5 Discretization of Transfer Operators

If we want to identify metastable subsets we have to compute certain eigenfunctions of the propagator P_τ. In the following we describe the discretization procedure of the eigenvalue problem $P_\tau v = \lambda v$. Throughout this section we assume that P_τ satisfies the conditions (C1) and (C2) defined in Section 3.2. Part of this section follows from [68, 69].

5.1 Galerkin Discretization

Let $\mathcal{D} = \{D_1, \ldots, D_n\}$ denote a decomposition of the state space and define the associated finite dimensional ansatz space by $\mathcal{V}_n = \text{span}\{\mathbf{1}_{D_1}, \ldots, \mathbf{1}_{D_n}\}$. Then, the **Galerkin projection** $\Pi_n : L^1(\mu) \to \mathcal{V}_n$ of $v \in L^1(\mu)$ is defined by

$$\Pi_n v \;=\; \sum_{k=1}^{n} \frac{\langle v, \mathbf{1}_{D_k}\rangle_\mu}{\langle \mathbf{1}_{D_k}, \mathbf{1}_{D_k}\rangle_\mu}\, \mathbf{1}_{D_k},$$

where $\langle \cdot, \cdot \rangle_\mu$ is the duality bracket between $L^1(\mu)$ and $L^\infty(\mu)$. The resulting discretized propagator $\Pi_n P_\tau \Pi_n$ induces an approximate eigenvalue problem $\Pi_n P_\tau \Pi_n v = \lambda \Pi_n v$ in $\mathcal{V}_n$. Using $v = \sum_{k=1}^n \nu_k \mathbf{1}_{D_k}$, the discretized eigenvalue problems reads in coordinate representation

$$\sum_{l=1}^{n} \langle \mathbf{1}_{D_l}, P_\tau \mathbf{1}_{D_k}\rangle_\mu \; \nu_l \;=\; \lambda \, \langle \mathbf{1}_{D_k}, \mathbf{1}_{D_k}\rangle_\mu \, \nu_k \tag{59}$$

for $k = 1, \ldots, n$. After division of (59) by $\langle \mathbf{1}_{D_k}, \mathbf{1}_{D_k}\rangle_\mu = \mu(D_k) > 0$, we obtain the convenient form

$$S\nu \;=\; \lambda\nu$$

with $\nu = (\nu_1, \ldots, \nu_n) \in \mathbf{C}^n$ and $n \times n$ stochastic transition matrix $S = (S_{kl})$, whose entries are given by the one–step transition probabilities from D_k to D_l within the time τ:

$$S_{kl} \;=\; \frac{\langle P_\tau \mathbf{1}_{D_k}, \mathbf{1}_{D_l}\rangle_\mu}{\langle \mathbf{1}_{D_k}, \mathbf{1}_{D_k}\rangle_\mu} \;=\; p(\tau, D_k, D_l). \tag{60}$$

Since P_τ is a Markov operator, its Galerkin discretization S is a (row) stochastic matrix, i.e., $S_{kl} \geq 0$ and $\sum_{l=1}^n S_{kl} = 1$ for every $k = 1, \ldots, n$. Hence, all its eigenvalues λ satisfy $|\lambda| \leq 1$. Moreover, we have the following three important properties [68, 69]:

(i) The row vector $\pi = (\pi_1, \ldots, \pi_n)$ with $\pi_k = \mu(D_k)$ represents the discretized invariant probability measure μ. It is a left eigenvector corresponding to the eigenvalue $\lambda = 1$, i.e., $\pi S = \pi$.

(ii) S is *irreducible and aperiodic*. As a consequence, the eigenvalue $\lambda = 1$ is *simple and dominant*. In particular, the discretized invariant density π is the *unique* invariant density of S.

(iii) If P_τ is reversible then S is *self–adjoint* w.r.t. the discrete scalar product $\langle u, v\rangle_\pi = \sum u_i \bar{v}_i \pi_i$. Equivalently, S satisfies the *detailed balance condition* $\pi_k S_{kl} = \pi_l S_{lk}$ for every $k, l \in \{1, \dots, n\}$. Consequently, all eigenvalues of S are real–valued and contained in the interval $[-1, 1]$.

The discretization of the propagator can be interpreted as a *coarse graining* procedure: Coarse graining the state space $\{x \in \mathbf{X}\} \to \{D_1, \dots, D_n\}$ results in a coarse graining of the propagator $P_\tau \to S$ corresponding to a coarse graining of the Markov process $p(\tau, x, C) \to p(\tau, D_k, D_l)$ with invariant measures $\mu \to \pi$. In doing so, the discretization inherits the most important properties of the propagator.

In numerical experiments, it is desirable to estimate the essential spectral radius $r_{\text{ess}}(P_\tau)$. Since $r_{\text{ess}}(P_\tau) \leq \Delta(P_\tau)$, we suggest the following heuristics to define an indicator $[\Delta(P_\tau)]$ for some upper bound on $r_{\text{ess}}(P_\tau)$. The basic idea is to use a decomposition $\mathcal{D} = \{D_1, \dots, D_n\}$ of the state space and a "discretized" version of

$$\Delta(P_\tau) \quad = \quad \limsup_{\mu(A)\to 0} \sup_{0\neq v\in L^1(\mu)} \frac{1}{\|v\|_1} \|\mathbf{1}_A \circ P_\tau v\|_1.$$

Hence, replacing suprema by maxima w.r.t. the decomposition $\mathcal{D}$ we get

$$\begin{aligned}
[\Delta(P_\tau)] &= \max_{D\in\mathcal{D}} \max_{0\neq v\in\mathcal{V}_n} \frac{1}{\|v\|_1} \|\mathbf{1}_D \circ P_\tau v\|_1 \\
&= \max_{j,k} \frac{1}{\mu(D_k)} \left\|\mathbf{1}_{D_j} \circ P_\tau \mathbf{1}_{D_k}\right\|_1 \\
&= \max_{j,k} S_{jk}.
\end{aligned}$$

Therefore, the maximal entry of the stochastic transition matrix S, obtained from discretizing the propagator P_τ, can be used as an indicator for $r_{\text{ess}}(P_\tau)$. In order to better capture the nature of the limit process $\mu(A) \to 0$, we suggest to use a sequence of decompositions $\mathcal{D}_1, \dots, \mathcal{D}_m$ that get finer and finer, and consider the corresponding sequence of indicators $[\Delta(P_\tau)]_{\mathcal{D}_1}, \dots, [\Delta(P_\tau)]_{\mathcal{D}_m}$. A proof about the reliability of the indicator seem to be possible under additional regularity conditions on the stochastic transition function. However, it should be clear that these regularity conditions conflict the fact that a non–vanishing essential spectral radius is related to singular and therefore irregular behavior of the underlying dynamics (see Theorem 4.9 and Lemma 4.29).

For a fixed decomposition, the indicator may still advantageous be exploited in a hierarchical context to indicate regions of further refinement: If $[\Delta(P_\tau)] = S_{jk} \approx 1$ for some pair (j, k), then we may suggests

(i) a refinement of the sampling corresponding to the jth box, since the approximation quality for the jth box was too bad, or

(ii) a further subdivision of the kth box, since the statistical weight $\mu(D_k)$ was too big.

If neither of the two strategies decreases the value of $[\Delta(P_\tau)]$, then the essential spectral radius may indeed be close to 1. The indicator works quite well for Markovian systems, as we are going to demonstrate it in Section 6; it has also already been successfully applied to biomolecular systems [7]. As it was pointed out by G. Froyland [24], it might be less useful for deterministic systems, in particular for hyperbolic deterministic systems, where we would expect the maximal entry of S to be approximately 0.5 due to expansion and contraction of the dynamics.

5.2 Convergence of Discrete Eigenvalues

We restrict our considerations to the important class of reversible propagators $P_\tau : L^2(\mu) \to L^2(\mu)$ satisfying the conditions (C1) and (C2). Under these assumptions, convergence results have been proved (see [68] for details).

Denote by $\sigma(P_\tau)$ the spectrum of P_τ in $L^2(\mu)$ and by $\sigma_{\text{discr}}(P_\tau) \subset \sigma(P_\tau)$ the discrete spectrum. We are interested in approximating a cluster of (real–valued) discrete eigenvalues $\lambda_c, \dots, \lambda_1 \in \sigma_{\text{discr}}(P_\tau)$ close to 1 and "outside" the disc with radius $r_{\text{ess}}(P_\tau)$. Assume that the eigenvalues are repeated according to their multiplicity with

$$r_{\text{ess}}(P_\tau) < \lambda_c \leq \cdots \leq \lambda_2 < \lambda_1 = 1.$$

and corresponding eigenfunctions $v_c, \cdots, v_1$, orthogonal w.r.t. $\langle \cdot, \cdot \rangle_\mu$. Furthermore, we require that the sequence of the Galerkin ansatz spaces $\mathcal{V}_1 \subset \mathcal{V}_2 \subset \dots$ is dense in $L^2(\mu)$ and the corresponding decompositions $\mathcal{D}_1, \mathcal{D}_2, \dots$ are getting finer and finer, i.e., $\max_{D \in \mathcal{D}_n} \operatorname{diam}(D) \to 0$ as $n \to \infty$. Denote by $S(\mathcal{V}_n)$ the stochastic transition matrix obtained from discretizing the propagator P_τ w.r.t. the ansatz space $\mathcal{V}_n$. Furthermore, denote the eigenvalues and corresponding eigenvectors of $S(\mathcal{V}_n)$ by $\lambda_i(\mathcal{V}_n)$ and $u_i(\mathcal{V}_n)$, respectively (ordered in decreasing magnitude and taken into account multiplicity). Under these assumptions the dominant eigenvalues of $S(\mathcal{V}_n)$ are good approximations of the dominant eigenvalues of P_τ, whenever the discretization is fine enough. In this case $S(\mathcal{V}_n)$ also has a cluster of eigenvalues $\lambda_c(\mathcal{V}_n) \leq \dots \leq \lambda_2(\mathcal{V}_n) < \lambda_1(\mathcal{V}_n) = 1$ close to 1. More precisely, for every $j = 1, \dots, c$, we get

$$\lambda_j(\mathcal{V}_n) \longrightarrow \lambda_j \quad \text{and} \quad u_j(\mathcal{V}_n) \longrightarrow u_j$$

in modulus and in the $L^2(\mu)$–norm, respectively, as $n \to \infty$ [68].

5.3 Evaluating the Stochastic Transition Matrix

We consider the evaluation of the stochastic transition matrix S obtained from discretizing P_τ.

Consider two elements B, C of some decomposition of the state space. Combining $p(\tau, x, C) = \mathbf{E}_x[\mathbf{1}_c(X_\tau)]$ with Eq. (32) we get

$$p(\tau, B, C) \quad = \quad \frac{1}{\mu(B)} \int_B \mathbf{E}_x[\mathbf{1}_c(X_\tau)]\, \mu(\mathrm{d}x), \tag{61}$$

which can be approximated within two steps:

(A1) approximation of the integral

$$\int_B g(x)\, \mu(\mathrm{d}x) \quad \approx \quad \sum_{k=1}^{N} \alpha_k\, g(x_k)$$

by some deterministic or stochastic integration scheme with partition points or random variables $x_1, \ldots, x_N$, respectively, and weights $\alpha_1, \ldots, \alpha_N$ [16, 29];

(A2) approximation of the expectation value

$$\mathbf{E}_x[\mathbf{1}_c(X_\tau)] \quad \approx \quad \frac{1}{M} \sum_{j=1}^{M} \mathbf{1}_C\left(X_\tau(\omega_j, x)\right)$$

by relative frequencies, where $X_\tau(\omega_k, x)$ denotes a realization of the Markov process at time τ with initial distribution $X_0 \sim x$ [52, Chapter 17].

A combination of the two steps (A1) and (A2) with $g(x) = \mathbf{E}_x[\mathbf{1}_c(X_\tau)]$ results in

$$p(\tau, B, C) \quad \approx \quad \frac{1}{M} \sum_{k=1}^{N} \sum_{j=1}^{M} \alpha_k\, \mathbf{1}_C\left(X_\tau(\omega_{kj}, x_k)\right);$$

hence, for each initial point x_k, the Markov process X_τ is realized M times. This allows us to approximate the entries of the stochastic transition matrix S due to $S_{jk} = p(\tau, D_j, D_k)$. The approximation quality of $p(\tau, B, C)$ depends on the interplay between the two approximation steps (A1) and (A2). Numerical experiments in low dimensions show that it is even possible to take $M = 1$, if the number of partition points N is chosen in such a way that the number of points per subset of the decomposition $\mathcal{D}$ is reasonable large. For high–dimensional problems, we in general will be forced to use stochastic integration schemes, such as Monte Carlo methods, to approximate the integral in (A1); for further details see Section 7.1. For low–dimensional problems, we may also use deterministic integration schemes:

Example 5.1 *Assume that* $\mathbf{X} = [a, b] \subset \mathbf{R}$ *and the invariant measure is absolutely continuous w.r.t. the Lebesgue measure, i.e.,* $\mu(\mathrm{d}x) = f(x)\mathrm{d}x$ *for some density* f*. For* $N \in \mathbf{Z}_+$ *define the partition points* $x_k = a + kh$ *with* $k = 1, \ldots, N$ *and* $h = (b-a)/N$*. Using the Trapezoid rule in (A1), and* M *realization of the Markov process in (A2) we get*

$$p(\tau, B, C) \approx \frac{1}{M\,\#[x_k \in B]} \sum_{j,\,k} \mathbf{1}_B(x_k)\, \mathbf{1}_C\big((X_\tau(\omega_j, x_k)\big)\, f(x_k),$$

where $\#[x_k \in B]$ *denotes the total number of partition points* x_k *in* B*.*

Already in moderately low dimensions the strategy presented in Example 5.1 may cause serious memory problems and unacceptable numerical effort. Here the adaptive discretization technique by Dellnitz and Junge [12, 13] can be of significant use. Developed to study in particular hyperbolic dynamical systems, its successful application to a small molecular system is described in [14]. For the analysis of biomolecules, e.g, small peptides, a different approach has to be chosen, as will be outlined in Section 7. Performing a realization of the discrete time Markov process $X_n = \{X_n\}_{n\in\mathbf{Z}_+}$, we may also combine the two approximation steps (A1) and (A2). Recall that we defined for some fixed $\tau > 0$ the time–reversed Markov process $Y_n = \{Y_n\}_{n\in\mathbf{Z}_+}$ with stochastic transition function q_τ and the time–symmetrized Markov process $Z_n = \{Z_n\}_{n\in\mathbf{Z}_+}$ with stochastic transition function r_τ, see Section 3.3 .

Example 5.2 *Let* $x_0, \ldots, x_N$ *denote a sequence of sampling points obtained from a realization of the discrete time Markov process* X_n*. Then*

$$p(\tau, B, C) \approx \frac{\#[x_k \in B \text{ and } x_{k+1} \in C]}{\#[x_k \in B]}, \tag{62}$$

where convergence is guaranteed for μ*–a.e. initial points* x_0 *by conditions (C1) and (C2) and the law of large numbers [52]. Since the reversed sampling* $x_N, \ldots, x_0$ *is a realization of the time–reversed Markov process* Y_n *with stochastic transition function* q_τ*, we have*

$$q_\tau(1, B, C) \approx \frac{\#[x_k \in B \text{ and } x_{k-1} \in C]}{\#[x_k \in B]}.$$

Hence, we can approximate the stochastic transition function r_τ *corresponding to the time–symmetrized Markov process* Z_n *by*

$$r_\tau(1, B, C) \approx \frac{\#[x_k \in B \text{ and } x_{k+1} \in C] + \#[x_k \in B \text{ and } x_{k-1} \in C]}{2\,\#[x_k \in B]}.$$

For a reversible *Markov process the identity* $q_\tau(1, B, C) = p_\tau(1, B, C)$ *and consequently* $r_\tau(1, B, C) = p_\tau(1, B, C)$ *holds. Thus, for reversible Markov processes we may "double" the information by considering both the original as well as the reversed sampling.*

Intuitively, it should be clear that the technique described in Example 5.2 becomes less efficient, if the Markov process admits a decomposition of the state space into very metastable subsets. Then, convergence to equilibrium is very slow so that we have to use a different approach (see Section 7).

5.4 The Numerical Identification Algorithm

We briefly outline the identification algorithm, a detailed description can be found in Deuflhard et al. [17].

The aim is to completely decompose the state space $\mathbf{X}$ into metastable subsets. Hence, given a decomposition $\mathcal{D} = \{D_1, \ldots, D_n\}$ of $\mathbf{X}$, we have to determine a clustering $\{C_1, \ldots, C_c\}$ into c metastable clusters by assigning each D_j with $j = 1, \ldots, n$ to some cluster C_k with $k = 1, \ldots, c$. This is done by exploiting the *almost constant level structure* of the dominant eigenvectors, which shows up in Theorem 3.1. Denote by $v_1, \ldots, v_c$ the eigenvectors corresponding to a cluster of c eigenvalues close to 1. Then, they are almost constant on each metastable subset, i.e., if D_i and D_j belong to the same metastable subset, then $v_k(D_i) \approx v_k(D_j)$ for $k = 1, \ldots, c$. Associate to each subset D_j the c–tuple of eigenvector components

$$D_j \quad \longmapsto \quad (v_1(D_j), \ldots, v_c(D_j))$$

and define a clustering $\{C_1, \ldots, C_c\}$ by collecting subsets D_j with almost identical c–tuples into the same cluster. Then, as it is shown in [17], this is sufficient to define a clustering into metastable clusters in the case of weak coupling. The identification of metastable clusters is reduced to cluster c–tuples w.r.t. geometrical similarity. We have implemented an algorithm, which also copes with larger perturbations in the eigenvector components due to stronger coupling between metastable subsets; for details see [17].