates of left eigenfunctions $\{l_i\}$ unless the stationary density $\pi$ is available or can be given by some density estimation algorithm. In contrast with MSMs, the proposed GMTMs can provide more smooth and flexible estimates of both left and right eigenfunctions by using Gaussian basis functions without the assumption of known $\pi$. Furthermore, the choice of discretization is still an open problem for MSMs and there is no systematic algorithm is available for optimizing discrete bins except trial-and-error, but for the GMTMs, the parameters of basis functions can be adaptively adjusted in the Bayesian framework.

*Remark* 6. In this paper, we calculate the stationary density $\pi$ a given MSM by

$$\pi(x) = \sum_i \frac{\pi_i}{\int_{A_i} \pi^{\mathrm{KDE}}(x')\,\mathrm{d}x'} 1_{A_i}(x)\,\pi^{\mathrm{KDE}}(x) \tag{I9}$$

such that $\pi_i = \int_{A_i} \pi(x)\,\mathrm{d}x$ holds for each bin $A_i$, where $\pi^{\mathrm{KDE}}(x)$ denotes the kernel density estimator[13] of the global distribution of simulation data.

## REFERENCES


[1] C. Schütte and M. Sarich, *Metastability and Markov State Models in Molecular Dynamics: Modeling, Analysis, Algorithmic Approaches*, Vol. 24 (AMS, 2013).

[2] G. J. Murphy, *$C^*$-algebras and operator theory* (Academic press San Diego, 1990).

[3] J. L. Martınez-Morales, J. Math. Anal. Appl. **279**, 71 (2003).

[4] M. Renardy and R. C. Rogers, *An introduction to partial differential equations*, Vol. 4 (Springer, 2004).

[5] Y. Du, *Order structure and topological methods in nonlinear partial differential equations* (World Scientific, 2006).

[6] J. Park and I. W. Sandberg, Neural Comput. **3**, 246 (1991).

[7] L. Elsner and S. Friedland, P. Am Math Soc. **123**, 2511 (1995).

[8] D. Sun and J. Berger, in *Proceedings of ISBA 8th World Meeting on Bayesian Statistics* (2007) pp. 1–38.

[9] B. Shaby and M. T. Wells, "Exploring an adaptive metropolis algorithm," Tech. Rep. (Department of Statistical Science, Duke University, 2010).

[10] A. Auger, in *Proceedings of the 11th Annual Conference Companion on Genetic and Evolutionary Computation Conference* (ACM, 2009) pp. 2447–2452.

[11] M. Sarich, F. Noé, and C. Schütte, SIAM Multiscale Model. Simul. **8**, 1154 (2010).

[12] J.-H. Prinz, H. Wu, M. Sarich, B. Keller, M. Senne, M. Held, J. D. Chodera, C. Schütte, and F. Noé, J. Chem. Phys. **134**, 174105 (2011).

[13] B. W. Silverman, *Density estimation for statistics and data analysis*, Vol. 26 (CRC press, 1986).