Konrad-Zuse-Zentrum
für Informationstechnik Berlin

Takustraße 7
D-14195 Berlin-Dahlem
Germany

MARCUS WEBER [1] AND JANNES QUER [1]

Estimating exit rates in rare event dynamical systems via extrapolation

[1] ZIB

ZIB Report 15-54 (November 2015)

Herausgegeben vom
Konrad-Zuse-Zentrum für Informationstechnik Berlin
Takustraße 7
D-14195 Berlin-Dahlem

Telefon: 030-84185-0
Telefax: 030-84185-125

e-mail: bibliothek@zib.de
URL: http://www.zib.de

ZIB-Report (Print) ISSN 1438-0064
ZIB-Report (Internet) ISSN 2192-7782

Estimating exit rates in rare event dynamical systems via extrapolation

Marcus Weber and Jannes Quer

December 15, 2015

Abstract

In this article we developed a new method for approximating exit rates for molecular dynamics with metastable regions. Due to the fact that Monte Carlo simulations perform quite poor and are very computational expensive in this setting we create several similar situations with a smoothed potential. For this we introduce a new parameter $\lambda \in [0, 1]$ ($\lambda = 1$ very smoothed potential, $\lambda = 0$ original potential) into the potential which controls the influence the smoothing. We then sample the exit rate for different parameters λ from a given region in the smoothed potential. We try to connect the λ to the exit rate to use this dependency to approximate the real exit rate. The method can be seen as something between hyperdynamics and temperature accelerated MC.

Keywords

rare event sampling, smoothing, membership functions, biased sampling

MSC Classification

82B80, 60G40, 60G10

Introduction

The exit rate out of a molecular state is one out of many things of interested in molecular dynamics. Because it is very hard to compute this quantity directly with Monte Carlo simulations there are several techniques available to accelerate the sampling of this exit rates see for example [Lelièvre, 2015] for a nice overview and mathematical investigation. Often these exit rates are very small and this is why it is hard to get a good estimation

for this rates this is why one can consider this exit rates as a rare event. In this article we consider a stochastic model for the atom movement. For this we use a system of overdamped Langevin equations

$$dX_t = -\nabla V(X_t) + \sqrt{(2D)}dB_t \tag{1.1}$$

where $\nabla V : \mathbb{R}^{3n} \to \mathbb{R}^{3n}$ is the gradient of the energy landscape, B_t is the Brownian motion and $D = \frac{k_B T}{\beta}$ with k_B being the Boltzmann constant, T the temperature and β the damping constant. Since the movement of the molecule in the energy landscape is described by an steepest descent with a random (Gaussian i.i.d) perturbation, the molecule will stay a long time in the conformation corresponding to this energy minimum. The transitions will only take place when there is a lot of randomness (Brownian motion). This, the cragginess and the high dimensionality makes it so hard to estimate the transition rate by a simple MD simulation.

In mathematical terms the exit rate can be described as a conditional probability. Without loss of generality we assume that we have two set $\mathcal{S}, \mathcal{T} \subset \Omega$ with $\mathcal{T} = \Omega \backslash \mathcal{S}$. We will call $\mathcal{S}$ the starting set and $\mathcal{T}$ the target set. The holding probability between these two sets is given by

$$\mathcal{P}(X_t \in \mathcal{T} | X_{t-1} \in \mathcal{S} \dots X_0 \in \mathcal{S}) \tag{1.2}$$

where we have assumed that the stochastic process has started in $\mathcal{S}$ and $X_\cdot$ are realizations of the stochastic differential equation (1.1) for different time steps $t \in [0, T]$ with $T < \infty$.

In order to find an approximated analytical expression for the exit rate we define the corresponding indicator function for the different sets. Let $\mathbb{1}_{\mathcal{S}} : \Omega \to \{0, 1\}$ denote the indicator function of the starting set and $\mathbb{1}_{\mathcal{T}} \to \{0, 1\}$ of the target set. Let X_t for $t > 0$ define a realization of the stochastic differential equation (1.1) and let $\mathbb{1}_0 : \mathbb{R} \to \{0, 1\}$ with $\mathbb{1}_0(x) = 1$ only for $x = 0$. With these preparations, the expression

$$\mathbb{1}_0\Big(\int_0^t \mathbb{1}_{\mathcal{T}}(X_s)ds\Big) \tag{1.3}$$

is 1, if and only if the stochastic process X_s never reached the (open) target set for $s < t$ otherwise it is zero. The exit rate of the starting set $\mathcal{S}$ is the quantity $\hat{\sigma} > 0$ which approximately solves

$$\exp(-\hat{\sigma}t)\mathbb{1}_{\mathcal{S}}(x) \approx \mathbb{E}\Big[\mathbb{1}_{\mathcal{S}}(X_t)\mathbb{1}_0\Big(\int_0^t \mathbb{1}_{\mathcal{T}}(X_s)ds\Big)|X_0 = x\Big], \quad \forall x \in \Omega, t > 0. \tag{1.4}$$

The expectation value on the right hand side is taken over a random variable (depending on realizations of the stochastic process X_t) which is either 1 or 0. The value is only 1, if the process has started in $x = x_0 \in \mathcal{S}$ and never reached $\mathcal{T}$ during the process (stopping

criterion). The probability that a stochastic process of length t meets this condition should decrease almost exponentially due to Arrhenius equation. Thus, the expectation value is approximated by an exponential function. A direct sampling of this quantity by computing realizations of the stochastic process X_t is impossible, if the exiting of the starting set $\mathcal{S}$ is a very rare event (or if the starting set $\mathcal{S}$ is a metastable set). In this case the switching from 1 to 0 is almost never observed, $\hat{\sigma}$ is almost 0.

Homotopy

Rare events occur in stochastic dynamics, if the potential energy function has high energy barriers between the basins or looks like a craggy landscape. Downscaling of the 'hills' is the way to turn rare events into frequent events. Usually, thermodynamic simulations are accelerated by increasing the temperature. Increasing the temperature is equivalent to a linear downscaling of the potential energy function $V : \Omega \to \mathbb{R}$, where Ω is the position space. For a parameter $\lambda \in [0, 1]$ this downscaling can be seen as a homotopy of the potential energy function given by

$$\hat{V}_\lambda(x) = (1 - \lambda)V(x)$$

This homotopy turns a potential energy function with local minima and local maxima step by step into a flat (constant) function, like it is shown in the Fig. 1 on the left side. By the downscaling the Boltzmann distribution of molecular states also turns step by step into a flat histogram on the position space, indicated in Fig. 1 on the right side.

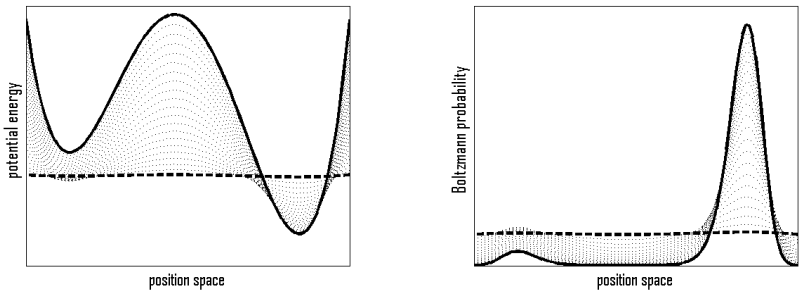

Figure 1: left: Downscaling of a potential energy function (solid line) to a flat function (dashed line). right: The Boltzmann distribution of these functions turn into a flat histogram.

This temperature-driven type of acceleration has some drawbacks see [Sorensen and Voter, 2000] for a general description of this approach. An almost flat histogram is not an easy function to sample from in a high-dimensional space. The reason is the following. Molecular potential energy functions usually have only very located basins of low energy surrounded by high and steep hills. The Boltzmann distribution (at "normal" temperature) has a low entropy - much lower than the entropy of a flat histogram. Another idea of transform-

ing the potential energy landscape is given by solving a boundary vale problem for the potential,

$$\Delta_x \bar{V}(x,\lambda) = \frac{\partial}{\partial \lambda}\bar{V}(x,\lambda), \quad \bar{V}(x,0) = V(x) \tag{1.5}$$

where Δ_x is the Laplace operator in x-direction. In this equation, the diffusion evens out the rough landscape. If we assume the position space to be unbounded, then potential energy function values are usually unbounded, too. This means that the solution of the heat equation is not converging against a constant function. From this one can conclude that the Boltzmann distribution of the solution of the boundary value problem is not a flat histogram, which would be quite hard to sample. Since the heat equation is well known in literature it is known that the solution can be formulated as a convolution with a Gaussian kernel cf. [Evans, 2010]. In the one-dimensional case ($\Omega = \mathbb{R}, \lambda > 0$) we get the homotopy of the form

$$\bar{V}(x,\lambda) = \frac{1}{\sqrt{4\pi\lambda}} \int_{\mathbb{R}} \exp\Big(\frac{-\|x-y\|^2}{4\lambda}\Big) V(y) dy. \tag{1.6}$$

This expression is also quite useful in order to apply our method in a high dimensional setting. If the potential V is know there is only a one dimensional formula to solve to get the transformed potential. In Fig.2 it is shown on the left side, how the diffusion turns a given non-convex potential energy function into a convex function. By this convexity, the Boltzmann distribution changes from a two-local maxima situation into a single bell-shaped function (Fig. 2, right), which should have a low entropy. Thus, it is easy to sample c.f. [Kostrowicki et al., 1991].

Using diffusion for a convexification of the potential energy function is an old idea. Also the fundamental solution of the diffusion equation offers some nice computational aspects. Since we want to use the smoothed potential for our simulations of the Langevin equation (1.1) we have to calculate the derivative of the new potential. Here we can use a well known property of the convolution formula

$$\begin{aligned}\frac{\partial}{\partial x}\bar{V}_\lambda(x) =& \frac{1}{\sqrt{4\pi\lambda}} \int_{\mathbb{R}} \frac{\partial}{\partial x} \exp\Big(\frac{-\|x-y\|^2}{4\lambda}\Big) V(y) dy \\ =& \frac{1}{\sqrt{4\pi\lambda}} \int_{\mathbb{R}} \exp\Big(\frac{-\|t\|^2}{4\lambda}\Big) \frac{\partial}{\partial x} V(x-t) dt.\end{aligned}$$

So in order to calculate the derivative it is also possible to calculate the derivative of the convolution kernel. This may be easier in some situations.

Numerical Aspects Since we consider distance depending potential we assume that we can write the potential in the following form

$$V(x) = \sum_{i>j}^{N} f_{ij}(x) = \sum_{i>j}^{N} H_{ij}(r_{ij}(x))$$

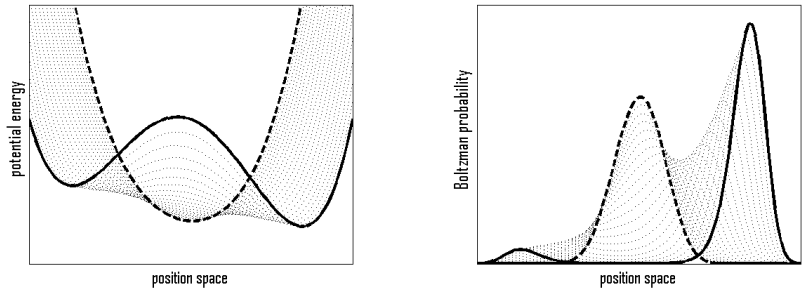

Figure 2: left: Diffusion of a potential energy function (solid line) to a convex function (dashed line). right: The Boltzmann distribution of these functions turn into a bell-shaped histogram with a low entropy.

with $H_{ij} : \mathbb{R}^3 \to \mathbb{R}$ being the potential function for a certain distance and $r_{ij}(x) : \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}$ the distance between two points in $\mathbb{R}^3$ defined by $r_{ij}(x) = \|x_i - x_j\|_2$. Due to the linearity of the Gauss transform it follows

$$[V]_\lambda(x) = \sum_{i>j}^{N} [H_{ij} \circ r_{ij}]_\lambda(x).$$

For simplification we introduce a distance vector $y^{ij} \in \mathbb{R}^3$ $y^{ij}_l = x_{il} - x_{jl}$ $l = 1, 2, 3$ so that we can rewrite H_{ij} with functions $h_{ij} : \mathbb{R}^3 \to \mathbb{R}$

$$H_{ij}(r_{ij}(x)) = H_{ij}(\|y^{ij}\|) = h_{ij}(y^{ij}).$$

For this type of functions More and Wu derived a special Gauss transformation formula

$$[H_{ij} \circ r_{ij}]_\lambda(x) = [h_i j]_{\sqrt{2\lambda}}(x).$$

For the case that H_{ij} is a even function the formula gets

$$[h_{ij}]_{\sqrt{2\lambda}}(x) = \frac{1}{\sqrt{\pi}\,\|y^{ij}\|} \int_{\mathbb{R}} (\|y^{ij}\| + \sqrt{2}\lambda s) H_{ij}(\|y^{ij}\| + \sqrt{2}\lambda s) e^{-s^2} ds. \tag{1.8}$$

This formula is very nice in the sense of numerical treatment. We use a Gauss quadrature to approximate the integral numerically for which there exists a error bound. Consider (1.8) for fixed λ and $r_{ij}(x)$. The integral (1.8) has the following general structure

$$G_{r,\lambda} = K_r \int_{\mathbb{R}} g_{r,\lambda}(s) e^{-s^2} ds. \tag{1.9}$$

Using a Gauss quadrature rule (1.9) can be approximated by a finite sum

$$\int_a^b g_{r,\lambda}(s) w(s) ds \approx \sum_{i=1}^{q} w_i g_{r,\lambda}(s_i).$$

We can choose the weights and the sampling points. We want to guarantee that polynomial of order $2q-1$ are integrated exactly by our approximation. This is why we choose the weight function $w(s) = e^{-s^2}$. Because of our claim for the approximation of our formula we have to choose the roots of Hermit polynomial as sampling points and the weights w_i have to be calculated by the integration over weighted Legendre coefficients. Because of the symmetry of the weight function we know that the samplings points have to be symmetric around 0. So we get the following approximation for (1.8)

$$[f]_\lambda(x) \approx [[f]]_\lambda(x) := \sum_{i>j}^{N} \frac{1}{\sqrt{\pi} r_{ij}(x)} \sum_{k=1}^{q} w_k (r_{ij}(x) + \sqrt{2}\lambda s_k) H_{ij}(r_{ij}(x) + \sqrt{2}\lambda s_k)$$

The error of this approximation is given by

$$|[f]_\lambda(x) - [[f]]_\lambda(x)| \leq \mu^l \lambda^l \psi(x)$$
$$\text{with } \psi(x) = \sqrt{\int_{\mathbb{R}} e^{-s^2} |f^l(x+\lambda s)|^2 ds}$$

with μ being some constant and f^l being some piecewise continuous function on $\mathbb{R}$ with $l \leq q$ cf. [Moré and Wu, 1995].

Identification of target sets and exit rates

We aim at a fast computation of (very small) exit rates $\hat{\sigma}$ from a given starting set to a given target set in position space. The homotopy of the potential energy function is used for an acceleration of the sampling procedure. So instead of sampling with (1.1) we sample the Langevin equation with a flattened potential

$$dX_t^\lambda = -\nabla \bar{V}(X_t^\lambda, \lambda) + \sqrt{(2D)} dB_t \quad X_0^\lambda = X_0,\ \lambda \in [0,1],$$

with $\bar{V}(X_t^\lambda, \lambda)$ as given in (1.6). Temperature scaling and diffusion turn the rare events into frequent events and allow for a statistical analysis. We want to correlate the exit rates with the homotopy parameter λ in order to use extrapolation techniques to estimate the exit rate at $\lambda = 0$. Instead of using fixed starting and target sets for the exit time samplings, we will have to change these sets according to λ too . The reason can clearly be seen in Fig.2, left. A homotopy might change the energy landscape such that starting and target sets should be selected accordingly. In order to get an algebraic expression for the sets depending on the infinitesimal generator L_λ of the samplings, we will use PCCA+ for identifying the metastable regions (starting and target set). The generator L_λ depends on λ because the potential energy does. One eigenfunction of the infinitesimal generator is the constant function $e : \Omega \to \{1\}$ with $L_\lambda e = 0$. First, the eigenvalue $-\sigma_\lambda$ which is closest to the eigenvalue 0 is searched for. The corresponding eigenfunction is denoted as

$v_{2,\lambda} : \Omega \to \mathbb{R}$, i.e.

$$L_\lambda v_{2,\lambda} = -\sigma v_{2,\lambda},$$

omitting the explicit λ-dependency in the followings for the sake of simplicity. PCCA+ uses a linear transform of the leading eigenfunctions (in our case a transform of e and v_2) in order to compute the membership functions of the metastable sets cf. [Deuflhard and Weber, 2005] [Weber, 1999]. Thus, the target set of our exit rate analysis is given by a membership function

$$\chi = \alpha_1 e + \alpha_2 v_2, \quad \alpha_1, \alpha_2 \in \mathbb{R}$$

which is not really a set in the strict sense. The membership function χ is more like a fuzzy set which is shown for the two different homotopy methods in Fig. 3. For this computation we restricted the (1-dimensional) position space to a fixed interval.

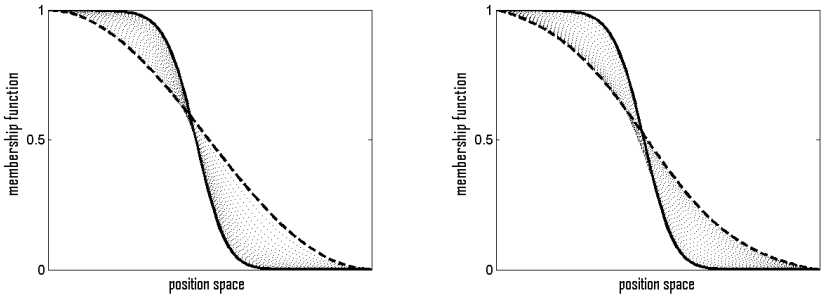

Figure 3: Plot of the membership function of the target set depending on the homotopy parameter λ. The solid line represents the membership function for the original potential energy function. The dashed line corresponds to the flattened potentials shown in Fig. 1 and Fig. 2: The rarer the transitions between the metastable regions the steeper the membership functions. Left: For the case of temperature scaling. Right: For the diffusion approach.

The eigenvalues of L represent the time-scales of the dynamical process, but they do not exactly correspond to the exit rates. We already have solved the problem to find an analytic relation between the (fuzzy) target set χ and the homotopy parameter λ by using PCCA+. A similar approach is now used in order to compute an analytic expression of the starting set $u : \Omega \to [0, 1]$ depending on λ For a given $\epsilon > 0$ we solve the eigenvalue problem

$$Lu - \epsilon(\chi u) = -\hat{\sigma} u, \tag{1.11}$$

such that $-\hat{\sigma}$ corresponds to the highest eigenvalue (which is close to 0) of $L - \epsilon D_\chi$, where D_χ is the multiplication operator based on the membership function χ. The corresponding eigenfunction u is scaled such that the maximal value of u is 1 on a given compact subset of the position space. For the discretized version of (1.11) it will be discussed later that the function u can restricted to only have non negative values. We conjecture, that is also a non negative eigenfunction for the continuous operator $L - \epsilon D_\chi$ and after a suitable

rescaling, u is restricted to have values only in the interval $[0,1]$. In Fig. 4 it is shown how these functions look like for the two homotopy approaches. The more the homotopy parameter converges against $\lambda = 0$, the more u can be interpreted as an indicator function of the start set, i.e., the complement of the starting set.

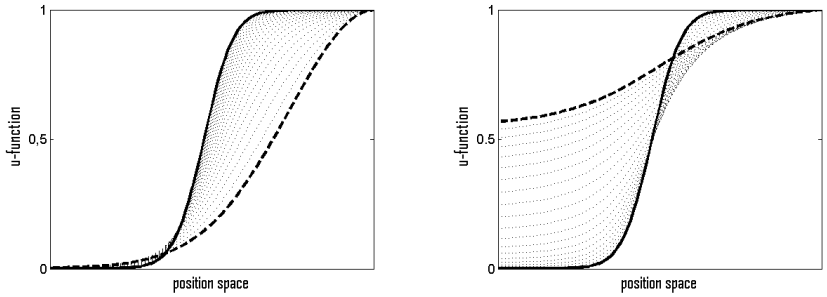

Figure 4: The function u is depending on the homotopy parameter λ. The solid line represents the u- function for the original potential energy function. The dashed line corresponds to the solution of (1.11) for the flattened potentials shown in Fig. 1 and Fig. 2.
Left: temperature scaling. Right: diffusion approach.

That u is indeed similar to a complement function of χ can be understood by looking at the eigenvalue of (1.11). If u is a complement of χ , then $\chi u \approx 0$. The equation approximately restricts to an eigenproblem $Lu \approx -\hat{\sigma}u$ which is solved by piecewise constant functions on the metastable sets. We will now see that it makes sense to regard $\hat{\sigma}$ as the exit rate of the process. By using the Feyman-Kac formula [Øksendal, 2003], the function u meets the equation

$$\exp(-\hat{\sigma}t)u(x) = \mathbb{E}\Big[u(X_t)\exp\Big(-\epsilon\int_0^t \chi(X_s)ds\Big)|X_0 = x\Big] \tag{1.12}$$

where the left hand side is a solution of the differential equations

$$\frac{\partial u}{\partial t} = Lu - \epsilon(\chi u) = -\hat{\sigma}u. \tag{1.13}$$

The above equation is a real equation and not only an approximation. It looks very similar to (1.4). The indicator function $\mathbb{1}_{\mathcal{S}}$ of the starting set is replaced by u , which is indeed regarded as the fuzzy or relaxed starting set function. The same holds for the indicator function $\mathbb{1}_{\mathcal{T}}$ of the target set and its relaxation χ. Instead of testing if the stochastic process has reached the target set by $\mathbb{1}_0\Big(\int_0^t \mathbb{1}_{\mathcal{T}}(X_s)ds\Big)$ in (1.4), this $\{0,1\}$-decision has been replaced by an exponential penalty term $\exp\Big(-\epsilon\int_0^t \chi(X_s)ds\Big) \in [0,1]$, which tends to zero the longer the process stays in χ. Thus, $\hat{\sigma}$ represents indeed a weak formulation of an exit rate out of the starting set. The rate of the rare event.

Instead of estimating the exit rate, one is often interested in finding the expected time τ_{arrive} which a process needs in order to arrive at a target set $\mathcal{T}$ when starting in a certain state x . Note, that the approximation (1.4) for the set-based case turns into an equation

(1.12) for the membership functions. For the starting point $X_0 = x$, the right hand side of (1.12) represents the ratio of states after simulation time t which are still in the starting "set" $u(\cdot)$ after simulation of time t. Thus, one can compute the estimated exit time into the target "set" $\chi(\cdot)$ (starting from a point with $u(x) = 1$) by a weighted mean value computation using the left hand side of

$$\mathbb{E}[\tau_{arrive}] = \frac{\int_0^\infty \exp(-\hat{\sigma}\tau)\tau d\tau}{\int_0^\infty \exp(-\hat{\sigma}\tau)d\tau} = \hat{\sigma}^{-1}.$$

Approximating the eigenproblems

Instead of a continuous trajectory X_s, we only get discrete time-steps from a realization $X_0, X_1, X_2, \ldots$ of a discrete-time and continuous-space Markov chain. The correspondence between the time-discrete transfer operator P^τ and the infinitesimal generator L can be written formally as

$$P^\tau = \exp(\tau L).$$

This transformation keeps the eigenfunctions L of unchanged but turns an eigenvalue $-\hat{\sigma}$ of L into an eigenvalue $\exp(-\tau\sigma)$ of P^τ . This means, for the eigenfunction v_2 of L that

$$P^\tau v_2 = \exp(\tau L)v_2 = \exp(-\tau\sigma)v_2 \tag{1.14}$$

holds. The eigenproblems of L used for the identification of χ are, thus, easily transferable to eigenproblems of P^τ. The eigenproblem (1.11), however, is a bit more complicated. It would be nice, if it would be possible to simulate trajectories according to a dissipative transfer operator $\hat{P}^\tau = \exp(\tau(L - \epsilon D_\chi))$ for the modified infinitesimal generator $L - \epsilon D_\chi$. In this paper, we will apply a heuristics instead, which works very well for our given examples. A time-discretization of the modified infinitesimal generator $L - \epsilon D_\chi$ can be approximated by

$$\hat{P}^\tau \approx \exp(\tau(L - \epsilon D_\chi)) = \exp(\tau L)\exp(-\tau\epsilon D_\chi) = P^\tau \exp(-\tau\epsilon D_\chi). \tag{1.15}$$

Equality holds, if L and D_χ commute. Commuting operators exist for non-ergodic processes, where χ is an indicator function of one of the ergodic compounds of the state space. For almost non-ergodic processes, thus, for metastable processes, this approximation is assumed to be good enough for our purposes. In fact, in the given examples below, the error which stems from this approximation is much smaller than the sampling error. This means, that given the transfer operator of the Markov chain P^τ, we can estimate the exit rate $\hat{\sigma}$ by computing the membership function χ via eigenfunction analysis of P^τ and,

then, solving the eigenfunction problem of $P^\tau \exp(-\tau\epsilon D_\chi)$. The result is

$$\hat{\sigma} = -\frac{1}{\tau}\log(\rho(\hat{P}^\tau)) \approx -\frac{1}{\tau}\log(\rho((P^\tau \exp(-\tau\epsilon D_\chi)))) \tag{1.16}$$

where $\rho(A)$ denotes the spectral radius of A.

Illustrative Example

In this example section we use two common approaches for the investigation of dynamical systems. First we use the homotopy approach to sample the generator of the stochastic process. In the second part we are going to assume that we can describe our dynamical system by a overdamped Langevin Equation. We wee use our accelerated sampling approach to differ the potential in which the process is living in.

In our first example we want to simulate a stochastic process on the double-well potential

$$V(x) = 8x^4 - \frac{44}{3}x^3 + 2x^2 + \frac{11}{3}x + 1.$$

This potential has exactly the shape presented in Fig. 1 and Fig. 2 (on the left, solid line). Instead of a continuous simulation we will directly sample from a discrete-space and discrete-time Markov chain. For the space discretization we divide the interval $[-0.5, 1.5]$ into equal units by $x_i = -0.48 + k0.02$ for $k = 1, \ldots, 101$. For the illustrative example we will need an exact solution of the above eigenvalue problems, thus, the infinitesimal generator is approximated by a 101-by-101-rate matrix L. Consecutive indices i an j have a positive entry

$$L_{ij} = \sqrt{\frac{\exp(-\beta V(x_i))}{\exp(-\beta V(x_j)}},$$

where the negative diagonal elements of L are determined such that the row sums of L are zero cf. [Lie et al., 2013]. All the other elements of the matrix L are zero. In order to obtain a metastable process we set the inverse temperature to be $\beta = 3$. The realization of a Markov chain according to this discrete infinitesimal generator is done by using a transition probability matrix

$$P := \exp(\tau L) \tag{1.17}$$

with a small time step of $\tau = 10$. Note that in realistic molecular simulations small time steps are mandatory, because of numerical reasons and because of the craggy potential energy landscape. Using this kind of construction of a Markov chain and for different homotopy parameters λ ranging between 0 and 0.15, we have simulated Markov chains for

the modified potential energy functions

$$\bar{V}_\lambda(x) = (1 - 6.6\lambda)V(x) \tag{1.18}$$

in case of the temperature-based homotopy approach where the potential energy function is simply scaled down, and

$$\bar{V}_\lambda(x) = V(x) + 96\lambda^2 + (4 - 88x + 96x^2)\lambda \tag{1.19}$$

for the convolution-based homotopy approach where the heat equation is solved analytically. The shape of these modified potential energy functions is also indicated in Fig.1 and Fig. 2 (left side, dotted lines). Based on the formula

$$S = -\sum_{i=1}^{N} \frac{\exp(-\beta V(x_i))}{\sum_j \exp(-\beta V(x_j))} \log\Big(\frac{\exp(-\beta V(x_i))}{\sum_j \exp(-\beta V(x_j))}\Big) \tag{1.20}$$

for the entropy of the Boltzmann distribution, it is possible to show that indeed the entropy of the convolution-based homotopy has a lower bound than the temperature-based homotopy, see Fig. 5. This will lead to a more efficient sampling of the corresponding distribution in the case of the convolution approach.

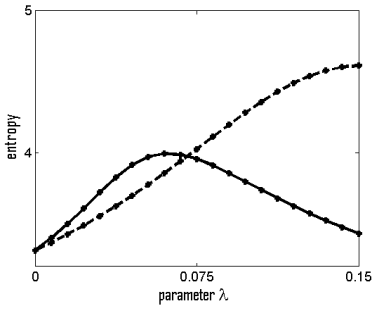

Figure 5: The entropy of Boltzmann distribution of the convolution-based homotopy method (solid line) has a lower upper bound than the entropy of the Boltzmann distribution of the temperature-based homotopy method. The reason is, that the simple downscaling of the potential energy function leads to a flat Boltzmann histogram with a high entropy value. The convolution approach allows for a more efficient sampling.

For the eigenvalue problem (1.11), the parameter ϵ was defined as $\epsilon = 0.01$. The smaller the parameter ϵ, the broader is the region of the position space, where the function u is almost 1 for $\lambda = 0$. The parameter ϵ should not be too small, in order to serve as a penalty expression in (1.12). The correct choice of ϵ is not part of this article. It turns out that the logarithm of $-\hat{\sigma}$ can be approximated with a polynomial of degree 3 in the case of the convolution approach in order to extrapolate the correct exit rate at $\lambda = 0$, if we sample from the modified potentials at $0.0375 \leq \lambda \leq 0.15$ Whereas, for the temperature

approach a polynomial of (at least) degree 2 is needed. The corresponding polynomials are plotted in Fig. 6. The advantage of using a convolution approach can be seen in Fig.

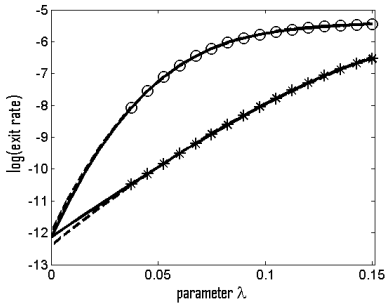

Figure 6: Extrapolation of the logarithm of the exit rate by using a polynomial regression. The real dependency of the exit rate from the parameter λ is plotted as a solid line. The approximated polynomial is plotted as dashed line. For the polynomial regression the indicated λ-values are used (circle: convolution approach; cross: temperature scaling).

7. Here the second largest eigenvalue of $Lv_2 = -\sigma v_2$ for different λ-values is divided by the eigenvalue at $\lambda = 0$. This ratio is denoted as acceleration, because it shows how the original dominant time-scales of the system are accelerated by the homotopy, thus, increasing the second largest eigenvalue σ.

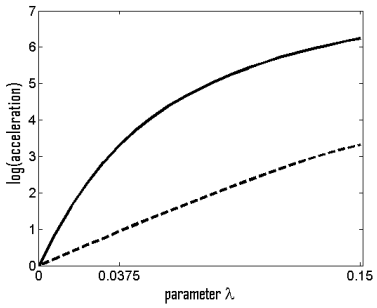

Figure 7: Acceleration of the dominant time-scales of the molecular system depending on λ. Solid line: convolution method. Dashed line: temperature scaling. Note that a logarithmic plot is shown, such that the convolution method is some order of magnitudes "faster" than the temperature scaling.

Simulation based examples

In the previous section we have discussed an illustrative example. For the presented computations, we used a known infinitesimal generator of the system and a time discretization of $\tau = 10$. In real-world applications we aim at using sampling routines for determining the transition probabilities between subsets of the state space. Thus, there is a statistical error connected to the creation of transition matrices. In Figure 8, we have generated a Markov chain of 15000 steps (for different homotopy parameters λ) in order to estimate the transition matrix P^τ. For low values of λ transitions between the right and the left basin

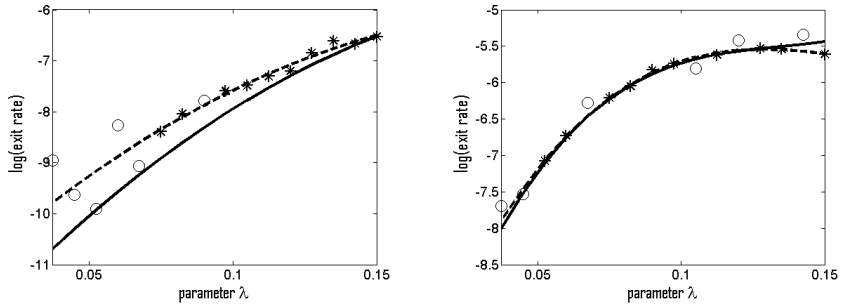

Figure 8: Using a sampling for the estimation of P^τ combined with the approach in (1.12) provides estimates for the exit rates of the molecular system. The potential energy functions have been sampled for 16 different λ-values between 0.04 and 0.15. Left: temperature rescaling. Right: convolution approach. The solid lines indicate the correct polynomial approximations (corresponding to Fig. 6). The dashed lines indicate the polynomial fitting according to the sampling data. For the polynomial fitting outliers have been identified first (circles).

did not occur due to metastability. Only for $\lambda \geq 0.045$ the samplings did provide a good estimate for temperature scaling as well as for the convolution approach. The estimated exit rates corresponding to the presented theory are shown in Fig.8.

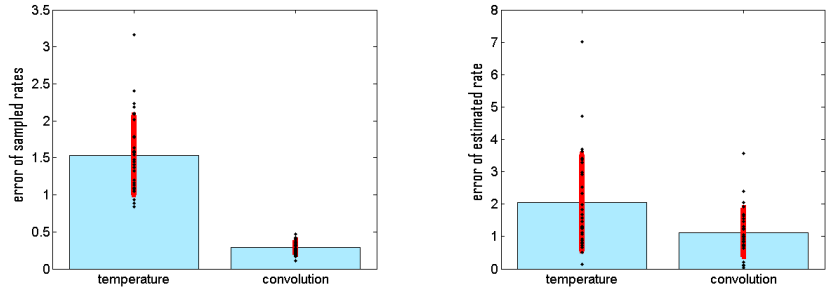

Figure 9: In this figure there is a sampling error for the approximated rates. The mean squared errors of the estimated logarithmic rates are plotted on the left for 30 different samplings (vertical dots). The standard deviation (red line) and the mean error (blue bar) are shown, too. On the right, it is shown how this error is increased by the polynomial extrapolation for finding the original exit rate at $\lambda = 0$. For the temperature scaling we used a polynomial of degree 2, for the convolution method we used a polynomial of degree 3. Better statistical results are provided by the convolution method.

This sampling error was always small for the convolution based method, but for the temperature scaling some sampling results were really bad (Fig. 9 left). Although, in order to extrapolate the exit rate from the samplings, a polynomial fit is needed. We used a polynomial of degree 2 to extrapolate in the case of temperature scaling, and we used a polynomial of degree 3 for the convolution method. A higher polynomial degree leads to a higher amplification of the statistical error (Fig. 9 right). Still the convolution method had a better performance.

For a more realistic example, we will numerically simulate trajectories of the stochastic

differential process. In this case, a discrete infinitesimal generator of the process is not explicitly given (in terms of a matrix) – only a trajectory. Additionally, we will fix the functions u and χ to be characteristic functions of the start and the target set, independent from λ. Furthermore, we will fix the starting point of the process.

We now want to estimate the probability that the process is leaving certain set around a minimum. For the toy example the set is $[-0.5, 0.5]$. The stochastic process which is satisfying the above SDE (1.1) starts at $X_0 = -0.25$. In order to find this probability we calculate 1000 short trajectories of length 1000 in different potentials depending on different homotopy parameters $\lambda \in [0, 0.16]$. We then count the processes which are in the well for every time step. From this we can see how likely it is for the process to leave the well depending on different flattening parameters. This information we can use to estimate the first exit rate for the process in the original potential.

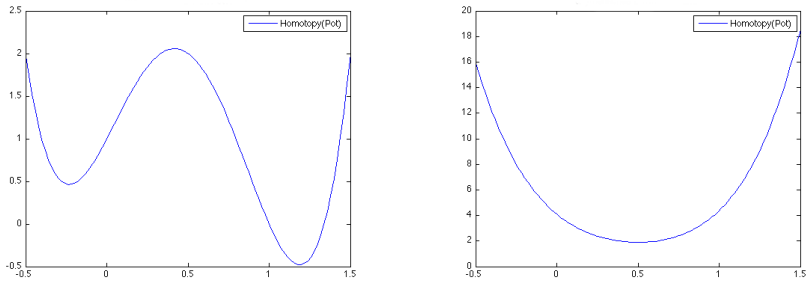

Figure 10: original potential (left), potential with $\lambda = 0.16$

The first result which is already visible form looking at the flattened potential is that the dynamic of the underlying system is faster then the in the original potential. The process sampled in $V_{0.16}$ (most flattened potential) only stays 145 time steps in the well on average while the process sampled in V_0 never the well within the 1000 steps. As one can see in Fig. 11 the number of steps increases while the parameter λ is going to zero. The SDE (1.1) is solved with a Euler Maruyama scheme with $dt = 1/1000$. We estimate the probability by fitting some curves in the data points we generated with the above method.

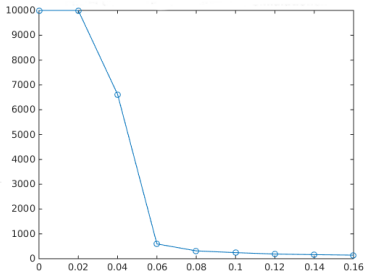

Figure 11: Number of steps still the hitting set is reached depending on λ

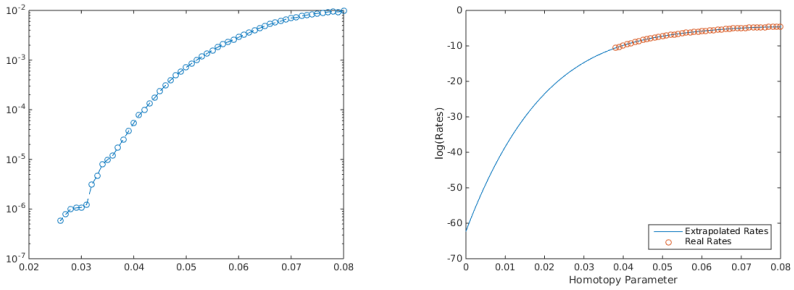

Figure 12: left: Exit rates which are different from 0 by sampling in different transformed potentials. right: Extrapolated exit rates by fitting an polynomial of order 5 in the data.

Butane

With the technique described above, we try to estimate the probability for butane starting with an torsion angle about 180° changing to 120°. In order to simulate butane we calculate the position in the next time step by a stochastic differential equation (1.1). The potential $V : \mathbb{R}^{42} \to \mathbb{R}$ is given by the atom atom interaction depending on the distance of the atoms $r(x, y) = \|x - y\|^2$, where $x, y \in \mathbb{R}^3$ are the atom positions. The derivative of the potential will give us the force field $\nabla V : \mathbb{R}^{42} \to \mathbb{R}^{42}$.

We consider 4 different types of atom atom interactions. For a direct connection of to atoms we calculate the binding energies

$$B_{1,2}(r) = K(r - d^2_{CH/CC})^2.$$

The constant $d_{CH/CC}$ specifies if a CH or CC bond is considered ($d_{CH} = 1.1, d_{CC} = 1.54$) and we use $K = 5$ for all our simulations. The second type of interactions is when there is a atom in between the two considered atoms. The formula for the binding energy looks quite similar to the case before but other constants are used

$$B_{1,3}(r) = K(r - d^2_{HH/CH/CC})^2.$$

Again the constant $d_{HH/CH/CC}$ specifies what kind of interaction is considered ($d_{HH} = 1.78, d_{CH} = 2.19, d_{CC} = 2.6$). The third kind of interaction is used if there are two atoms between the considered atoms. In this case the torsion angle and the Lennard Jones potential have to be taken into account. The Lennard Jones potential is weighted with $l = 0.01$ so that it does not dominate the energy in this case. The formula is given by

$$TLJ(r) = l(\frac{A_{HH/CH/CC}}{r^6} - \frac{B_{HH/CH/CC}}{r^3}) + T(r - d^2_{HH/CH/CC_{180}})^2(r - d^2_{HH/CH/CC_{60}})\cdot$$

Again the constants $A_{HH/CH/CC}, B_{HH/CH/CC}, d_{HH/CH/CC_{180}}, d_{HH/CH/CC_{60}}$ specify which kind of bond is considered. They are given in the table 1. For our simulation we choose

	$A_{HH/CH/CC}$	$B_{HH/CH/CC}$	$d_{HH/CH/CC_{180}}$	$d_{HH/CH/CC_{60}}$
HH	7220	76	2.5	3.1
CH	37430	127.4	3.58	8.1
CC	285800	372.5	4.74	11.1

Table 1: Constants for the TLJ potential and for the LJ potential

the constant $T = 5$. In the last case all interactions where there are more than 4 atoms are between the considered ones the Lennard Jones potential is used. The term is given by

$$LJ(r) = \frac{A_{HH/CH/CC}}{r^6} - \frac{B_{HH/CH/CC}}{r^3}.$$

The constants $A_{HH/CH/CC}, B_{HH/CH/CC}$ are specified by which atoms are interacting. The vales for the different cases are the same as in the case before and can also be looked up in the table 1. The whole potential of butane is given by the summation of the different energies calculated for all atoms. The force field is given by the derivative of the energies.

To calculate the new position we have to solve the SDE. For this we use a Leap Frog scheme with 50 interim stages. We calculated 10 trajectories of length 1000 for each parameter λ. As a first result one can see that also in the high dimensional case the dynamics of the system is faster when the flattened potential is considered see the above figure.

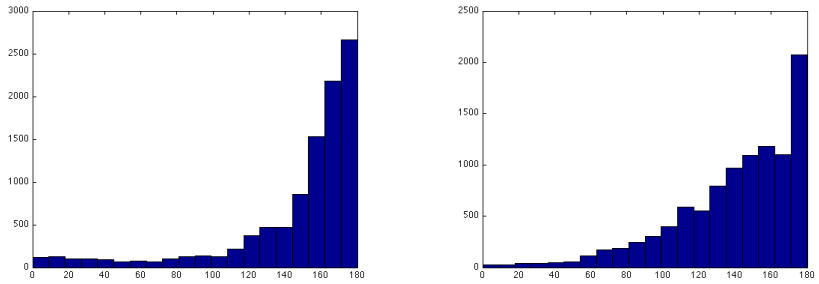

Figure 13: left: Average over 10 trajectories of a system in the original potential.
right: Average over 10 trajectories in the smoothed potential with $\lambda = 1$

With the above setting we can also try to estimate the probability for this event.

Conclusions

We are able to show that homotopy method can be used to calculate exit rates of rare events in dynamical systems. In our case we considered dynamical system which can be discretized in two sets. One starting and one hitting set. In this case the homotopy only has to guarantee that the discretization keeps this two sets separated. In our further investigations we will look at systems with have more than two sets. We will investigate in the question how these sets scale under the homotopy.

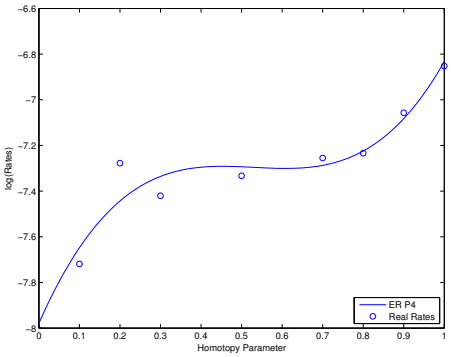

Figure 14: Estimation of exit rate in the original potential

References

[Deuflhard and Weber, 2005] Deuflhard, P. and Weber, M. (2005). Robust perron cluster analysis in conformation dynamics. In *Lin. Alg. Appl. ? Special Issue on Matrices and Mathematical Biology*, volume 398, pages 161 – 184.

[Evans, 2010] Evans, L. (2010). *Partial Differential Equations*. Graduate studies in mathematics. American Mathematical Society.

[Kostrowicki et al., 1991] Kostrowicki, J., Piela, L., Cherayil, B. J., and Scheraga, H. A. (1991). Performance of the diffusion equation method in searches for optimum structures of clusters of lennard-jones atoms. *The Journal of Physical Chemistry*, 95(10).

[Lelièvre, 2015] Lelièvre, T. (2015). Accelerated dynamics: Mathematical foundations and algorithmic improvements. *The European Physical Journal Special Topics*, 224(12):2429–2444.

[Lie et al., 2013] Lie, H. C., Fackeldey, K., and Weber, M. (2013). A square root approximation of transition rates for a markov state model. *SIAM. J. Matrix Anal. Appl.*, 34(2):738 – 756.

[Moré and Wu, 1995] Moré, J. J. and Wu, Z. (1995). Global continuation for distance geometry problems. *SIAM J. OPTIMIZATION*, 7(7):814–836.

[Øksendal, 2003] Øksendal, B. (2003). *Stochastic Differential Equations: An Introduction with Applications*. Hochschultext / Universitext. Springer.

[Sorensen and Voter, 2000] Sorensen, M. R. and Voter, A. F. (2000). Temperature-accelerated dynamics for simulation of infrequent events. *The Journal of Chemical Physics*, 112(21):9599–9606.

[Weber, 1999] Weber, M. (1999). Die Gausstransformation als Homotopiemethode zur globalen Optimierung von Molekülpotentialfunktionen.

Acknowledgement

The authors would like to thank Tony Lelievre for his comments and remarks on our paper.

The authors would like to thank the DFG for founding their work in the CRC-1114.