The Geometric Minimum Action Method:

A Least Action Principle on the Space of Curves

by

Matthias Heymann

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Department of Mathematics

New York University

September 2007

Eric Vanden-Eijnden—Advisor

To my parents.

Acknowledgements

How miraculous an invention the science of mathematics is: it enables us to express some of the most beautiful and ingenious ideas of mankind by placing only a few carefully selected symbols into the right order. Even more miraculous it seems however that at the same time these symbols entirely fail to express the blood, sweat and tears that oftentimes accompany the creation of such pieces of work. Instead, the last evidence for those can be found in the only part that is written in plain text: the acknowledgements.

Little to none of this particular work would exist without the help or influence of several people who crossed my way in the decisive moments.

An important role during the first two years of my PhD was played by Prof. Jonathan Goodman and Mr. Robert Fechter who invested both their time and money into my scientific and non-scientific ideas, thus being the first people at NYU to take me serious as a collaborator. Tamar Arnon gave me both logistic support and much needed encouragement when late in my third year I had to switch my advisor for the second time, and the fellowship committee gave me the opportunity to use my then remaining time more effectively by generously lowering my teaching load.

Especially during that time I was very grateful that I could always rely on the moral support of my parents and my girlfriend Sarah all of whom stubbornly believed in me without truly understanding a single page of my work, and that I could always count on my friend Giuseppe when I needed someone to explore New York City with.

The key player for this thesis however was certainly Prof. Eric Vanden-Eijnden who has been more than only a scientific advisor in the past two years. Always trying to get the best for and out of me, he spent lots of time with me in numerous one-on-one sessions (in person and via skype) to introduce me to large deviations theory, and to fill many of my gaps in numerical analysis. When necessary, he himself was working feverishly to push our project forward, throwing in many decisive tricks to make my algorithms work, while at other times he gave me all the freedom to explore more analytical questions by using my own set of tools. In the end it was just this combination of our very distinct backgrounds that made this collaboration so fruitful.

Finally, I want to thank Prof. Oliver Bühler, Prof. Weinan E, Prof. Robert Kohn, Prof. Maria Reznikoff and Prof. Marco Avellaneda for their useful comments and suggestions.

"La la . . . . . la la!"

little girl on swing, Washington Square Park

Abstract

Dynamical systems with small noise (e.g. SDEs or continuous-time Markov chains) allow for rare events that would not be possible without the presence of noise, e.g. for transitions from one stable state into another. Large deviations theory provides the means to analyze both the frequency of these transitions and the maximum likelihood transition path. The key object for the determination of both is the quasipotential,

$$V(x_1, x_2) = \inf_{T,\psi} S_T(\psi),$$

where $S_T(\psi)$ is the action functional associated to the system, and where the infimum is taken over all $T > 0$ and all paths $\psi : [0, T] \to \mathbb{R}^n$ leading from x_1 to x_2. The numerical evaluation of $V(x_1, x_2)$ however is made difficult by the fact that in most cases of interest no minimizer exists.

Here, this issue is resolved by introducing the action $\hat{S}(\varphi)$ on the space of curves (i.e. $\hat{S}$ is independent of the parametrization of φ) and proving the alternative geometric formulation of the quasipotential

$$V(x_1, x_2) = \inf_{\varphi} \hat{S}(\varphi),$$

where the infimum is taken over all curves $\varphi : [0,1] \to \mathbb{R}^n$ leading from x_1 to x_2. In this formulation a minimizer exists, and we use this formulation to build a flexible algorithm (the geometric minimum action method, gMAM) for finding the maximum likelihood transition curve.

We demonstrate on several examples that the gMAM performs well for SDEs, SPDEs and continuous-time Markov chains, and we show how the gMAM can be adjusted to solve also minimization problems with endpoint constraints or endpoint penalties.

Finally, we apply the gMAM to problems from mathematical finance (the valuation of European options) and synthetic biology (the design of reliable standard genetic parts). For the latter, we develop a new tool to identify sources of instability in (genetic) networks that are modelled by continuous-time Markov chains.

Contents

List of Figures

List of Tables

List of Appendices

Chapter 1

Introduction and main results

Dynamical systems are often subject to random perturbations. Even when these perturbations have small amplitude, they have a profound impact on the dynamics on the appropriate time scale. For instance, perturbations result in transitions between regions around the stable equilibrium points of the deterministic dynamical system which would otherwise be impossible. Such transitions are responsible for metastable phenomena observed in many systems: regime changes in climate, nucleation events during phase transitions, conformation changes of biomolecules and bistable behavior in genetic switches are just a few examples among many others.

When the amplitude of the random perturbations is small, Freidlin-Wentzell theory of large deviations provides the right framework to understand their effects on the dynamics [12, 29, 31]. In a nutshell, the theory builds on the property that events with very little likelihood, when they occur, do so with high probability by following the pathway that is least unlikely. This makes rare events

predictable, in a way that Freidlin-Wentzell theory of large deviations quantifies. The central object in the theory is an action functional whose minimum (subject to appropriate constraints) gives an estimate of the probability and the rate of occurrence of the rare event and whose minimizer gives the pathway of maximum likelihood by which this event occurs. A key practical question then becomes how to compute the minimum and minimizer of the Freidlin-Wentzell action functional. This question is the main topic of this thesis. As we will see, it will lead us to reformulate the action minimization problem in a form which is convenient for numerical purposes but will also shed light on some interesting analytical properties of the minimizer.

Before going there, however, we begin with a brief summary of the main results of Freidlin-Wentzell theory of large deviations which we will use. For simplicity of exposition, we focus here on the finite dimensional case, but the theory can be extended to infinite-dimension (e.g. to situations where (1.3) below is replaced by a stochastic partial differential equation defining a stochastic process X^ε with values in some suitable Hilbert space [5] – situations of this type are considered in Section 4.2).

1.1 Freidlin-Wentzell theory of large deviations

As mentioned above, the central object in the theory is an action functional: if the state space of the dynamical system is embedded in $\mathbb{R}^n$ and if $C(0,T)$ denotes the space of all continuous functions mapping from $[0,T]$ into $\mathbb{R}^n$, this action can

be written as

$$S_T(\psi) = \begin{cases} \int_0^T L(\psi, \dot{\psi})\, dt & \text{if } \psi \in C(0,T) \text{ is absolutely continuous} \\ & \text{and the integral converges,} \\ +\infty & \text{otherwise,} \end{cases} \tag{1.1}$$

where the Lagrangian $L(x, y)$ is given by

$$L(x, y) = \sup_{\theta \in \mathbb{R}^n} \big(\langle y, \theta \rangle - H(x, \theta) \big). \tag{1.2}$$

Here $\langle \cdot, \cdot \rangle$ denotes the Euclidean scalar product in $\mathbb{R}^n$ and $H(x, \theta)$ is the Hamiltonian whose specific form depends on the dynamical system at hand. There are two important classes we shall consider here. The first consists of stochastic differential equations (SDEs) on $\mathbb{R}^n$ with drift vector b and diffusion tensor $a = \sigma\sigma^T$, i.e.

$$dX^\varepsilon(t) = b(X^\varepsilon(t))dt + \sqrt{\varepsilon}\, \sigma(X^\varepsilon(t))dW(t). \tag{1.3}$$

Their Hamiltonian H and Lagrangian L are given by

$$\begin{aligned} H(x, \theta) &= \langle b(x), \theta \rangle + \tfrac{1}{2} \langle \theta, a(x)\theta \rangle, \\ \text{and} \quad L(x, y) &= \big\langle y - b(x), a^{-1}(x)(y - b(x)) \big\rangle. \end{aligned} \tag{1.4}$$

Additional restrictions are required on b and σ in order that large deviations theory applies, but these are usually mild – for instance it suffices that a and b be bounded and uniformly continuous, and that a be uniformly elliptic, i.e. $\exists m > 0\ \forall \xi \in \mathbb{R}^n : \langle \xi, a(x)\xi \rangle \geq m|\xi|^2$, see [12, Chapter 5.3]. The second class

consists of continuous-time Markov jump processes on $\varepsilon\mathbb{Z}^n$ with a generator Q defined for every test function $f : \mathbb{R}^n \to \mathbb{R}$ by

$$(Qf)(x) = \varepsilon^{-1} \sum_{j=1}^{N} \nu_j(x)\big(f(x + \varepsilon e_j) - f(x)\big), \tag{1.5}$$

where $\nu_j : \mathbb{R}^n \to (0, \infty)$, $j = 1, \ldots, N$, are the rates (or propensities) and $e_j \in \mathbb{Z}^n$, $j = 1, \ldots, N$, are the change (or stoichiometric) vectors. The Hamiltonian H for this type of dynamics is given by

$$H(x, \theta) = \sum_{j=1}^{N} \nu_j(x)\big(e^{\langle \theta, e_j \rangle} - 1\big), \tag{1.6}$$

and L must be obtained via (1.2) – in this case, no closed-form expression for L is available in general. Here too, some mild restrictions are necessary in order that large deviations theory applies, e.g. that ν_j be uniformly bounded away from 0 and $+\infty$ [29].

Large deviations theory gives a rough estimate for the probability that the trajectory $X^\varepsilon(t)$, $t \in [0, T]$, $T < \infty$, of the random dynamical system, be it the SDE (1.3), the Markov chain with generator (1.5) or any other system whose action functional can be expressed as (1.1), lies in a small neighborhood around a given path $\psi \in C(0, T)$. The theory asserts that, for δ and ε sufficiently small,

$$\mathbb{P}_x\Big\{ \sup_{0 \le t \le T} |X^\varepsilon(t) - \psi(t)| \le \delta \Big\} \approx \exp\left(-\varepsilon^{-1} S_T(\psi)\right), \tag{1.7}$$

where $\mathbb{P}_x$ denotes the probability conditional on $X^\varepsilon(0) = x$ and we assumed that

$\psi(0) = x$. The estimate (1.7) can be made precise in terms of lower and upper bounds on the probability [12, 29, 31], but for our purpose here it suffices to say that it implies that the probability of various events can be evaluated by constrained minimization. For instance, if B is a Borel subset of $\mathbb{R}^n$, we have

$$\mathbb{P}_x \{X^\varepsilon(T) \in B\} \asymp \exp\Big(- \varepsilon^{-1} \inf_\psi S_T(\psi)\Big) \tag{1.8}$$

where $f(\varepsilon) \asymp g(\varepsilon)$ iff $\log f(\varepsilon)/\log g(\varepsilon) \to 1$ as $\varepsilon \to 0$, and the infimum is taken over all paths ψ such that $\psi(0) = x$ and $\psi(T) \in B$. The minimizer of $S_T(\psi)$ in (1.8) is then the path of maximum likelihood by which the process X^ε ends in B at time T starting from x.

1.2 The role of the quasipotential

In (1.8), T is finite, but large deviations theory can be generalized to make predictions on long time intervals $[0, T(\varepsilon)]$, with $T(\varepsilon) \asymp \exp(\varepsilon^{-1}C)$ and $C > 0$. On these time scales, the effects of the noise become ubiquitous in the sense that it makes likely the occurence of events which are otherwise prohibited by the deterministic dynamics. For this reason these are often the natural time scales over which to analyze the dynamics, and these are the time scales on which we shall mostly focus in this thesis. To understand what happens then, the relevant object is the quasipotential

$$V(x_1, x_2) = \inf_{T>0} \inf_{\psi \in \bar{C}_{x_1}^{x_2}(0,T)} S_T(\psi), \tag{1.9}$$

where $\bar{C}_{x_1}^{x_2}(0,T)$ denotes the space of all absolutely continuous functions $f:[0,T]\to\mathbb{R}^n$ such that $f(0)=x_1$ and $f(T)=x_2$.

A detailed exposition of the significance of the quasipotential is beyond the scope of this thesis and can be found in [12, Chapter 6]. Let us simply say that the quasipotential roughly measures the difficulty to go from point x_1 to point x_2, as made apparent by the following alternative definition (see [12, p. 161])

$$V(x_1,x_2)=\lim_{T\to\infty}\lim_{\delta\to 0}\lim_{\varepsilon\to 0}\left(-\varepsilon\log\mathbb{P}_{x_1}\{\tau_{\delta,x_2}(X^\varepsilon)\le T\}\right) \tag{1.10}$$

where

$$\tau_{\delta,x_2}(X^\varepsilon):=\inf\{t>0\,|\,X^\varepsilon(t)\in B_\delta(x_2)\} \tag{1.11}$$

denotes the first time at which the process X^ε starting from x_1 enters the ball $B_\delta(x_2)$ of radius δ around x_2.

The quasipotential allows one to understand the limiting dynamics over exponentially long intervals of time. For instance, suppose that the deterministic systems associated with (1.3) and (1.5), that is

$$\dot{X}(t)=b(X(t)) \tag{1.12}$$

and, according to Kurtz's Theorem [29],

$$\dot{X}(t)=\sum_{j=1}^{N}\nu_j(X(t))e_j, \tag{1.13}$$

respectively, possess exactly two stable equilibrium points x_1 and x_2, the basins of

attraction of which form a complete partition of $\mathbb{R}^n$. Then on large time intervals the dynamics can be reduced to that of a continuous-time Markov chain on the state space $\{x_1, x_2\}$ with rates

$$k_{1,2} \asymp \exp\left(-\varepsilon^{-1} V(x_1, x_2)\right), \qquad k_{2,1} \asymp \exp\left(-\varepsilon^{-1} V(x_2, x_1)\right). \tag{1.14}$$

These are reminiscent of the Arrhenius law. Similar reductions are possible when (1.12) and (1.13) possess more than two stable equilibrium points or even other stable equilibrium structures such as limit cycles, etc. [12] The quasipotential $V(x_1, x_2)$ is also the key object to characterize the equilibrium distribution of the process in the limit as $\varepsilon \to 0$. For instance, if x_1 is the only stable equilibrium point of (1.12) or (1.13) and it is globally attracting, then

$$\mu^\varepsilon(B) \asymp \exp\Big(- \varepsilon^{-1} \inf_{x_2 \in B} V(x_1, x_2)\Big) \tag{1.15}$$

where B is any Borel set in $\mathbb{R}^n$ and μ^ε is the equilibrium distribution of the process. (1.15) is reminiscent of the Gibbs distribution associated with a potential. Similar statements can again be made in more general situations [12].

1.3 Geometric reformulation

One of our main theoretical results is to show that the variational problem (1.9) which defines the quasipotential admits a geometric reformulation. This reformulation will prove quite useful for numerical purposes. Since it is actually quite simple to understand in the context of SDEs, let us outline the argument in

this case (a similar argument is given in [30] in the different context of quantum tunneling, and it is also at the core of [12, Lemma 3.1, p. 120]). For an SDE like (1.3), (1.9) reduces to (using (1.1) and (1.4))

$$V(x_1, x_2) = \tfrac{1}{2} \inf_{T>0} \inf_{\psi \in \bar{C}^{x_2}_{x_1}(0,T)} \int_0^T |\dot{\psi}(t) - b(\psi(t))|^2_{a(\psi)}\, dt, \tag{1.16}$$

where, for any $u, v, x \in \mathbb{R}^n$, $\langle u, v\rangle_{a(x)} = \langle u, a^{-1}(x)v\rangle$ is the inner product associated with the diffusion tensor $a = \sigma\sigma^T$ and $|u|_{a(x)} = \langle u, u\rangle^{1/2}_{a(x)}$ is the associated norm. Clearly, expanding the square under the integral in (1.16) and using $|u|^2_{a(x)} + |v|^2_{a(x)} \geq 2|u|_{a(x)}|v|_{a(x)}$, we deduce that

$$\begin{aligned} V(x_1, x_2) &\geq \inf_{T,\psi} \int_0^T \big(|\dot{\psi}(t)|_{a(\psi)}|b(\psi(t))|_{a(\psi)} - \langle \dot{\psi}(t), b(\psi(t))\rangle_{a(\psi)}\big) dt \\ &= 2 \inf_{T,\psi} \int_0^T |\dot{\psi}(t)|_{a(\psi)}|b(\psi(t))|_{a(\psi)} \sin^2 \tfrac{1}{2}\eta(t)\, dt, \end{aligned} \tag{1.17}$$

where $\eta(t)$ is the angle between $\dot{\psi}(t)$ and $b(\psi(t))$ in the metric induced by $\langle \cdot, \cdot\rangle_{a(\psi(t))}$. On the other hand, there is a matching upper bound since equality between the integrals at the left-hand sides of (1.16) and (1.17) is achieved in the special case when ψ is constrained so that $|\dot{\psi}(t)| = |b(\psi(t))|$. Thus, the inequality sign in (1.17) can be replaced by an equality sign (this conclusion is proven rigourously in Chapter 2). Now, the key observation is that the integral in (1.17) has become independent of the particular way in which ψ is parametrized by time. In other words, (1.17) offers a geometric expression for the quasipotential as

$$V(x_1, x_2) = 2 \inf_{\gamma} \int_{\gamma} |b|_a \sin^2 \tfrac{1}{2}\eta\, ds, \tag{1.18}$$

where the integral at the right-hand side is the line integral along the curve γ (ds being the arclength element along this curve), η is angle between γ and b at location s along the curve, and the infimum is taken over all curves γ connecting x_1 to x_2.

As shown below, (1.18) can be generalized to dynamical systems with action functional (1.1) which are not SDEs (like e.g. Markov jump processes with generator (1.5)). In these cases too, the key idea is to reformulate (1.9) geometrically in terms of curves $\gamma = \{\varphi(\alpha) \,|\, \alpha \in [0,1]\}$, where $\varphi : [0,1] \to \mathbb{R}^n$ is an arbitrary parametrization of the curve γ. Our main result in this direction is that the quasipotential (1.9) can be expressed as

$$V(x_1, x_2) = \inf_{\varphi \in \bar{C}_{x_1}^{x_2}(0,1)} \hat{S}(\varphi), \quad \text{with} \quad \hat{S}(\varphi) = \sup_{\substack{\vartheta : [0,1] \to \mathbb{R}^n \\ H(\varphi, \vartheta) \equiv 0}} \int_0^1 \langle \varphi', \vartheta \rangle \, d\alpha \tag{1.19}$$

(see Proposition 1 below for a precise statement and other representations of $\hat{S}(\varphi)$).

The action $\hat{S}(\varphi)$ in (1.19) is parametrization-free, i.e. it is left invariant under reparametrization of φ, so it can be interpreted as an action on the space of curves. Compared to (1.9), the minimizer of (1.19) over all φ exists in more general circumstances (since the issue of infinite T does not exist in this formulation), and this makes (1.19) also more suitable for computations, as explained next. The curve $\gamma^\star$ associated with the minimizer $\varphi^\star$ of (1.19) can then be interpreted as the curve of maximum likelihood by which transitions from x_1 to x_2 occur (see Proposition 2).

1.4 Numerical aspects

One of our main points of focus in this thesis is the numerical counterpart to Freidlin-Wentzell theory, i.e. how can one efficiently compute the quasipotential $V(x_1, x_2)$ in (1.19) and the corresponding the minimizer $\varphi^\star$ (the maximum likelihood transition curve)?

The Shooting method. Perhaps the simplest way to minimize the Freidlin-Wentzel action is to use a shooting method (see e.g. [19]) to solve as an initial value problem the boundary value problem for the Hamilton equation associated with the minimization problem in (1.8) or (1.9). Working with the Hamiltonian is an advantage since it is typically known explicitly. On the other hand, in practice this approach quickly becomes inefficient when the dimension of the system increases, and it can be inapplicable in infinite dimension. In addition, the shooting method is better suited for the fixed T problem and leads to additional difficulties when T is optimized upon as well.

The original MAM and the string method. In [8], a numerical technique, termed minimum action method (MAM), was introduced for situations where the minimization of $S_T(\psi)$ is sought over a fixed time interval of length T. The MAM is a relaxation method and is a generalization of previous techniques, such as the one used in [25]. The MAM, however, is not very well-suited for the double minimization problem over ψ and T required to compute the quasipotential $V(x_1, x_2)$ defined in (1.9). The main reason is that the functional may have no minimizer because the infimum is only "achieved" when $T \to \infty$. (In fact,

this will always happen in the typical case in which the prescribed start and end points x_1 and x_2 are critical points of the deterministic dynamics (1.12) or (1.13), see Lemma 3 (ii) below.)

In the special case of an SDE (1.3) in which b is minus the gradient of some potential, $b = -\nabla U$, and σ is the identity, the string method (introduced in [7] and generalizing the nudged elastic band method introduced in [18]) circumvents this problem by taking advantage of the fact that for such systems transition paths are always parallel to the drift $b = -\nabla U$. This allows for a geometric reformulation of the problem and leads to a numerical algorithm in which a discretized curve (or string) is evolved by iterating over the following two step procedure: in the first step the discretization points along the curve are evolved independently in the direction of the flow $b = -\nabla U$; in the second step the curve is reparametrized by redistributing the discretization points at equally spaced positions along the curve (an idea which we will borrow in our approach). Unfortunately, for a generic SDE, transition paths are generally not parallel to the drift, and therefore the string method is not applicable.

The gMAM. The geometric minimum action method (gMAM) presented in this thesis merges and further develops ideas from both the original MAM and the string method. It also has the advantage that it is formulated in terms of the Hamiltonian $H(x, \theta)$. The gMAM resolves the problem of infinite T analytically, leading to the equivalent minimization problem (1.19), which can then be performed in various ways. Here we will use what is, in essence, a pre-conditioned steepest descent algorithm (that is, it is based on a semi-implicit spatio-temporal

discretization of the Euler-Lagrange equation for (1.19)). The only nonstandard aspects of the procedure are that (i) it requires performing first the maximization over ϑ, which we do in an inner loop using a quadratically convergent version of a Newton-Raphson-like algorithm, and (ii) it requires to control the parametrization of the curve φ since the latter is non-unique. Here we opt for parametrizing φ by normalized arclength (as it was done in the string method), meaning that φ satisfies $|\varphi'| = cst$ a.e. on $[0, 1]$.

The gMAM can be applied to generic SDEs, continuous-time Markov chains, and other types of dynamics whose Hamiltonians are known analytically and fulfill the Assumptions 1–3 below. The gMAM can also be applied to SPDEs, as illustrated here, and the underlying strategy may even apply to the minimization of integrals that do not directly fit into the framework of this thesis (see Remark 4).

1.5 Organization, notations and assumptions

The remainder of this thesis, the core results of which will be published in [16], is organized as follows.

In Chapter 2 we first establish and discuss the theoretical results mentioned above, and we also show how to recover the optimal time parametrization once the minimizing curve $\gamma^\star$ is found.

In Chapter 3, we propose and discuss the gMAM algorithm for computing $V(x_1, x_2)$ and the maximum likelihood transition curve $\gamma^\star$.

In Chapter 4 we illustrate these algorithms on several examples: in Section 4.1

we consider an example with bistable behavior first analyzed in [22] in the context of an SDE; in Section 4.2 we consider two SPDE generalizations of this example; and in Section 4.3 we consider a Markov chain used in [1, 28] which arises in the context of the genetic toggle switch.

In Chapter 5 we provide the numerical tools necessary to handle action minimization with either constraints or penalties associated to the location of the endpoint of the curve (Sections 5.1 and 5.2); then we discuss one SDE example to benchmark these methods (Section 5.3), and another example illustrating their use in the context of a problem from mathematical finance (Section 5.4).

In Chapter 6 we give an introduction to the field of synthetic biology (Section 6.1), with focus on possible uses of large deviations theory in the design process of artificial genetic networks (Section 6.2). We then use the gMAM to investigate a more elaborate six-dimensional version of the genetic switch model from Section 4.3, and in this context we develop a tool to detect the sources of instabilities in networks (Sections 6.3 and 6.4).

Finally, we draw some conclusions in Chapter 7.

For the reader's convenience, our most technical calculations and proofs are deferred to several appendices which we recommend to skip on first reading. We do however recommend to read the proof of Proposition 1 in Section 2.3 since it provides valuable insights into the workings of our method. Note also that the Appendices A-D are sorted by their order of dependence and should thus be read in the order in which they appear.

Notations and assumptions. Throughout this thesis we make the following assumptions on the Hamiltonian $H(x,\theta) : D\times\mathbb{R}^n \to \mathbb{R}$ in (1.2), where the domain D is an open connected subset of $\mathbb{R}^n$ (in the introduction we used $D = \mathbb{R}^n$ for simplicity):

Assumption 1. *For every $x \in D$ we have $H(x,0) \leq 0$.*

Assumption 2. *$H(\cdot,\cdot)$ is twice continuously differentiable.*

Assumption 3. *$H_{\theta\theta}(x,\cdot)$ is uniformly elliptic on compact sets, i.e. there exists a function $m(x)$ such that for $\forall \xi,\theta \in \mathbb{R}^n : \langle \xi, H_{\theta\theta}(x,\theta)\xi\rangle \geq m(x)|\xi|^2$, and for every compact set $K \subset D$ we have $m_K := \inf_{x\in K} m(x) > 0$.*

Remark 1. *Equivalently, one can rephrase these assumptions in terms of the Lagrangian $L(x,y)$ defined in (1.2) by requiring that $L(x,y) \geq 0$ for every $x \in D$ and $y \in \mathbb{R}^n$, and that Assumptions 2 and 3 hold with H replaced by L.*

For every $T > 0$ we denote by $C(0,T)$ the space of all continuous functions $f : [0,T] \to D$ equipped with the supremum norm $|f|_{[0,T]} := \sup_{t\in[0,T]} |f(t)|$, and by $\bar{C}(0,T)$ the subspace of all such functions which are absolutely continuous. For every $x_1, x_2 \in D$ we further define the subspaces

$$\begin{aligned}
\bar{C}_{x_1}(0,T) &= \{f \in \bar{C}(0,T) \,|\, f(0) = x_1\},\\
\bar{C}_{x_1}^{x_2}(0,T) &= \{f \in \bar{C}(0,T) \,|\, f(0) = x_1,\ f(T) = x_2\}.
\end{aligned}$$

For every function $f \in \bar{C}(0,T)$ we denote its graph as

$$\gamma(f) := \{f(t) \,|\, t \in [0,T]\}.$$

We say that two functions $f_1 \in \bar{C}(0,T_1)$ and $f_2 \in \bar{C}(0,T_2)$ traverse the same curve and write

$$\gamma(f_1) = \gamma(f_2)$$

if there exists a (necessarily unique) curve $\varphi \in \bar{C}(0,1)$ with $|\varphi'| \equiv cst$ a.e. and two absolutely continuous rescalings $\alpha_1 : [0,T_1] \to [0,1]$ and $\alpha_2 : [0,T_2] \to [0,1]$ with $\alpha_1', \alpha_2' \geq 0$ a.e. such that

$$\begin{aligned} \varphi(\alpha_1(t)) &= f_1(t) \qquad \text{for every } t \in [0,T_1] \\ \text{and} \qquad \varphi(\alpha_2(t)) &= f_2(t) \qquad \text{for every } t \in [0,T_2]. \end{aligned}$$

We denote the space of all functions g traversing the curve of a given function $f \in \bar{C}(0,T_f)$ within a fixed time T by

$$\bar{C}_f(0,T) := \{g \in \bar{C}(0,T) \,|\, \gamma(g) = \gamma(f)\}.$$

To express the distance between two functions $f_1 \in \bar{C}(0,T_1)$ and $f_2 \in \bar{C}(0,T_2)$, we either use the pointwise distance $|f_1 - f_2|_{[0,T_1]}$ (if $T_1 = T_2$), or for more geometrical statements we use the Fréchet distance defined as

$$\rho(f_1,f_2) = \inf_{\substack{t_1:[0,1]\to[0,T_1] \\ t_2:[0,1]\to[0,T_2]}} \left| f_1 \circ t_1 - f_2 \circ t_2 \right|_{[0,1]}, \tag{1.20}$$

where the infimum is taken over all weakly increasing continuous surjective reparametrizations t_1 and t_2 only. (One can quickly check that $\rho(f_1,f_2) = 0$ if $\gamma(f_1) = \gamma(f_2)$.)

We use subscripts to denote differentiation, i.e. $H_\theta(x,\theta) = \partial H/\partial\theta$, etc., and regard all vectors (including gradients) as column vectors.

We introduce the following notation for the Lagrangian:

$$L(x,y) = \sup_{\theta\in\mathbb{R}^n} \big(\langle y,\theta\rangle - H(x,\theta)\big) \tag{1.21}$$

$$= \langle y,\theta^\star(x,y)\rangle - H(x,\theta^\star(x,y)), \tag{1.22}$$

where the maximizer $\theta^\star(x,y)$ is implicitly (and due to Assumption 3 uniquely) defined by

$$H_\theta(x,\theta^\star(x,y)) = y. \tag{1.23}$$

Finally, we call a point $x \in D$ a critical point if

$$H(x,0) = 0 \qquad \text{and} \qquad H_\theta(x,0) = 0. \tag{1.24}$$

Remark 2. *Note that in the examples treated in this thesis, i.e. for the Hamiltonians* (1.4) *and* (1.6)*, we actually have equality in Assumption 1 so that the requirement* $H(x,0) = 0$ *in* (1.24) *is redundant. However, the weaker Assumption 1 allows for a broader class of applications, as we will show in the conclusions (Chapter 7). In fact, in the example in Chapter 7 it is the* first *condition in* (1.24) *that is the decisive one, whereas the second one is fulfilled by every point* $x \in D$*. Our motivation to define critical points in the general case via* (1.24) *will become clear later in Lemmas 1 and 3 (see Remark 5).*

Chapter 2

Theoretical background

2.1 A large deviations action on the space of curves

We collect our main theoretical results regarding the quasipotential $V(x_1, x_2)$ defined in (1.9), with S_T given by (1.1), in the following proposition whose proof will be carried out in Section 2.3.

Proposition 1. *(i) Under the Assumptions 1-3 the following two representations of the quasipotential are equivalent:*

$$V(x_1, x_2) = \inf_{T>0} \inf_{\psi \in \bar{C}_{x_1}^{x_2}(0,T)} S_T(\psi) \qquad \textit{and} \tag{2.1}$$

$$V(x_1, x_2) = \inf_{\varphi \in \bar{C}_{x_1}^{x_2}(0,1)} \hat{S}(\varphi), \tag{2.2}$$

where for every $\varphi \in \bar{C}(0,1)$ the action $\hat{S}(\varphi)$ is given by any of the following four

equivalent expressions:

$$\hat{S}(\varphi) = \inf_{T>0} \inf_{\psi \in \bar{C}_\varphi(0,T)} S_T(\psi), \tag{2.3}$$

$$\hat{S}(\varphi) = \sup_{\substack{\vartheta:[0,1]\to\mathbb{R}^n \\ H(\varphi,\vartheta)\equiv 0}} \int_0^1 \langle \varphi', \vartheta \rangle \, d\alpha, \tag{2.4}$$

$$\hat{S}(\varphi) = \int_0^1 \langle \varphi', \hat{\vartheta}(\varphi, \varphi') \rangle \, d\alpha, \tag{2.5}$$

$$\hat{S}(\varphi) = \int_0^1 L(\varphi, \lambda\varphi')/\lambda \, d\alpha, \qquad \lambda = \lambda(\varphi, \varphi'). \tag{2.6}$$

Here $L(x, y)$ is the Lagrangian associated with the Hamiltonian $H(x, \theta)$ via (1.21), and the functions $\hat{\vartheta}(x, y)$ and $\lambda(x, y)$ are implicitly defined for all $x \in D$ and $y \in \mathbb{R}^n \setminus \{0\}$ as the unique solution $(\hat{\vartheta}, \lambda) \in \mathbb{R}^n \times [0, \infty)$ of the system

$$H(x, \hat{\vartheta}) = 0, \qquad H_\theta(x, \hat{\vartheta}) = \lambda y, \qquad \lambda \geq 0. \tag{2.7}$$

When $\varphi' = 0$ or $\lambda(\varphi, \varphi') = 0$, the integrands in (2.5) and (2.6) are interpreted as 0.

(ii) The functional $\hat{S}(\varphi)$ is invariant under reparametrization of φ, in the sense that $\hat{S}(\varphi_1) = \hat{S}(\varphi_2)$ if $\gamma(\varphi_1) = \gamma(\varphi_2)$. Thus the infimum in (2.2) may be taken subject to some additional constraint on the parametrization of φ, e.g. that $|\varphi'| = cst$ almost everywhere.

(iii) Assume that the sequence $\big((T_k, \psi_k)\big)_{k\in\mathbb{N}}$, with $T_k > 0$ and $\psi_k \in \bar{C}_{x_1}^{x_2}(0, T_k)$ for every $k \in \mathbb{N}$, is a minimizing sequence of (2.1) and that the lengths of the

curves of ψ_k are uniformly bounded, i.e.

$$\lim_{k\to\infty} S_{T_k}(\psi_k) = V(x_1, x_2) \qquad \textit{and} \qquad \sup_{k\in\mathbb{N}} \int_0^{T_k} |\dot{\psi}_k|\, dt < \infty. \tag{2.8}$$

Then the infimum in (2.2) has a minimizer $\varphi^\star$, and for some subsequence $(\psi_{k_l})_{l\in\mathbb{N}}$ we have that

$$\lim_{l\to\infty} \rho(\psi_{k_l}, \varphi^\star) = 0, \tag{2.9}$$

where ρ denotes the Fréchet distance. If $\varphi^\star$ is unique up to reparametrization (i.e. if $\gamma(\tilde{\varphi}) = \gamma(\varphi^\star)$ for every minimizer $\tilde{\varphi}$ of $\hat{S}$) then the full sequence $(\psi_k)_{k\in\mathbb{N}}$ converges to $\varphi^\star$ in the Fréchet distance.

At the end of this section we specialize the results in this proposition to the case of diffusions and derive expression (1.18). The probabilistic interpretation of Proposition 1 is discussed in Section 2.2.

We comment on the meaning of the various quantities entering Proposition 1, starting with the action $\hat{S}(\varphi)$. $\hat{S}(\varphi)$ can be viewed as the rate function on the space of curves (constructed in a way reminiscent of the contraction principle of large deviations theory, see (2.3)). Indeed, $\hat{S}(\varphi)$ is invariant under reparametrization of φ (this is part (ii) of Proposition 1 which is a consequence of (2.3)), and one should think of the integrals (2.4)-(2.6) as line integrals such as

$$\hat{S}(\varphi) = \int_{\gamma(\varphi)} \langle \hat{\vartheta}(z, \hat{\tau}_z), dz\rangle = \int_{\gamma(\varphi)} L(z, \hat{\tau}_z\lambda)/\lambda\, dz, \quad \lambda = \lambda(z, \hat{\tau}_z), \tag{2.10}$$

where $\hat{\tau}_z$ denotes the unit tangent vector along the curve. Note that this invariance also means that in order to make the minimizer $\varphi^\star$ (as opposed to $\gamma(\varphi^\star)$)

unique, one has to decide for a constraint on its parametrization, such as $|\varphi'| = cst$ almost everywhere.

Consider now the other two important quantities entering Proposition 1, $\hat{\vartheta}(x, y)$ and $\lambda(x, y)$. The meaning of the system of equations (2.7) which defines the functions $\hat{\vartheta}(x, y)$ and $\lambda(x, y)$ can be understood as follows: $\hat{\vartheta}(x, y)$ is the extremal point of the set $\{\theta \in \mathbb{R}^n \,|\, H(x, \theta) \leq 0\}$ in the direction y (see the illustration in Figure 2.1); therefore $H_\theta(x, \hat{\vartheta})$ is parallel to y, and $\lambda(x, y)$ is the factor such that $H_\theta(x, \hat{\vartheta}) = \lambda y$. One can quickly check the properties

$$\hat{\vartheta}(x, cy) = \hat{\vartheta}(x, y) \quad \text{and} \quad c\lambda(x, cy) = \lambda(x, y) \quad \text{for } \forall c > 0, \tag{2.11}$$

which are used to show that the representations (2.5) and (2.6) of $\hat{S}$ are invariant under reparametrization of φ (Proposition 1 (ii), see also Lemma 10 in Appendix A).

Two further interpretations of λ are: (i) $\lambda(x, y)y$ is the optimal speed for moving into a given direction y starting from point x (this will become clear in the proof of Proposition 1), and (ii) λ is the Lagrange multiplier used to enforce the constraint in (2.4). λ can also be used as an indicator that tells us where the curve passes a critical point, as part (i) of the following Lemma shows.

Lemma 1. *(i) Let $y \neq 0$. Then x is a critical point if and only if $\lambda(x, y) = 0$.*

(ii) If x is a critical point and $y \neq 0$ then $\hat{\vartheta}(x, y) = 0$.

(iii) If x is a critical point then for $\forall y \in \mathbb{R}^n$: $\lim_{\lambda \to 0+} L(x, \lambda y)/\lambda = 0$.

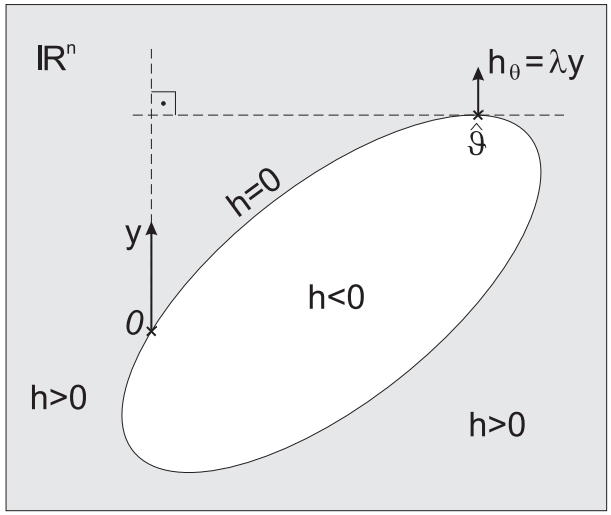

Figure 2.1: Illustration of the system of equations (2.7) for fixed x and y, with $h(\theta) := H(x,\theta)$: $\hat{\vartheta}$ is the extremal point of $\{\theta \in \mathbb{R}^n \,|\, h(\theta) \leq 0\}$ in direction y, λ is then the value such that $h_\theta(\hat{\vartheta}) = \lambda y$.

(iv) Thus the local action of $\hat{S}$, given by any of the integrands in (2.4)–(2.6), vanishes as φ passes a critical point.

The proof of Lemma 1 is carried out in Appendix B. Note also that parts (i) and (iii) explain why we have to interpret the integrand in (2.6) as 0 if $\lambda = 0$.

The case of diffusion processes. If we specialize to diffusion processes whose dynamics is given by (1.3), $L(x,y)$ is available explicitly (see (1.4)), and this allows one to give a closed form formula also for $\hat{S}(\varphi)$. A quick calculation shows that in this case

$$\hat{\vartheta}(x,y) = a^{-1}(x)\Big(\frac{|b(x)|_a}{|y|_a}y - b(x)\Big) \tag{2.12}$$

$$\text{and} \qquad \lambda(x,y) = \frac{|b(x)|_a}{|y|_a}. \tag{2.13}$$

As a result, $\hat{S}(\varphi)$ as given by (2.5) and (2.6) reduces to

$$\hat{S}(\varphi) = \int_0^1 \big(|\varphi'|_a |b(\varphi)|_a - \langle \varphi', b(\varphi)\rangle_a\big) d\alpha \qquad \text{(SDE)}, \tag{2.14}$$

consistent with (1.18).

2.2 Probabilistic interpretation

The minimizing curve $\gamma(\varphi^\star)$ in Proposition 1 (iii) has also a probabilistic interpretation: Let $\{(X_t^\varepsilon)_{t\geq 0},\ \varepsilon > 0\}$ be a family of processes that for every fixed $T > 0$ in the limit $\varepsilon \to 0$ satisfy a large deviations principle with respect to the metric induced by $|\cdot|_{[0,T]}$, and let the associated rate function S_T be of the form (1.1), where the Hamiltonian H fulfills the Assumptions 1-3. Then the statement in Proposition 2 holds.

Put simply, Proposition 2 says the following (under certain technical assumptions): Given that a transition from x_1 to a small ball around x_2 occurs before some finite time T close to some $T^\star = T^\star(\varphi^\star)$, the probability that the process follows $\gamma(\varphi^\star)$ throughout this transition goes to 1 as $\varepsilon \to 0$. ($T^\star$ can be interpreted as the "optimal transition time", see Section 2.4.)

The precise statement is as follows:

Proposition 2. *Assume that the action $\hat{S}$ has a minimizer $\varphi^\star$ among all curves leading from $x_1 \in D$ to $x_2 \in D$, and that this minimizer is unique up to reparametrization. Define $T^\star := \int_0^1 1/\lambda(\varphi^\star, \varphi^{\star\prime})\, d\alpha \in (0, \infty]$, and for any $\delta > 0$ and any path $\psi : [0, T] \to D$, $T > 0$, let $\tau_\delta(\psi)$ denote the first time the path ψ*

hits the closed ball $B_\delta(x_2)$.

Further assume that for all $\delta, T > 0$ *sufficiently close to* 0 *and* $T^\star$*, respectively, there exists a unique path* $\psi_{\delta,T} \in \bar{C}_{x_1}(0,T)$ *such that*

$$S_T(\psi_{\delta,T}) = \inf_{\substack{\psi \in \bar{C}_{x_1}(0,T) \\ \tau_\delta(\psi) \le T}} S_T(\psi),$$

and such that the length and the endpoint of this path fulfills

$$\limsup_{\substack{\delta \to 0+ \\ T \to T^\star}} \int_0^T |\dot{\psi}_{\delta,T}|\, dt < \infty \qquad \text{and} \qquad \lim_{\substack{\delta \to 0+ \\ T \to T^\star}} \psi_{\delta,T}(T) = x_2. \tag{2.15}$$

Then for every $\eta > 0$ *we have*

$$\lim_{\substack{\delta \to 0+ \\ T \to T^\star}} \liminf_{\varepsilon \to 0} \mathbb{P}\Big(\rho\big(X^\varepsilon|_{[0,\tau_\delta(X^\varepsilon)]}, \varphi^\star\big) \le \eta \;\Big|\; \tau_\delta(X^\varepsilon) \le T \Big) = 1, \tag{2.16}$$

where ρ *is the Fréchet distance. Equation* (2.16) *remains true if* $X^\varepsilon|_{[0,\tau_\delta(X^\varepsilon)]}$ *is replaced by* $X^\varepsilon|_{[0,T]}$.

Remark 3. *(i) The form of Equation* (2.16) *was chosen to resemble the formula* (1.10)*. We actually prove a stronger statement, namely that for* δ *and* T *sufficiently close to* 0 *and* $T^\star$ *the limit in* (2.16) *as* $\varepsilon \to 0$ *is already* equal *to* 1.

(ii) The second condition in (2.15) *can be shown to hold whenever* x_1 *is a critical point, or under other technical assumptions (such as bounded lengths of the paths* $\psi_{\delta,T}$ in phase space *if* $T^\star$ *is finite). We decided not to go any further here in order to keep the length of the proof within reasonable limits.*

Sketch of Proof. The proof of this proposition relies on three rather technical

statements (Steps 1-3 below) whose proofs can be found in Appendix D. Note that the proof of Step 2 will require techniques that we will only develop in the proof of Proposition 1 in Section 2.3.

First we would like to estimate the probability in (2.16) below by the same expression with $X^\varepsilon|_{[0,\tau_\delta(X^\varepsilon)]}$ replaced by $X^\varepsilon|_{[0,T]}$ (and then show that the resulting expression still converges to 1). Since replacing $X^\varepsilon|_{[0,\tau_\delta(X^\varepsilon)]}$ by $X^\varepsilon|_{[0,T]}$ can potentially decrease the Fréchet distance and thus increase the probability, we have to decrease η at the same time. This is done in

Step 1: There exists an $\tilde{\eta} > 0$ such that for small enough δ

$$\rho(X^\varepsilon|_{[0,T]}, \varphi^\star) \leq \tilde{\eta} \quad \text{and} \quad \tau_\delta(X^\varepsilon) \leq T \quad \Rightarrow \quad \rho(X^\varepsilon|_{[0,\tau_\delta(X^\varepsilon)]}, \varphi^\star) \leq \eta. \tag{2.17}$$

We then want to replace $\varphi^\star$ by $\psi_{\delta,T}$ since we expect X^ε to be close to $\psi_{\delta,T}$ (even pointwise). This can be achieved by showing

Step 2: $\lim_{(T,\delta)\to(T^\star,0+)} \rho(\psi_{\delta,T}, \varphi^\star) = 0$.

Putting both steps together, we find that if δ is small enough so that (2.17) is true, and if δ and T are also close enough to 0 and $T^\star$ so that $\rho(\psi_{\delta,T}, \varphi^\star) \leq \frac{1}{2}\tilde{\eta}$, then we have

$$\begin{aligned} 1 &\geq \mathbb{P}\big(\rho(X^\varepsilon|_{[0,\tau_\delta(X^\varepsilon)]}, \varphi^\star) \leq \eta \mid \tau_\delta(X^\varepsilon) \leq T\big) \\ &\geq \mathbb{P}\big(\rho(X^\varepsilon|_{[0,T]}, \varphi^\star) \leq \tilde{\eta} \mid \tau_\delta(X^\varepsilon) \leq T\big) \\ &\geq \mathbb{P}\big(\rho(X^\varepsilon|_{[0,T]}, \psi_{\delta,T}) \leq \tfrac{1}{2}\tilde{\eta} \mid \tau_\delta(X^\varepsilon) \leq T\big) \\ &\geq \mathbb{P}\big(|X^\varepsilon - \psi_{\delta,T}|_{[0,T]} \leq \tfrac{1}{2}\tilde{\eta} \mid \tau_\delta(X^\varepsilon) \leq T\big). \end{aligned} \tag{2.18}$$

Now suffices to show

Step 3: For all $T, \delta > 0$ the set of paths $\{\psi \in \bar{C}_{x_1}(0,T) \,|\, \tau_\delta(\psi) \leq T\}$ is regular with respect to S_T

(i.e. the minimal action on the closure of that set is the same as the one on its interior, see [12, p. 85]). Then, since $\psi_{\delta,T}$ is the unique minimizer of the action S_T on the set of paths $\{\psi \in \bar{C}_{x_1}(0,T) \,|\, \tau_\delta(\psi) \leq T\}$, by [12, p. 86, Thm. 3.4] the lower bound in (2.18) converges to 1 as $\varepsilon \to 0$, terminating the proof. □

2.3 Lower semi-continuity of $\hat{S}$, proof of Proposition 1

The proof of part (iii) of Proposition 1 relies on parts (ii) and (iii) of the following lemma which states some important technical properties of the action $\hat{S}(\varphi)$. The proof of this lemma will be carried out in Appendix C.

Lemma 2. *(i) For every $M > 0$ and every compact set $X \subset D$, the set*

$$C_{X,M} := \Big\{\varphi \in \bar{C}(0,1) \;\Big|\; \varphi(0) \in X,\ |\varphi'| \leq M \ \ a.e.\Big\} \tag{2.19}$$

is a compact subset of $C(0,1)$.

(ii) For every $x_1, x_2 \in D$ and every $M > 0$, the set

$$C_M^{x_1,x_2} := \Big\{\varphi \in \bar{C}_{x_1}^{x_2}(0,1) \;\Big|\; |\varphi'| \leq M \ \ a.e.\Big\} \tag{2.20}$$

is a compact subset of $C(0,1)$.

(iii) For every $M > 0$ *and every compact* $X \subset D$*, the functional* $\hat{S} : C_{X,M} \to \mathbb{R}$ *defined by* (2.4) – (2.6) *is lower semicontinuous with respect to uniform convergence.*

(iv) $\hat{S}$ *attains its infimum on every non-empty closed subset of* $C_{X,M}$*. Specifically, it attains its infimum on the sets* $C_M^{x_1,x_2}$*.*

We can now begin with the proof of Proposition 1 which generalizes the heuristic argument given in Section 1.3 for the case of diffusions.

Proof of Proposition 1. (i) and (ii): The main idea of the proof is to rewrite the quasipotential as

$$\begin{aligned} V(x_1,x_2) &= \inf_{T>0} \inf_{\psi \in \bar{C}_{x_1}^{x_2}(0,T)} S_T(\psi) \\ &= \inf_{T>0} \inf_{\varphi \in \bar{C}_{x_1}^{x_2}(0,1)} \inf_{\psi \in \bar{C}_{\varphi}(0,T)} S_T(\psi) \\ &= \inf_{\varphi \in \bar{C}_{x_1}^{x_2}(0,1)} \Big(\inf_{T>0} \inf_{\psi \in \bar{C}_{\varphi}(0,T)} S_T(\psi) \Big) \\ &=: \inf_{\varphi \in \bar{C}_{x_1}^{x_2}(0,1)} \hat{S}(\varphi) \end{aligned} \tag{2.21}$$

where $\hat{S}(\varphi)$ is defined by (2.3). Note that it is clear from the definition (2.3) that $\hat{S}(\varphi)$ actually only depends on the curve that φ traverses and not on the specific parametrization of φ (which already proves part (ii)). Therefore, to prove that $\hat{S}(\varphi)$ as defined in (2.3) can also be written in the forms (2.4)-(2.6), it is enough to restrict ourselves to all those functions $\varphi \in \bar{C}_{x_1}^{x_2}(0,1)$ with the additional property that $|\varphi'| \equiv cst$ almost everywhere, and then to show that the representations (2.4)-(2.6) are invariant under reparametrization as well.

To do so, let $\varphi \in \bar{C}_{x_1}^{x_2}(0,1)$ be given with $|\varphi'| \equiv cst$ almost everywhere. To get a lower bound for $\hat{S}(\varphi)$, we estimate, for any $T > 0$ and any path $\psi \in \bar{C}_\varphi(0,T)$:

$$\begin{aligned} S_T(\psi) &= \int_0^T L(\psi, \dot{\psi})\, dt = \int_0^T \sup_{\theta \in \mathbb{R}^n} \big(\langle \dot{\psi}, \theta \rangle - H(\psi, \theta) \big)\, dt \\ &\geq \int_0^T \sup_{\substack{\theta \in \mathbb{R}^n \\ H(\psi,\theta)=0}} \big(\langle \dot{\psi}, \theta \rangle - H(\psi, \theta) \big)\, dt = \int_0^T \sup_{\substack{\theta \in \mathbb{R}^n \\ H(\psi,\theta)=0}} \langle \dot{\psi}, \theta \rangle\, dt \\ &= \int_0^1 \sup_{\substack{\theta \in \mathbb{R}^n \\ H(\varphi,\theta)=0}} \langle \varphi', \theta \rangle\, d\alpha, \end{aligned}$$

where in the last step we applied Lemma 10 in Appendix A, with $\ell(x,y) := \sup_{\theta \in \mathbb{R}^n,\, H(x,\theta)=0} \langle y, \theta \rangle$. Since the last expression only depends on φ, this shows that the representations (2.4) and (2.5) are lower bounds for $\hat{S}(\varphi)$:

$$\begin{aligned} \hat{S}(\varphi) &= \inf_{T>0} \inf_{\psi \in \bar{C}_\varphi(0,T)} S_T(\psi) \\ &\geq \int_0^1 \sup_{\substack{\theta \in \mathbb{R}^n \\ H(\psi,\theta)=0}} \langle \varphi', \theta \rangle\, d\alpha \geq \int_0^1 \langle \varphi', \hat{\vartheta}(\varphi, \varphi') \rangle\, d\alpha, \end{aligned} \tag{2.22}$$

where we used the first equation in (2.7). To obtain an upper bound on $\hat{S}(\varphi)$, define a minimizing sequence $\big((T_k, \psi_k)\big)_{k \in \mathbb{N}}$ as follows. For every $k \in \mathbb{N}$ let

$$\begin{aligned} \lambda_k(\alpha) &:= \max\big\{ \lambda(\varphi(\alpha), \varphi'(\alpha)), \tfrac{1}{k} \big\}, && \alpha \in [0,1], \\ G_k(\alpha) &:= \int_0^\alpha 1/\lambda_k\, da, && \alpha \in [0,1], \\ T_k &:= G_k(1), \\ \psi_k(t) &:= \varphi(G_k^{-1}(t)), && t \in [0, T_k]. \end{aligned} \tag{2.23}$$

Since for the rescaling $\alpha(t) := G_k^{-1}(t)$ we have $\alpha'(t) = \lambda_k(\alpha(t))$ and thus $\frac{1}{k} \leq \alpha'(t) \leq |\lambda_k|_\infty < \infty$ for every $t \in [0, T_k]$ (see Lemma 12 in Appendix A), $\alpha(t)$ is absolutely continuous. Therefore we see from (2.23) that $\gamma(\psi_k) = \gamma(\varphi)$, i.e. $\psi_k \in \bar{C}_\varphi(0, T_k)$.

To compute $S_{T_k}(\psi_k)$, we perform the change of variables $t = t(\alpha) = G_k(\alpha)$, so that $dt = d\alpha/\lambda_k$, $\varphi(\alpha) = \psi_k(G_k(\alpha))$ and $\varphi'(\alpha) = \dot{\psi}_k(t)G_k'(\alpha) = \dot{\psi}_k(t)/\lambda_k(\alpha)$, and we find that

$$S_{T_k}(\psi_k) = \int_0^{T_k} L(\psi_k, \dot{\psi}_k)\, dt = \int_0^1 L(\varphi, \varphi'\lambda_k)/\lambda_k\, d\alpha. \tag{2.24}$$

Since the integrand on the right-hand side is uniformly bounded in k (see again Lemma 12), to compute the limit of (2.24) as $k \to \infty$, we can exchange limit and integral and obtain the upper bound

$$\begin{aligned} \hat{S}(\varphi) &= \inf_{T>0} \inf_{\psi \in \bar{C}_\varphi(0,T)} S_T(\psi) \\ &\leq \lim_{k\to\infty} S_{T_k}(\psi_k) = \int_0^1 L(\varphi, \varphi'\lambda)/\lambda\, d\alpha, \qquad \lambda = \lambda(\varphi, \varphi'), \end{aligned} \tag{2.25}$$

which is representation (2.6), where we interpret $L(\varphi, \varphi'\lambda)/\lambda = 0$ if $\lambda = 0$, due to Lemma 1 (i) and (iii).

To show that the integrands of the lower bound in (2.22) and the upper bound in (2.25) are the same, consider first the case $\lambda > 0$. The maximizing θ in the expression

$$L(\varphi, \varphi'\lambda)/\lambda = \sup_{\theta \in \mathbb{R}^n} \big(\langle \varphi', \theta\rangle - H(\varphi, \theta)/\lambda\big)$$

has to fulfill the first- and second-order conditions that

$$\varphi' - H_\theta(\varphi,\theta)/\lambda = 0 \quad \text{and that} \quad -H_{\theta\theta}(\varphi,\theta)/\lambda \text{ is negative definite.}$$

By Assumption 3 and the second equation in (2.7), both conditions are fulfilled by $\theta = \hat{\vartheta}(\varphi,\varphi')$, so that in fact

$$L(\varphi,\varphi'\lambda)/\lambda = \langle \varphi', \hat{\vartheta}\rangle - H(\varphi,\hat{\vartheta})/\lambda = \langle \varphi', \hat{\vartheta}\rangle, \qquad \hat{\vartheta} = \hat{\vartheta}(\varphi,\varphi'), \tag{2.26}$$

using also the first relation in (2.7). When $\lambda = 0$, Equation (2.26) holds as well because then $\hat{\vartheta} = 0$ due to Lemma 1 (i) and (ii) and since we agreed on interpreting the left-hand side of (2.26) as zero if $\lambda = 0$.

Therefore the lower bound in (2.22) and the upper bound in (2.25) are the same, and thus all four representations (2.3)-(2.6) of $\hat{S}(\varphi)$ are equal if $|\varphi'| \equiv cst$ almost everywhere.

To end the proof of part (i), it now only remains to show that the expressions (2.4)-(2.6) are invariant under reparametrization, i.e. for $\hat{S}$ given by any of the representations (2.4)-(2.6) and for any $\tilde{\varphi} \in \bar{C}_\varphi(0,1)$ we have $\hat{S}(\tilde{\varphi}) = \hat{S}(\varphi)$. But this is a direct consequence of Lemma 10, the observations (2.11), and our agreement in the statement of Proposition 1 to interpret the integrands in (2.5) and (2.6) as zero if $\varphi' = 0$.

The proof of part (iii) of Proposition 1 follows a standard argument based on the lower-semicontinuity of the functional $\hat{S}$ and the compactness of an appropriate set of functions, both of which were established in parts (ii) and (iii) of

Lemma 2.

Let a sequence $\big((T_k, \psi_k)\big)$ be given with the properties stated in Proposition 1 (iii), and define the functions $\varphi_k \in \bar{C}_{\psi_k}(0,1)$ such that they traverse the curves $\gamma(\psi_k)$ at normalized unit speed, i.e. $|\varphi_k'| \equiv L_k$ a.e., where L_k is the length of the curve ψ_k, as follows:

Let $\alpha_k : [0, T_k] \to [0,1]$ be defined as $\alpha_k(t) := \frac{1}{L_k}\int_0^t |\dot{\psi}_k(\tau)|\, d\tau$, where $L_k = \int_0^{T_k} |\dot{\psi}_k(\tau)|\, d\tau$, define its "inverse" as $\alpha_k^{-1}(\alpha) := \inf\{t \in [0,1] \,|\, \alpha_k(t) \geq \alpha\}$, and set $\varphi_k := \psi_k \circ \alpha_k^{-1}$. Then we have $\varphi_k(\alpha_k(t)) = \psi_k(t)$ for all $t \in [0, T_k]$, and both α_k and φ_k are absolutely continuous, with $\alpha_k' = |\dot{\psi}_k|/L_k$ and $|\varphi'| = \big(|\dot{\psi}_k|/\alpha_k'\big) \circ \alpha_k^{-1} \equiv L_k$. Therefore we have $\gamma(\varphi_k) = \gamma(\psi_k)$, i.e. $\psi_k \in \bar{C}_{\varphi_k}(0, T_k)$.

Using (2.3), this gives us the estimate

$$\inf_{\varphi \in \bar{C}_{x_1}^{x_2}(0,1)} \hat{S}(\varphi) \leq \hat{S}(\varphi_k) = \inf_{T>0} \inf_{\psi \in \bar{C}_{\varphi_k}(0,T)} S_T(\psi) \leq S_{T_k}(\psi_k)$$

for every $k \in \mathbb{N}$. In the limit as $k \to \infty$, the right-hand side converges to the left-hand side, and it follows that

$$\lim_{k\to\infty} \hat{S}(\varphi_k) = \inf_{\varphi \in \bar{C}_{x_1}^{x_2}(0,1)} \hat{S}(\varphi). \tag{2.27}$$

Since $M := \sup_k |\varphi_k'| = \sup_k L_k < \infty$, the sequence $(\varphi_k)_{k\in\mathbb{N}}$ lies in the compact set $C_M^{x_1,x_2}$ defined in Lemma 2 (ii), and thus there exists a subsequence (φ_{k_l}) that converges uniformly to some limiting function $\varphi^\star \in C_M^{x_1,x_2}$. Now since

$C_M^{x_1,x_2} \subset C_{\{x_1\},M}$, by Lemma 2 (iii) and Equation (2.27) we have

$$\hat{S}(\varphi^\star) \leq \liminf_{l\to\infty} \hat{S}(\varphi_{k_l}) = \inf_{\varphi\in\bar{C}_{x_1}^{x_2}(0,1)} \hat{S}(\varphi)$$

and thus $\hat{S}(\varphi^\star) = \inf_\varphi \hat{S}(\varphi)$, i.e. $\varphi^\star$ is a minimizer of $\hat{S}$.

Since the functions φ_{k_l} are time-rescaled versions of the functions ψ_{k_l} and converge uniformly to $\varphi^\star$, this implies that

$$\rho(\psi_{k_l}, \varphi^\star) = \rho(\varphi_{k_l}, \varphi^\star) \leq |\varphi_{k_l} - \varphi^\star|_{[0,1]} \to 0 \quad \text{as} \quad l \to \infty,$$

proving the first statement of part (iii).

To prove also the second statement, assume now that the minimizer of $\hat{S}$ is unique up to reparametrization, let $\varphi^\star$ be the limit of some converging subsequence of (ψ_k) from the first part of the proof, and suppose that $\rho(\psi_k, \varphi^\star) \nrightarrow 0$ as $k \to \infty$. Then we could construct a subsequence (ψ_{k_l}) such that

$$\inf_{l\in\mathbb{N}} \rho(\psi_{k_l}, \varphi^\star) > 0. \tag{2.28}$$

But by the same arguments as above, this subsequence would have a subsubsequence $(\psi_{k_{l_m}})$ that converges in the Fréchet metric to some limit $\tilde{\varphi}$ that is a minimizer of $\hat{S}$. Now the uniqueness of the minimizer implies that $\gamma(\tilde{\varphi}) = \gamma(\varphi^\star)$, and thus we have $\rho(\psi_{k_{l_m}}, \varphi^\star) = \rho(\psi_{k_{l_m}}, \tilde{\varphi}) \to 0$ as $m \to \infty$, contradicting (2.28). □

Remark 4. *The formulas for $\hat{S}$, $\hat{\vartheta}$ and λ can also be derived as follows: Every $\psi \in \bar{C}_{x_1}^{x_2}(0,T)$ can be written as $\psi = \varphi \circ G^{-1}$, where $\varphi \in \bar{C}_{x_1}^{x_2}(0,1)$ follows the*

path of ψ at constant speed, and $G : [0,1] \to [0,T]$ is an appropriately chosen time-rescaling. Minimizing over all ψ and T is therefore equivalent to minimizing over all functions φ and G. But after a change of variables we see that

$$S_T(\psi) = \int_0^T L(\psi, \dot{\psi})\, dt = \int_0^1 L(\varphi, \varphi'/g) g\, d\alpha,$$

where $g = G' : [0,1] \to (0,\infty)$. The second expression can now easily be minimized over all g (and thus over all G) by setting the derivative of the integrand equal to zero, which leads us directly to the representations (2.5) and (2.6) and the equations (2.7) for $\hat{\vartheta}$ and $\lambda := 1/g$.

This trick may also be useful for problems that do not directly fit into the framework of this paper.

2.4 Recovering the time parametrization

Since $\hat{S}(\varphi)$ is parametrization-free, its minimizer $\varphi^\star$ only gives us information about the graph of the minimizer $\psi^\star$ of the original action $S_T(\psi)$ over both ψ and T (assuming that $\psi^\star$ exists), but not its parametrization by time. However, if the minimizing $T^\star$ is finite we can recover $T^\star$ and the path $\psi^\star$ parametrized by physical time afterwards by defining $G(\alpha) := \int_0^\alpha 1/\lambda(\varphi^\star, \varphi^{\star\prime})\, d\alpha$ for $\alpha \in [0,1]$, $T^\star := G(1)$, and setting $\psi^\star(t) := \varphi^\star(G^{-1}(t))$ for $t \in [0, T^\star]$, since then we have

$$\begin{aligned} S_{T^\star}(\psi^\star) &= \int_0^{T^\star} L(\psi^\star, \psi^{\star\prime})\, dt = \int_0^1 L(\varphi^\star, \varphi^{\star\prime}\lambda)/\lambda\, d\alpha \\ &= \hat{S}(\varphi^\star) = \inf_{\varphi \in \bar{C}^{x_2}_{x_1}(0,1)} \hat{S}(\varphi) = \inf_{T>0} \inf_{\psi \in \bar{C}^{x_2}_{x_1}(0,T)} S_T(\psi), \end{aligned}$$

where we performed the change of variables $t = G(\alpha)$ and used (2.1) and (2.2).

If $T^\star = \int_0^1 1/\lambda(\varphi^\star, \varphi^{\star\prime})\, d\alpha = \infty$ (i.e. if "the minimizing $T^\star$ is infinite"), then no minimizer $(T^\star, \psi^\star)$ of the original action $S_T(\psi)$ exists, but we can still extract information from $\lambda(\varphi^\star, \varphi^{\star\prime})$ by splitting the curve into pieces on which λ is nonzero (i.e. into pieces that do not contain any critical points): The following proposition says that if we recover the parametrization on any such piece as above, then the resulting path will give us the optimal way to move from the starting point to the end point of that piece.

Proposition 3. *Let $\varphi^\star$ be a minimizer of the functional $\hat{S}(\varphi)$ defined by (2.3) – (2.6), parametrized such that $|\varphi^{\star\prime}| \equiv cst$ almost everywhere. Let $\alpha_1, \alpha_2 \in [0,1]$ be such that there is no critical point on $\gamma(\varphi^\star)$ between $\tilde{x}_1 := \varphi^\star(\alpha_1)$ and $\tilde{x}_2 := \varphi^\star(\alpha_2)$.*

Define the rescaling $G(\alpha) := \int_{\alpha_1}^{\alpha} 1/\lambda(\varphi^\star, \varphi^{\star\prime})\, da$, $\alpha \in [\alpha_1, \alpha_2]$, and set $\tilde{\psi}^\star(t) := \varphi^\star(G^{-1}(t))$ for $t \in [0, \tilde{T}^\star]$, $\tilde{T}^\star := G(\alpha_2)$. Then we have

$$V(\tilde{x}_1, \tilde{x}_2) = \inf_{T>0} \inf_{\psi \in \bar{C}_{\tilde{x}_1}^{\tilde{x}_2}(0,T)} S_T(\psi) = S_{\tilde{T}^\star}(\tilde{\psi}^\star). \tag{2.29}$$

Proposition 3, which is proven at the end of this section, is relevant because $T^\star = \int_0^1 1/\lambda(\varphi^\star, \varphi^{\star\prime})\, d\alpha$ is infinite in most cases of interest, e.g. if at least one end point of the path is a critical point, or if the path has to pass a critical point to connect the two given states: Part (ii) in the following lemma, which is a slightly stronger statement than Lemma 1 (i), tells us that the minimizing path needs infinite time to leave, pass through, or reach any critical point of the system:

Lemma 3. *Suppose that φ is parametrized such that $|\varphi'| \equiv cst$ a.e., and let $\alpha_c \in [0,1]$ be such that $\varphi(\alpha_c)$ is a critical point. Then*

(i) $\lambda = \lambda(\varphi, \varphi')$ is Lipschitz continuous at α_c in the sense that there exists a constant $C > 0$ such that for a.e. $\alpha \in [0,1]$ we have $\lambda\big(\varphi(\alpha), \varphi'(\alpha)\big) \leq C|\alpha - \alpha_c|$.

(ii) $1/\lambda$ is not locally integrable at α_c. In particular, if the curve $\gamma(\varphi)$ contains a critical point then $T^\star = \int_0^1 1/\lambda \, d\alpha = \infty$.

The proof of this lemma is technical and is carried out in Appendix B.

Remark 5. *Recall in this context that for SDE's and continuous-time Markov chains critical points are those points x with vanishing drift, $H_\theta(x,0) = 0$. For other Hamiltonians (see e.g. Chapter 7) this may not be enough: Only if in addition that point x fulfills $H(x,0) = 0$ then it is a critical point, and it takes infinite time to pass the point. If however $H(x,0) < 0$, then by Lemma 1 we have $\lambda(x,y) \neq 0$ for any direction $y \neq 0$, and therefore such a point is passed in finite time. This finally justifies our definition of critical points via (1.24): A point fulfills the properties (1.24) if and only if it is passed in infinite time.*

Proof of Proposition 3. First, note that since $\lambda(x,y)$ is continuous (even differentiable, see Lemma 13 in Appendix E), we have $\operatorname{essinf}_{\alpha_1 \leq \alpha \leq \alpha_2} \lambda(\varphi^\star, \varphi^{\star\prime}) > 0$, and thus G is well-defined.

Now assume that (2.29) does not hold. Then there exist $\hat{T} > 0$ and $\hat{\psi} \in \bar{C}_{\tilde{x}_1}^{\tilde{x}_2}(0, \hat{T})$ such that $S_{\hat{T}}(\hat{\psi}) < S_{\tilde{T}^\star}(\tilde{\psi}^\star)$, i.e. $\eta := S_{\tilde{T}^\star}(\tilde{\psi}^\star) - S_{\hat{T}}(\hat{\psi}) > 0$. Observe that

$$
\begin{aligned}
S_{\tilde{T}^\star}(\tilde{\psi}^\star) &= \int_0^{\tilde{T}^\star} L(\tilde{\psi}^\star, \tilde{\psi}^{\star\prime})\, dt \\
&= \int_{\alpha_1}^{\alpha_2} L(\varphi^\star, \varphi^{\star\prime}\lambda)/\lambda\, d\alpha, \qquad \lambda = \lambda(\varphi^\star, \varphi^{\star\prime}) \\
&= \int_{\gamma(\varphi^\star|_{[\alpha_1,\alpha_2]})} L(z, \hat{\tau}_z\lambda)/\lambda\, d\alpha, \qquad \lambda = \lambda(z, \hat{\tau}_z) \\
&= \inf_{T>0} \inf_{\psi \in \bar{C}_{\varphi^\star|_{[\alpha_1,\alpha_2]}}(0,T)} S_T(\psi).
\end{aligned}
$$

We will now use the path $\hat{\psi}$ to construct a contradiction to the minimizing property of $\varphi^\star$. To do so, define the sequence $\big((\psi_k, T_k)\big)$ with $\psi_k \in \bar{C}_{\varphi^\star}(0, T_k)$ for all $k \in \mathbb{N}$ and with $\lim_{k\to\infty} S_{T_k}(\psi_k) = \hat{S}(\varphi^\star)$, as in the proof of Proposition 1 (i). Now let T_1^k and T_2^k be such that $\psi_k(T_1^k) = \tilde{x}_1$ and $\psi_k(T_1^k + T_2^k) = \tilde{x}_2$, set $T_3^k := T_k - T_1^k - T_2^k$, and define the pieces ψ_1^k, ψ_2^k and ψ_3^k by

$$
\begin{aligned}
\psi_1^k(t) &= \psi_k(t), && t \in [0, T_1^k], \\
\psi_2^k(t) &= \psi_k(T_1^k + t), && t \in [0, T_2^k], \\
\psi_3^k(t) &= \psi_k(T_1^k + T_2^k + t), && t \in [0, T_3^k].
\end{aligned}
$$

Finally, define the sequence $(\bar{T}_k, \bar{\psi}_k)$ by replacing the piece of ψ_k between $\tilde{x}_1$ and $\tilde{x}_2$ by $\hat{\psi}$ in order to reduce its action, i.e. let

$$
\bar{\psi}_k(t) := \begin{cases}
\psi_1^k(t), & t \in [0, T_1^k], \\
\hat{\psi}(t - T_1^k), & t \in [T_1^k, T_1^k + \hat{T}], \\
\psi_3^k(t - T_1^k - \hat{T}), & t \in [T_1^k + \hat{T}, T_1^k + \hat{T} + T_3^k],
\end{cases}
$$

and $\bar{T}_k := T_1^k + \hat{T} + T_3^k$. For this path we have

$$
\begin{aligned}
S_{\bar{T}_k}(\bar{\psi}_k) &= S_{T_1^k}(\psi_1^k) + S_{\hat{T}}(\hat{\psi}) + S_{T_3^k}(\psi_3^k) \\
&= S_{T_1^k}(\psi_1^k) + S_{\tilde{T}^\star}(\tilde{\psi}^\star) - \eta + S_{T_3^k}(\psi_3^k) \\
&= S_{T_1^k}(\psi_1^k) + \inf_{T>0} \inf_{\psi \in \bar{C}_{\varphi^\star|_{[\alpha_1,\alpha_2]}}(0,T)} S_T(\psi) - \eta + S_{T_3^k}(\psi_3^k) \\
&\leq S_{T_1^k}(\psi_1^k) + S_{T_2^k}(\psi_2^k) - \eta + S_{T_3^k}(\psi_3^k) \\
&= S_{T_k}(\psi_k) - \eta \\
&\to \hat{S}(\varphi^\star) - \eta = \inf_{\varphi \in \bar{C}_{x_1}^{x_2}(0,1)} \hat{S}(\varphi) - \eta \\
&= \inf_{T>0} \inf_{\psi \in \bar{C}_{x_1}^{x_2}(0,T)} S_T(\psi) - \eta
\end{aligned}
$$

as $k \to \infty$. But this means that for sufficiently large k we have

$$
S_{\bar{T}_k}(\bar{\psi}_k) < \inf_{T>0} \inf_{\psi \in \bar{C}_{x_1}^{x_2}(0,T)} S_T(\psi),
$$

and since $\bar{\psi}_k \in \bar{C}_{x_1}^{x_2}(0, T_k)$ for every $k \in \mathbb{N}$ we have a contradiction. $\square$

Chapter 3

Numerical algorithms

The main objective of this section is to design a numerical algorithm to compute the quasipotential $V(x_1, x_2)$ via minimization of $\hat{S}(\varphi)$ and identify the minimizer $\varphi^\star$ such that

$$V(x_1, x_2) = \inf_{\varphi \in \bar{C}_{x_1}^{x_2}(0,1)} \hat{S}(\varphi) = \hat{S}(\varphi^\star), \tag{3.1}$$

where $\hat{S}(\varphi)$ is the action functional given by (2.3)-(2.6). We are primarily interested in cases where x_1 and x_2 in (3.1) are stable equilibrium points of the deterministic dynamics, $\dot{X}(t) = H_\theta(X(t), 0)$, though the algorithm presented below can also be applied to situations where x_1 and/or x_2 are not critical points.

Several strategies can be used for the minimization problem in (3.1). Here, we will proceed as follows. In Section 3.1, starting from the representation (2.5) we will derive the Euler-Lagrange equation associated with the minimization of $\hat{S}(\varphi)$, assuming that $\hat{\vartheta}(\varphi, \varphi')$ is known. In Section 3.2 we will then design a (preconditioned) steepest descent algorithm for the solution of the Euler-Lagrange equation. If no explicit formula for $\hat{\vartheta}(x, y)$ is available, this algorithm will com-

pute $\hat{\vartheta}(x, y)$ in an inner loop, using an efficient quadratically convergent routine which is derived in Section 3.4. The steepest descent algorithm is based on a proper discretization of the Euler-Lagrange equation and uses an interpolation-reparametrization step, similar to the one used in the string method [7], to enforce exactly a constraint on the parametrization of φ, such as $|\varphi'| = cst$. (As we mentioned in Section 2.1, such a constraint is necessary to make the minimizer $\varphi^\star$ of $\hat{S}$ unique.)

We note that the strategy above may not be the most efficient one: for instance, the nonlinear minimization problem in (2.2) could be tackled by discretizing the action $\hat{S}(\varphi)$ first, then use other techniques than steepest descent (like e.g. a quasi-Newton method such as BFGS or conjugate gradient, or a multi-grid method, cf. [24]). However, the approach that we take here has the advantage that it gives some insight about the nature of the action $\hat{S}(\varphi)$. It was also sufficient for our purpose: even the problem considered in Section 4.2 which involves a stochastic partial differential equation (in which case the path φ is not defined in $\mathbb{R}^n$ but rather in some Hilbert space) can be handled by our algorithm in a few minutes using Matlab on a standard workstation.

3.1 The Euler-Lagrange equation and the steepest descent flow

We have the following result whose proof is carried out in Appendix E:

Proposition 4. *The Euler-Lagrange-equation associated with the minimization*

problem in (2.2) can be written in the following two ways:

$$\begin{cases} 0 = P_{\varphi'}\big(-\lambda^2\varphi'' + \lambda H_{\theta x}\varphi' - H_{\theta\theta}H_x\big) \\ \quad = -\lambda^2\varphi'' + \lambda H_{\theta x}\varphi' - H_{\theta\theta}H_x - \lambda\lambda'\varphi', \\ \varphi(0) = x_1,\ \varphi(1) = x_2, \end{cases} \tag{3.2}$$

$$\textit{where} \quad \lambda = \frac{\langle H_\theta, \varphi'\rangle}{|\varphi'|^2} \quad \textit{and} \quad P_{\varphi'} = I - \frac{\varphi' \otimes H_{\theta\theta}^{-1}\varphi'}{\langle \varphi', H_{\theta\theta}^{-1}\varphi'\rangle},$$

and where H_x, $H_{\theta x}$ and $H_{\theta\theta}$ are evaluated at $(\varphi, \hat{\vartheta}(\varphi, \varphi'))$.

The right-hand side in (3.2) is in fact $\lambda H_{\theta\theta} D\hat{S}(\varphi)$, where $D\hat{S}(\varphi)$ is the gradient of $\hat{S}(\varphi)$ with respect to the L^2 inner product. Note that by taking the Euclidean inner product of (3.2) with $\lambda^{-1}H_{\theta\theta}^{-1}\varphi'$ one can see that $\langle D\hat{S}(\varphi), \varphi'\rangle = 0$ for all α, i.e. the variation $D\hat{S}(\varphi)$ is everywhere perpendicular to the path with respect to the Euclidean metric. This is a simple consequence of the fact that $\hat{S}(\varphi)$ is parametrization-free: if id denotes the identity mapping on $[0,1]$, then for any test function $\eta \in C_c^\infty(0,1)$ and sufficiently small $h > 0$ we have

$$\begin{aligned} 0 &= h^{-1}\big[\hat{S}(\varphi \circ (id + h\eta)) - \hat{S}(\varphi)\big] \\ &= h^{-1}\big[\hat{S}(\varphi + h\eta\varphi' + o(h)) - \hat{S}(\varphi)\big] \\ &\to \langle D\hat{S}(\varphi), \eta\varphi'\rangle_{L^2([0,1];\,\mathbb{R}^n)} = \big\langle \langle D\hat{S}(\varphi), \varphi'\rangle, \eta\big\rangle_{L^2([0,1];\,\mathbb{R})} \qquad \text{as } h \to 0. \end{aligned}$$

The algorithm presented in Section 3.2 finds the solution of (3.2) using a relax-

ation method based on a discretized version of the equation:

$$\begin{cases} \dot{\varphi} = P_{\varphi'}\big(\lambda^2\varphi'' - \lambda H_{\theta x}\varphi' + H_{\theta\theta}H_x\big) + \mu\varphi' \\ \quad = \lambda^2\varphi'' - \lambda H_{\theta x}\varphi' + H_{\theta\theta}H_x + \lambda\lambda'\varphi' + \mu\varphi', \\ \varphi(\tau,0) = x_1, \quad \varphi(\tau,1) = x_2, \quad \varphi(0,\alpha) = \varphi^0(\alpha), \end{cases} \tag{3.3}$$

for $\alpha \in [0,1]$ and $\tau \geq 0$. Here $\varphi = \varphi(\tau,\alpha)$ where τ is the artificial relaxation time, $\dot{\varphi} = \partial\varphi/\partial\tau$, $\varphi' = \partial\varphi/\partial\alpha$, $\varphi'' = \partial^2\varphi/\partial\alpha^2$, and $\mu\varphi'$ is a Lagrange multiplier term added to enforce some constraint on the parametrization of φ, e.g. by normalized arclength (in which case $|\varphi'(\tau,\cdot)| = cst(\tau)$ and the initial condition $\varphi(0,\alpha) = \varphi^0(\alpha)$ must be consistent with this constraint). Adding the term $\mu\varphi'$ has no effect on the graph $\gamma(\varphi)$ of the solution since $\hat{S}(\varphi)$ is parametrization-free.

The simple form of (3.3) is a result of us building the flow on $\lambda H_{\theta\theta}D\hat{S}(\varphi)$ rather than $D\hat{S}(\varphi)$ alone, which is legitimate since $H_{\theta\theta}$ is a positive-definite matrix by Assumption 1 and $\lambda \geq 0$. As we shall show in Chapter 4 where we analyze examples, this choice allows one to design an algorithm that achieves a good balance among speed, stability, and accuracy.

To further understand the properties of this flow, let us note that (3.3) can also be derived independently of the theory developed in Chapter 2:

Remark 6. *Another interpretation of the right-hand side of* (3.3) *is the following. Suppose that ψ is a minimizer of the original action $S_T(\psi)$ for fixed T, i.e. it satisfies the Hamiltonian system of ODEs, $\dot{\psi} = H_\theta(\psi,\theta)$, $\dot{\theta} = -H_x(\psi,\theta)$, subject to some boundary conditions. Let $\varphi(\alpha) = \psi(G(\alpha))$ and differentiate it twice in α*

to get

$$\lambda\varphi' = \dot{\psi} \circ G \quad \textit{and} \quad \lambda^2\varphi'' + \lambda\lambda'\varphi' = \ddot{\psi} \circ G,$$

where $\lambda := 1/G'$. *Now use the Hamilton equations for* ψ *to obtain the following second-order ODE for* φ*:*

$$\begin{aligned} \ddot{\psi} &= H_{\theta x}\dot{\psi} + H_{\theta\theta}\dot{\theta} \\ &= H_{\theta x}\dot{\psi} - H_{\theta\theta}H_x \qquad (3.4) \\ \Leftrightarrow \qquad \lambda^2\varphi'' + \lambda\lambda'\varphi' &= H_{\theta x}\lambda\varphi' - H_{\theta\theta}H_x \\ \Leftrightarrow \qquad \lambda^2\varphi'' - \lambda H_{\theta x}\varphi' + H_{\theta\theta}H_x + \lambda\lambda'\varphi' &= 0. \qquad (3.5) \end{aligned}$$

The derivatives of H *have to be evaluated at* (φ, θ)*, which has to fulfill* $H(\varphi, \theta) = cst$ *and* $H_\theta(\varphi, \theta) = \dot{\psi} = \lambda\varphi'$.

This shows the following property of the steady state solutions of (3.3):

Lemma 4. *The flow in* (3.3) *has reached steady state if and only if*

(i) $\mu \equiv 0$*, and*

(ii) the functions ψ *corresponding to* (φ, λ) *in the sense of Proposition 3 (i.e. defined by pieces of* (φ, λ) *on which* $\lambda \neq 0$*) solve the second-order Hamiltonian ODE given by* H *on the energy level* $H = 0$*, i.e. they fulfill* (3.4) *and* $H \equiv 0$*, where* H *and all its derivatives are evaluated at* $\big(\psi, \theta^\star(\psi, \dot{\psi})\big)$.

Proof. Looking at the first representation of the flow in (3.3), we multiply the equation by $P_{\varphi'}$ and by $I - P_{\varphi'}$ to conclude that at steady state we must have

$\mu\varphi' \equiv 0$ (and thus $\mu \equiv 0$) and

$$P_{\varphi'}\big(\lambda^2\varphi'' - \lambda H_{\theta x}\varphi' + H_{\theta\theta}H_x\big) = 0.$$

But as shown in Proposition 4, this equation is the same as (3.5), which by Remark 6 is equivalent to the second-order Hamiltonian ODE (3.4) for $\psi(t) = \varphi(G^{-1}(t))$, where $G' = 1/\lambda$ with

$$\lambda = \big\langle \varphi', H_\theta\big(\varphi, \hat{\vartheta}(\varphi, \varphi')\big)\big\rangle / |\varphi'|^2 = \big\langle \varphi', \lambda(\varphi, \varphi')\varphi'\big\rangle / |\varphi'|^2 = \lambda(\varphi, \varphi'),$$

as in the construction of Proposition 3. The energy level is zero because we have

$$H\big(\psi, \theta^\star(\psi, \dot{\psi})\big) \circ G \quad = \quad H\big(\varphi, \theta^\star(\varphi, \varphi'\lambda)\big) = H\big(\varphi, \hat{\vartheta}(\varphi, \varphi')\big) \equiv 0.$$

□

The fact that at steady state we have $\mu \equiv 0$ also eliminates possible inaccuracies in the result of an algorithm based on discretizing (3.3), and it opens the door to methods for achieving second-order accuracy in α or higher even if the curve passes through critical points, see Section 3.3.

Also note that one could have designed the algorithm without the discussion in Chapter 2, solely based on the observation in Remark 6. Within that approach, however, to show that (3.3) converges one would then have to look for a corresponding Lyapunov function for this equation, which is in fact given by $\hat{S}(\varphi)$.

3.2 The outer loop

To solve (3.3) in practice, we discretize first $\varphi(\tau, \alpha)$ both in τ and α, i.e. we define $\varphi_i^k = \varphi(k\Delta\tau, i\Delta\alpha)$, $k \in \mathbb{N}_0$, $i = 0, \ldots, N$, where $\Delta\tau$ is the time step and $\Delta\alpha = 1/N$ if we discretize the curve into $N+1$ points. Then we discretize the initial condition $\varphi(0, \alpha)$ to obtain $\{\varphi_i^0\}_{i=0,\ldots,N}$ and, for $k \geq 0$, use the following two-step method to update these points.

1. Given φ_i^k and $\varphi_i'^k = (\varphi_{i+1}^k - \varphi_{i-1}^k)/(2/N)$, compute $\hat{\vartheta}_i^k = \hat{\vartheta}(\varphi_i^k, \varphi_i'^k)$,

$$\lambda_i^k = \frac{\langle H_\theta(\varphi_i^k, \hat{\vartheta}_i^k), \varphi_i'^k \rangle}{|\varphi_i'^k|^2},$$

   and finally $\lambda_i'^k = (\lambda_{i+1}^k - \lambda_{i-1}^k)/(2/N)$ for $i = 1, \ldots, N-1$.

2. Let $\{\tilde{\varphi}_i\}_{i=0,\ldots,N}$ be the solution of the linear system

$$\begin{cases} \dfrac{\tilde{\varphi}_i - \varphi_i^k}{\Delta\tau} = (\lambda_i^k)^2 \dfrac{\tilde{\varphi}_{i+1} - 2\tilde{\varphi}_i + \tilde{\varphi}_{i-1}}{1/N^2} - \lambda_i^k H_{\theta x} \varphi_i'^k \\ \qquad\qquad + H_{\theta\theta} H_x + \lambda_i^k \lambda_i'^k \varphi_i'^k, \qquad i = 1, \ldots, N-1, \\ \tilde{\varphi}_0 = x_1, \\ \tilde{\varphi}_N = x_2, \end{cases} \tag{3.6}$$

   where $H_{\theta x}$, $H_{\theta\theta}$ and H_x are evaluated at $(\varphi_i^k, \hat{\vartheta}_i^k)$.

3. Interpolate a curve across $\{\tilde{\varphi}_i\}_{i=0,\ldots,N}$ and discretize this curve to find $\{\varphi_i^{k+1}\}_{i=0,\ldots,N}$ so that the prescribed constraint on the parametrization of φ be satisfied.

4. Repeat until some stopping criterion is fulfilled.

Step 1 requires the computation of $\hat{\vartheta}(\varphi, \varphi')$. In the case of a diffusion, $\hat{\vartheta}(\varphi, \varphi')$ is given by (2.12). If $\hat{\vartheta}(\varphi, \varphi')$ is not available explicitly, it is computed using the algorithm given in Section 3.4 in an inner loop.

Step 2 uses semi-implicit updating for stability: as will be shown via the examples in Chapter 4, proceeding this way makes the time step $\Delta\tau$ required for stability independent of $\Delta\alpha = 1/N$ (in contrast, an explicit step would require $\Delta\tau = O(\Delta\alpha^2)$). As a result, it accelerates the convergence rate (see the discussion in Section 3.3) and, in effect, amounts to pre-conditioning appropriately the steepest descent scheme [23, 24]. In practice it turns out that it is not necessary to treat the term $\lambda\lambda'\varphi'$ (which, when written out, contains the term φ'' as well) implicitly, since changes of the curve φ in the direction of φ' do not carry any information. Notice also that Step 2 is computationally straightforward since λ_i^k is scalar: hence the linear system can be solved component by component using e.g. the Thomas algorithm [23]. Finally notice that a simple modification of (3.6) in Step 2 can be used to have the two endpoints of the curve fall into the nearest stable state, by setting

$$\frac{\tilde{\varphi}_i - \varphi_i^k}{\Delta\tau} = H_\theta(\varphi_i^k, 0) \qquad (i = 0, N). \tag{3.7}$$

Step 3 is the interpolation-reparametrization step used to enforce the constraint on the parametrization of the curve $\gamma(\varphi)$. For instance, if we parametrize the curve by normalized arclength so that $|\varphi'| = cst$, this step amounts to redistributing the images along the interpolated curve in such a way that the points

φ_i^k, $i = 0, \ldots, N$, be equidistant. Consistent with the order of accuracy at which we discretize the derivatives φ', φ'', etc., Step 3 can be done using linear interpolation which is second-order accurate if $\varphi \in C^1(0,1)$ (see the discussion about accuracy in Section 3.3).

Finally, the stopping criterion in Step 4 can be based on a slowdown in the movement of φ, or in the decay of the action $\hat{S}(\varphi)$. Plotting the function $\alpha \mapsto \lambda(\varphi(\alpha), \varphi'(\alpha))$ at each iteration can further help to determine whether the algorithm has already converged: if one knows that the curve that one is looking for has to pass a saddle point, then λ must have a root in $(0,1)$ by Lemma 1 (i) (see also Figure 4.3 in Section 4.1).

All in all, this algorithm is a blend between the original MAM [8] and the string method [7].

Recovering the parametrization, evaluating the action. Other quantities of interest include the actual value of the action, the optimal transition time $T^\star$, and the path $\psi^\star$ parametrized by physical time. To compute those, one can then add the following steps:

5. Given $\{\varphi_i^k\}_{i=0,\ldots,N}$, compute ${\varphi'}_i^k$, $\hat{\vartheta}_i^k$ and λ_i^k as in Step 1, for every $i = 0, \ldots, N$.

6. Return the action

$$\hat{S} = \frac{1}{2N} \sum_{i=1}^{N} \left(\langle {\varphi'}_i^k, \hat{\vartheta}_i^k \rangle + \langle {\varphi'}_{i-1}^k, \hat{\vartheta}_{i-1}^k \rangle \right).$$

7. Set $t_0 = 0$, and $t_i = \frac{1}{2\lambda_0^k} + \frac{1}{\lambda_1^k} + \cdots + \frac{1}{\lambda_{i-1}^k} + \frac{1}{2\lambda_i^k}$ for $i = 1, \ldots, N$.

8. Return the transition time $T^\star = t_N$.

9. To add $D+1$ points at equidistant times to the graph of φ (if $T^\star < \infty$), interpolate the function $G^{-1}(t)$ given by the points $(t, G^{-1}(t)) = (t_i, \frac{i}{N})$, $i = 0, \ldots, N$, at the values $\tilde{t}_d = \frac{d}{D} T^\star$, $d = 0, \ldots, D$, to obtain values $\alpha_d = G^{-1}(\tilde{t}_d)$, and then discretize the curve interpolated from $\{\varphi_i^k\}_{i=0,\ldots,N}$ at those values α_d.

If $T^\star$ is infinite or very large (i.e. if $\lambda_i^k \approx 0$ for some index i), Step 9 can be performed on trimmed values $\tilde{\lambda}_i^k := \max\{\lambda_i^k, \eta\}$ (for some small $\eta > 0$): Proposition 3 shows that this will still lead to representative dots away from the critical points (where $\lambda \geq \eta$); only close to the critical points (where $\lambda < \eta$) this will simply lead to dots that are equidistant *in space*.

3.3 Accuracy and efficiency of the outer loop

A rigorous discussion of the accuracy and efficiency of the gMAM is beyond the scope of the present paper. However, we find it useful to make a few heuristic comments.

The discussion is complicated by the fact that the minimum action path (i.e. the steady state solution $\varphi^\star$ of (3.3) which is also the minimizer of $\hat{S}(\varphi)$) will, in general, go through critical points, and the path may not be smooth at these points. We first discuss the case when this does not happen, i.e. when $\varphi^\star$ is smooth, then explain what to do if that is not the case.

If $\varphi^\star$ is smooth, we expect the algorithm to identify it with second-order accuracy in $\Delta\alpha = 1/N$. This is because the derivatives of φ and λ are computed by central differences, which are second-order accurate, and the linear interpolation in Step 3 is second-order accurate as well. This was confirmed in the example below. As long as we achieve stability (which requires to take a small enough value for $\Delta\tau$, but one that is independent of N), we observe that the error on the curve can be made $O(\Delta\alpha^2) = O(1/N^2)$, in the sense that

$$\rho(\varphi_{\text{interp}}, \varphi^\star) \leq CN^{-2} \tag{3.8}$$

for some constant $C > 0$, where φ_{interp} is the curve linearly interpolated from $\{\varphi_i^k\}_{i=0,\ldots,N}$ after convergence. Assuming linear convergence in time, the number of steps until convergence is then $O(\log N)$ which gives a total cost, measured in number of operations till convergence, scaling as

$$\text{cost} = O(N \log N). \tag{3.9}$$

Notice that this estimate takes into account that the interpolation step requires $O(N)$ operations if there are $N + 1$ discretization points along the curve. The estimate (3.9) was confirmed in our numerical examples, see Chapter 4.

Consider now what happens if the steady state solution of (3.3) (i.e. the solution of (3.2)) goes through one or more critical points and is not smooth at these points. Notice first that this leads to no problem with (3.2). Indeed, by Lemma 1 (i), if the point $\varphi(\alpha_c)$ is a critical point for some $\alpha_c \in [0, 1]$ then $\lambda \to 0$

as $\alpha \to \alpha_c$. All the terms in (3.2) involving φ', φ'' and λ' are multiplied by λ or λ^2 and so these products tend to 0 as $\alpha \to \alpha_c$. Furthermore, because of the Lipschitz continuity of λ their discretizations still approximate the correct value 0 at the critical point. However, they will do so only up to first-order accuracy, $O(\Delta\alpha) = O(1/N)$, unless one makes sure that one discretization point of the curve falls onto the critical point (in which case we obtain the exact value 0).

In the algorithm, this problem is also aggravated by the interpolation-reparametrization Step 3. The non-differentiability of the curve will reduce the order of accuracy of the linear interpolation procedure to first-order as well, unless we take some extra care of how we handle the critical points on the curve.

One possible way to solve these problems and restore the second-order accuracy is to identify the location(s) of the critical point(s) along the curve and treat the pieces on each side separately by the same basic algorithm above. This can be done on-the-fly using the following procedure. Specify a small threshold value for λ, say $\lambda_0 > 0$, such that if $\lambda < \lambda_0$ there is likely to be a critical point in the vicinity. Then let the curve evolve as above, but as soon as $\lambda < \lambda_0$ at some point, say $\varphi^k_{i^\star}$, split the curve into the two pieces at the left and right of $\varphi^k_{i^\star}$. Continue the algorithm with modified Steps 2 and 3: in Step 2 replace the equation for $\varphi^k_{i^\star}$ in (3.6) by

$$\frac{\tilde{\varphi}_{i^\star} - \varphi^k_{i^\star}}{\Delta\tau} = -H_{x\theta}H_\theta - HH_x \quad \Big(= -\tfrac{1}{2}\nabla_x\big(|H_\theta|^2 + H^2\big) \Big), \tag{3.10}$$

with all functions on the right evaluated at $(\varphi^k_{i^\star}, 0)$, so that $\varphi^k_{i^\star}$ is attracted by the critical point where $H_\theta(x,0)$ and $H(x,0)$ vanish; in Step 3 redistribute the points

on both parts of the curve separately, without changing $\tilde{\varphi}_{i^\star}$ (i.e. $\varphi_{i^\star}^{k+1} = \tilde{\varphi}_{i^\star}$). Observe that in the cases of an SDE or an SPDE with unit noise the modification (3.10) can be achieved by *setting* $\lambda_{i^\star}^k = 0$ after completion of Step 1 (to see this, set $\lambda = 0$ in (4.3) and (4.10) below).

This procedure, which we found to work well on all examples treated in Chapter 4 and which is further discussed in Section 4.1, restores the second-order accuracy in $\Delta\alpha = 1/N$ even if there are critical points along the curve. Note that there is no *a priori* difficulty to design more accurate schemes by using a higher-order stencil for the derivatives, choosing a higher-order interpolation method, and taking care of the critical points along the curve as explained above.

3.4 The inner loop (computing $\hat{\vartheta}(\varphi, \varphi')$)

In order to compute $\hat{\vartheta}(\varphi, \varphi')$ from (2.7) we must solve the following problem: Given the strictly convex and twice differentiable function $h(\cdot) = H(\varphi, \cdot)$ with $h(0) \leq 0$, and given a direction φ', we want to find the unique point $\hat{\vartheta}$ with

$$h(\hat{\vartheta}) = 0 \qquad \text{and} \qquad h_\theta(\hat{\vartheta}) = \lambda\varphi' \quad \text{for some} \quad \lambda \geq 0. \tag{3.11}$$

This problem has a simple geometric interpretation, as was illustrated before in Figure 2.1 (with $y = \varphi'$): it amounts to finding the point of the convex zero-level set of h where the normal to that level set is parallel to and points into the same direction as φ'.

Since the region $\{\theta \in \mathbb{R}^n \,|\, h(\theta) \leq 0\}$ can be potentially very thin and long,

one must make use of the underlying geometry of the problem. One very efficient strategy for finding a smart update for an initial guess $\hat{\vartheta}^0$ is a procedure similar in spirit to a higher-order version of the standard Newton-Raphson algorithm. However, while the Newton-Raphson method typically computes in each iteration the exact solution of the first-order approximation of the problem, we *must* use a second-order approximation since the solution of our problem is only well-defined for strictly convex functions h.

The procedure is thus as follows. For $p \geq 0$:

1. Compute $h(\hat{\vartheta}^p)$, $h_\theta(\hat{\vartheta}^p)$ and $h_{\theta\theta}(\hat{\vartheta}^p)$.

2. Find the unique quadratic function $f(\theta)$ such that $f(\hat{\vartheta}^p) = h(\hat{\vartheta}^p)$, $f_\theta(\hat{\vartheta}^p) = h_\theta(\hat{\vartheta}^p)$ and $f_{\theta\theta}(\hat{\vartheta}^p) = h_{\theta\theta}(\hat{\vartheta}^p)$.

3. If the region $\{\theta \in \mathbb{R}^n \mid f(\theta) < 0\}$ is non-empty, let $\hat{\vartheta}^{p+1}$ be the solution of

$$f(\hat{\vartheta}) = 0 \qquad \text{and} \qquad f_\theta(\hat{\vartheta}) = \lambda\varphi' \quad \text{for some} \quad \lambda \geq 0,$$

   i.e. of (3.11) with h replaced by its approximation f. Otherwise, let $\hat{\vartheta}^{p+1} := \operatorname{argmin}_\theta f(\theta)$.

4. Repeat until convergence.

Steps 1–3 in this procedure can be done analytically, and so it provides us with a closed-form update formula. The computation is carried out in Appendix F and

gives:

$$\hat{\vartheta}^{p+1} := \hat{\vartheta}^p + h_{\theta\theta}^{-1}\big(\tilde{\lambda}(\hat{\vartheta}^p)\varphi' - h_\theta\big) \quad \text{with} \quad \tilde{\lambda}(\hat{\vartheta}^p) := \left(\frac{\langle h_\theta, h_{\theta\theta}^{-1} h_\theta\rangle - 2h}{\langle \varphi', h_{\theta\theta}^{-1}\varphi'\rangle}\right)_+^{1/2}, \quad (3.12)$$

where $w_+^{1/2} = \sqrt{w}$ if $w \geq 0$ and $w_+^{1/2} = 0$ otherwise, and where h, h_θ and $h_{\theta\theta}$ are evaluated at $\hat{\vartheta}^p$. Note also that by definition of the algorithm, if h is quadratic to begin with, then the algorithm converges after only one iteration (since then $f = h$). This will happen if the underlying process is a diffusion process.

Once $\hat{\vartheta}$ has been determined, the value of λ in (3.11) can then be computed as a simple function of $\hat{\vartheta}$ via

$$\lambda = \frac{\langle h_\theta(\hat{\vartheta}), \varphi'\rangle}{|\varphi'|^2} = \frac{\langle H_\theta(\varphi, \hat{\vartheta}), \varphi'\rangle}{|\varphi'|^2}. \quad (3.13)$$

Next we show that the sequence generated by (3.12) has $\hat{\vartheta}$ as its unique fixed point and is quadratically convergent if h is smooth enough. The latter is not surprising since the standard Newton-Raphson algorithm has the same rate of convergence.

Lemma 5 (Uniqueness of fixed point $\hat{\vartheta}$)**.** *We have $\hat{\vartheta}^{p+1} = \hat{\vartheta}^p$ if and only if $\hat{\vartheta}^p = \hat{\vartheta}$, i.e. if $\hat{\vartheta}^p$ is the solution of the system (3.11). In that case, the value of λ in (3.11) is given by $\tilde{\lambda}(\hat{\vartheta}^p)$.*

Proof. "$\Leftarrow$": If $h(\hat{\vartheta}^p) = 0$ and $h_\theta(\hat{\vartheta}^p) = \lambda\varphi'$ for some $\lambda \geq 0$ then

$$\tilde{\lambda}(\hat{\vartheta}^p)\varphi' = \left(\frac{\langle h_\theta, h_{\theta\theta}^{-1} h_\theta\rangle - 2h}{\langle \varphi', h_{\theta\theta}^{-1}\varphi'\rangle}\right)_+^{1/2}\varphi' = \left(\frac{\langle h_\theta, h_{\theta\theta}^{-1} h_\theta\rangle}{\langle \varphi', h_{\theta\theta}^{-1}\varphi'\rangle}\right)_+^{1/2}\varphi' = \lambda\varphi' = h_\theta(\hat{\vartheta}^p),$$

so that $\hat{\vartheta}^{p+1} = \hat{\vartheta}^p$.

"$\Rightarrow$": Now let $\hat{\vartheta}^{p+1} = \hat{\vartheta}^p$. Then $h_\theta(\hat{\vartheta}^p) = \tilde{\lambda}(\hat{\vartheta}^p)\varphi'$ and clearly $\tilde{\lambda}(\hat{\vartheta}^p) \geq 0$, so it remains to show that $h(\hat{\vartheta}^p) = 0$. If $\beta := \langle h_\theta, h_{\theta\theta}^{-1} h_\theta \rangle - 2h \geq 0$ (where here and in the next line h, h_θ and $h_{\theta\theta}$ are evaluated at $\hat{\vartheta}^p$) then we can compute that

$$\langle h_\theta, h_{\theta\theta}^{-1} h_\theta \rangle = \tilde{\lambda}(\hat{\vartheta}^p)^2 \langle \varphi', h_{\theta\theta}^{-1} \varphi' \rangle = \langle h_\theta, h_{\theta\theta}^{-1} h_\theta \rangle - 2h \quad \Rightarrow \quad h(\hat{\vartheta}^p) = 0.$$

If $\beta \leq 0$, then $\tilde{\lambda}(\hat{\vartheta}^p) = 0$ and thus $h_\theta(\hat{\vartheta}^p) = 0$, i.e. $\hat{\vartheta}^p$ is the minimum of h. Since we know that $h(0) \leq 0$, this implies $h(\hat{\vartheta}^p) \leq 0$. On the other hand, $0 \geq \beta = -2h(\hat{\vartheta}^p)$, so $h(\hat{\vartheta}^p)$ must be zero. □

Lemma 6 (Quadratic convergence)**.** *If $h \in C^4(\mathbb{R}^n)$ then there exists a neighborhood $U_\epsilon(\hat{\vartheta})$ of the solution $\hat{\vartheta}$ and a constant $c > 0$ such that for $\forall \hat{\vartheta}^p \in U_\epsilon(\hat{\vartheta})$ we have*

$$|\hat{\vartheta}^{p+1} - \hat{\vartheta}| \leq c|\hat{\vartheta}^p - \hat{\vartheta}|^2.$$

Proof. Starting from (3.12), we write

$$|h_{\theta\theta}(\hat{\vartheta}^p)(\hat{\vartheta}^{p+1} - \hat{\vartheta})| = \left| h_{\theta\theta}(\hat{\vartheta}^p)(\hat{\vartheta}^p - \hat{\vartheta}) + \tilde{\lambda}(\hat{\vartheta}^p)\varphi' - h_\theta(\hat{\vartheta}^p) \right|. \quad (3.14)$$

We approximate the expression $\tilde{\lambda}(\hat{\vartheta}^p)\varphi' - h_\theta(\hat{\vartheta}^p)$ by its first-order Taylor expansion around $\hat{\vartheta}$ and estimate the remainder involving its second derivative (and thus the fourth derivative of h) by $O(|\hat{\vartheta}^p - \hat{\vartheta}|^2)$. Since the zeroth-order term vanishes, i.e. $\tilde{\lambda}(\hat{\vartheta})\varphi' - h_\theta(\hat{\vartheta}) = 0$ (this was shown in the first part of the proof of

Lemma 5), the right-hand side of (3.14) is equal to

$$\Big|h_{\theta\theta}(\hat{\vartheta}^p)(\hat{\vartheta}^p-\hat{\vartheta})+\big(\varphi'\otimes\nabla\tilde{\lambda}(\hat{\vartheta})-h_{\theta\theta}(\hat{\vartheta})\big)(\hat{\vartheta}^p-\hat{\vartheta})\Big|+O(|\hat{\vartheta}^p-\hat{\vartheta}|^2).$$

We show below that $\nabla\tilde{\lambda}(\hat{\vartheta})=0$, so that

$$\begin{aligned}\big|h_{\theta\theta}(\hat{\vartheta}^p)(\hat{\vartheta}^{p+1}-\hat{\vartheta})\big| &= \big|(h_{\theta\theta}(\hat{\vartheta}^p)-h_{\theta\theta}(\hat{\vartheta}))(\hat{\vartheta}^p-\hat{\vartheta})\big|+O(|\hat{\vartheta}^p-\hat{\vartheta}|^2)\\ &= O(|\hat{\vartheta}^p-\hat{\vartheta}|^2).\end{aligned}$$

Since $|\hat{\vartheta}^{p+1}-\hat{\vartheta}|\le|h_{\theta\theta}^{-1}(\hat{\vartheta}^p)|\,|h_{\theta\theta}(\hat{\vartheta}^p)(\hat{\vartheta}^{p+1}-\hat{\vartheta})|$, we are done.

To show that $\nabla\tilde{\lambda}(\hat{\vartheta})=0$, consider first the case $\lambda>0$, pick any $i\in\{1,\dots,n\}$ and use that at $\theta=\hat{\vartheta}$ we have $h=0$ and $h_\theta=\lambda\varphi'$:

$$\begin{aligned}&\partial_{\theta_i}\left(\frac{\langle h_\theta,h_{\theta\theta}^{-1}h_\theta\rangle-2h}{\langle\varphi',h_{\theta\theta}^{-1}\varphi'\rangle}\right)^{1/2}\Bigg|_{\theta=\hat{\vartheta}}\\ &\quad=\left(\frac{1}{2(\dots)^{1/2}}\,\partial_{\theta_i}\frac{\langle h_\theta,h_{\theta\theta}^{-1}h_\theta\rangle-2h}{\langle\varphi',h_{\theta\theta}^{-1}\varphi'\rangle}\right)\Bigg|_{\theta=\hat{\vartheta}}\\ &\quad=\frac{1}{2\lambda}\langle\varphi',h_{\theta\theta}^{-1}\varphi'\rangle^{-2}\Big[\Big(\langle h_\theta,(\partial_{\theta_i}h_{\theta\theta}^{-1})h_\theta\rangle\ +\underbrace{2h_\theta^Th_{\theta\theta}^{-1}h_{\theta\theta_i}-2h_{\theta_i}}_{=0}\Big)\langle\varphi',h_{\theta\theta}^{-1}\varphi'\rangle\\ &\qquad\qquad-\Big(\langle h_\theta,h_{\theta\theta}^{-1}h_\theta\rangle-\underbrace{2h}_{=0}\Big)\langle\varphi',(\partial_{\theta_i}h_{\theta\theta}^{-1})\varphi'\rangle\Big]\Big|_{\theta=\hat{\vartheta}}\\ &\quad=\frac{1}{2\lambda}\langle\varphi',h_{\theta\theta}^{-1}\varphi'\rangle^{-2}\Big[\langle h_\theta,(\partial_{\theta_i}h_{\theta\theta}^{-1})h_\theta\rangle\langle\varphi',h_{\theta\theta}^{-1}\varphi'\rangle-\langle h_\theta,h_{\theta\theta}^{-1}h_\theta\rangle\langle\varphi',(\partial_{\theta_i}h_{\theta\theta}^{-1})\varphi'\rangle\Big]\Big|_{\theta=\hat{\vartheta}}\\ &\quad=0.\end{aligned}$$

The case $\lambda=0$ can be treated by checking that in this case the function $\langle h_\theta,h_{\theta\theta}^{-1}h_\theta\rangle-2h$ as well as its first two derivatives vanish at $\theta=\hat{\vartheta}$, so that its square root is of order $o(|\hat{\vartheta}^p-\hat{\vartheta}|)$. □

Chapter 4

Examples

4.1 SDE: The Maier-Stein model

As a first test for our method, we use the following example of a diffusion process (SDE) first proposed by Maier and Stein [22]:

$$\begin{cases} du = (u - u^3 - \beta u v^2)dt + \sqrt{\varepsilon}\, dW_u(t) \\ dv = -(1 + u^2)v dt + \sqrt{\varepsilon}\, dW_v(t) \end{cases} \tag{4.1}$$

where W_u and W_v are independent Wiener processes, and $\beta > 0$ is a parameter. (In [22], Maier and Stein use two parameters: μ, which we set to 1 in this treatment, and α, which we call β in order to avoid confusion with the variable used to parametrize the path $\varphi(\alpha)$.)

For all values of $\beta > 0$, the SDE (4.1) has two stable equilibrium points at $(u, v) = (\pm 1, 0)$ and an unstable equilibrium point at $(u, v) = (0, 0)$ (see Figure 4.1). The drift vector field

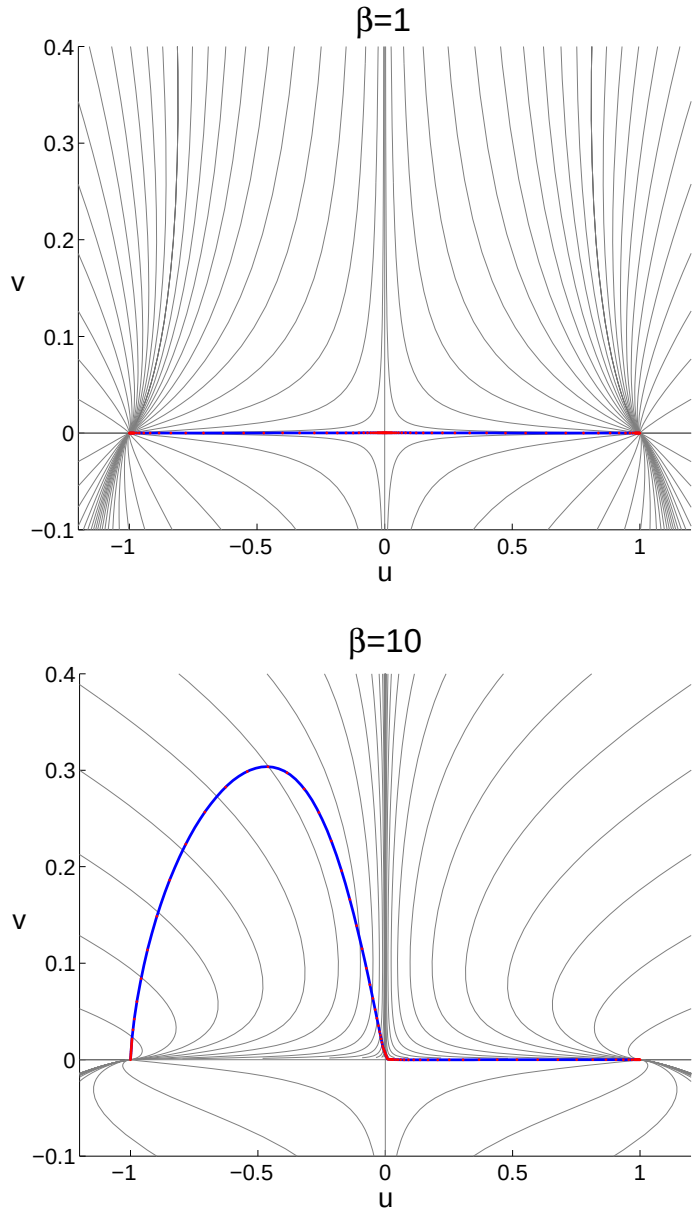

Figure 4.1: The minimum action paths from $(u, v) = (-1, 0)$ to $(u, v) = (1, 0)$ for the Maier-Stein model (4.1) shown on the top of the flow lines of the determinitic velocity field (gray lines). The parameters are $\beta = 1$ (first panel) and $\beta = 10$ (second panel). When $\beta = 1$, the minimum action path is simply the heteroclinic orbit joining $(\pm 1, 0)$ via $(0, 0)$; when $\beta = 10$, non-gradient effects take over, and the minimum action path is different from the heteroclinic orbit.

$$b(u,v) = \begin{pmatrix} u - u^3 - \beta u v^2 \\ -(1+u^2)v \end{pmatrix} \tag{4.2}$$

is the gradient of a potential if and only if $\beta = 1$.

When the noise amplitude ε is small, (4.1) displays bistability. Any initial condition with $u < 0$ is rapidly attracted toward a small neighborhood of $(u,v) = (-1,0)$ whereas any initial condition with $u > 0$ is rapidly attracted toward a small neighborhood of $(u,v) = (1,0)$. As a result, the equilibrium distribution of the process defined by (4.1) is concentrated in small neighborhoods around $(\pm 1, 0)$ and the process switches between these two regions only rarely. When it does so, large deviations theory tells us that, with probability 1 in the limit as $\varepsilon \to 0$, the trajectory remains in an arbitrarily small tube around the miminizer $\varphi^\star$ of $\hat{S}(\varphi)$ connecting $(u,v) = (-1,0)$ to $(u,v) = (1,0)$ or the other way around – in other words, the minimum action curve $\varphi^\star$ is the maximum likelihood pathway of switching (see Section 2.2). In addition, large deviations theory tells us that the frequency of these hopping events is roughly $\exp(-\varepsilon^{-1}\hat{S}(\varphi^\star))$.

Maier and Stein studied (4.1) for various values of β. They noted that the minimum action path from $(u,v) = (-1,0)$ to $(u,v) = (1,0)$ is the heteroclinic orbit joining these two points via $(u,v) = (0,0)$ when $\beta < \beta_{crit} = 4$ (this is consistent with the system being not too far from the gradient regime in these cases). However, when $\beta > \beta_{crit} = 4$, the piece of the minimum action path in the region $u < 0$ (i.e. in the basin of attraction of $(u,v) = (-1,0)$ by the deterministic dynamics) stops being the heteroclinic orbit. Some intuition for

why this change of behavior occurs can be gained by looking at the deterministic flow lines shown in Figure 4.1. Here we confirm these results using our method to find the minimum action path, as shown in Figure 4.1.

Before discussing the accuracy, stability and efficiency of our method in the context of the Maier-Stein model in detail, let us note that when applying our technique to a diffusion process such as (4.1) where the diffusion tensor is the identity, we can use the following explicit formulas for $\hat{\vartheta}$ and λ:

$$\hat{\vartheta} = \lambda\varphi' - b(\varphi), \qquad \lambda = \frac{|b(\varphi)|}{|\varphi'|}.$$

Since $H(x,\theta) = \langle b(x), \theta\rangle + \frac{1}{2}|\theta|^2$, we also have

$$H_{\theta x} = \nabla b, \qquad H_{\theta\theta} = I, \qquad H_x = (\nabla b)^T\hat{\vartheta} = (\nabla b)^T(\lambda\varphi' - b),$$

and so the equation (3.3) can be written explicitly as

$$\dot{\varphi} = \lambda^2\varphi'' - \lambda(\nabla b - (\nabla b)^T)\varphi' - (\nabla b)^T b + \lambda\lambda'\varphi' + \mu\varphi'. \tag{4.3}$$

Equation (4.3) can be integrated using a straight forward modification of the algorithm presented in Section 3.2. If b is a gradient field, $b = -\nabla U$, then $\nabla b - (\nabla b)^T = 0$ and the right-hand side simplifies further:

$$\dot{\varphi} = \lambda^2\varphi'' - \nabla\nabla U\,\nabla U + \lambda\lambda'\varphi' + \mu\varphi'. \tag{4.4}$$

The steady state of this equation is the minimum energy path, i.e. the path such

that $\nabla U^{\perp} = 0$ along it. Using gMAM to integrate (4.4) may represent a useful alternative to the string method [7].

Stability, accuracy and efficiency. We discuss the case $\beta = 10$ when the minimum action path is nontrivial. To obtain a benchmark solution, we first ran the algorithm with $N = 10^5$ discretization points at decreased stepsize to obtain a curve which we regarded as the true solution connecting $(-1, 0)$ and $(1, 0)$ (to get this benchmark, we actually ran the code between $(-1, 0)$ and $(0, 0)$ and then extended the path by the straight line between $(0, 0)$ and $(1, 0)$ since we know that piece exactly).

Next we ran the code for 300 iterations for $N = 100, 200, \ldots, 900, 1000, 2000, \ldots, 10000$ at a fixed time step $\Delta\tau = 0.1$, to find both the minimum action path connecting $(-1, 0)$ to $(0, 0)$ and the one connecting $(-1, 0)$ to $(1, 0)$ via the critical point $(0, 0)$. For the initial condition, we used a semicircle in the upper half plane connecting the critical points $(-1, 0)$ and $(0, 0)$ in the first case, or $(-1, 0)$ and $(1, 0)$ in the second case. The error was estimated by computing the maximum distance of each of the curves interpolated between the points to the benchmark curve obtained before. (Note that we took as many as 300 iterations because we are interested in measuring the accuracy of the algorithm here, but convergence is already achieved up to an error of 10^{-6} in the action after less than 20 iterations, see Figure 4.4.)

Since the left part of the path is smooth (there is no critical point along the path), we expect that the unmodified algorithm of Section 3.2 be second-order accurate in N if we seek for the minimum action path connecting $(-1, 0)$ to $(0, 0)$:

this is confirmed by results shown in Figure 4.2a. On the other hand, since the path connecting $(-1,0)$ to $(1,0)$ has to pass the critical point $(0,0)$ and is not differentiable at that point, we expect the unmodified algorithm of Section 3.2 to be only first-order accurate for these runs, and this is confirmed by the upper curve in Figure 4.2b. However, when we modified the algorithm as proposed in Section 3.3 and ran it for an additional 300 steps with the reparametrization step treating the right and the left side separately, second-order accuracy was restored, as shown by the lower curve in Figure 4.2b.

To check convergence of the curve between $(-1,0)$ and $(1,0)$, we also plotted $\lambda(\varphi(\alpha), \varphi'(\alpha))$, see Figure 4.3. As expected, after convergence it has one root in the interval $(0,1)$, corresponding to the critical point on the curve φ. The plot also confirms Lemma 3 (i), which says that λ is Lipschitz continuous at that point. Furthermore, we observed that with the unmodified algorithm the value of λ at that point only reaches $O(N^{-1})$ (since the points of the discretized curve φ only approximate the critical point to that order), whereas our method to achieve second-order accuracy brings it all the way down to zero (up to machine precision).

The time step to achieve stability in all the runs was found to be largely independent of the choice of N. For the gridsizes N as above, the maximum value for $\Delta\tau$ before visible oscillations occured stayed constant at about 0.3.

The required number of iterations at $\Delta\tau = 0.2$ until the change of the action per iteration became less than 10^{-7} varies insignificantly between 31 and 33 for the gridsizes N as above. For fixed N, the action decreases exponentially to its limiting value, as can be seen in Figure 4.4 which shows the decay for the various

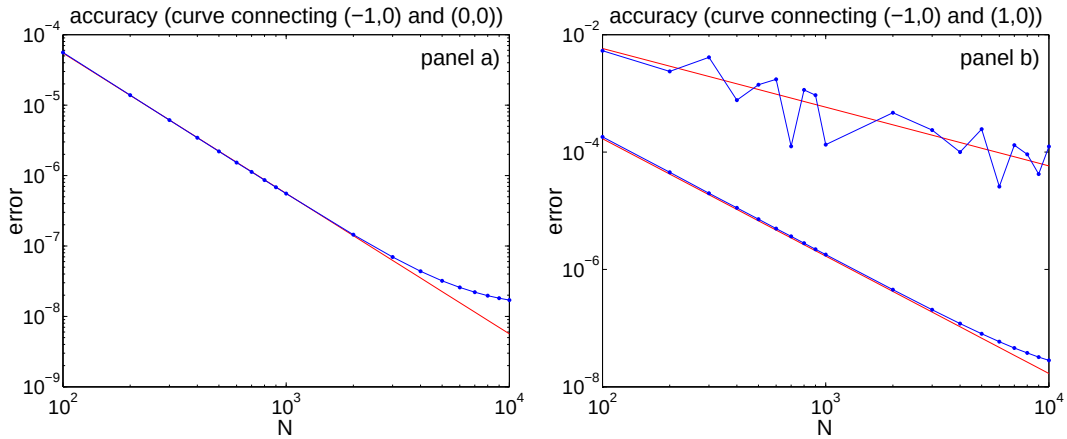

Figure 4.2: The accuracy measurements for the Maier-Stein model on double-logarithmic scale, for the left path (panel a) and the whole path (panel b, using the modified and the unmodified algorithm). The red lines have slope -1 (panel b, upper curve) and -2, indicating accuracies of order $O(1/N)$ and $O(1/N^2)$, respectively. The noise in the measurements of the unmodified algorithm on the whole path (panel b, upper curve) is due to the fact that the error strongly depends on whether a grid point happens to lie close to the unstable equilibrium point; however, modifying the algorithm as explained in Section 3.3 restores the second-order accuracy.

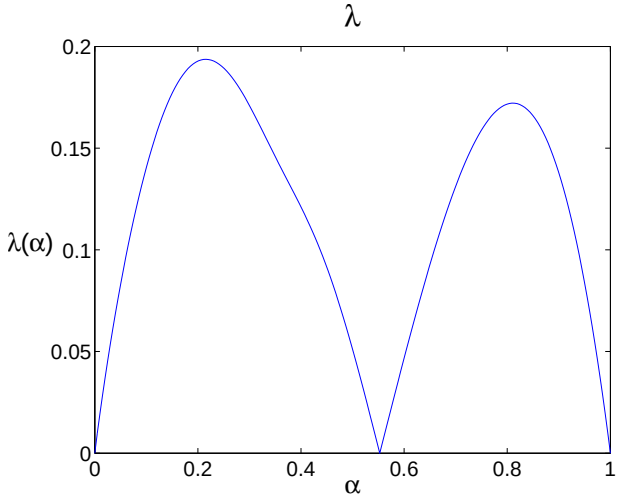

Figure 4.3: The function $\lambda(\varphi(\alpha), \varphi'(\alpha))$. The root in the middle corresponds to the value α where $\varphi(\alpha)$ is a critical point.

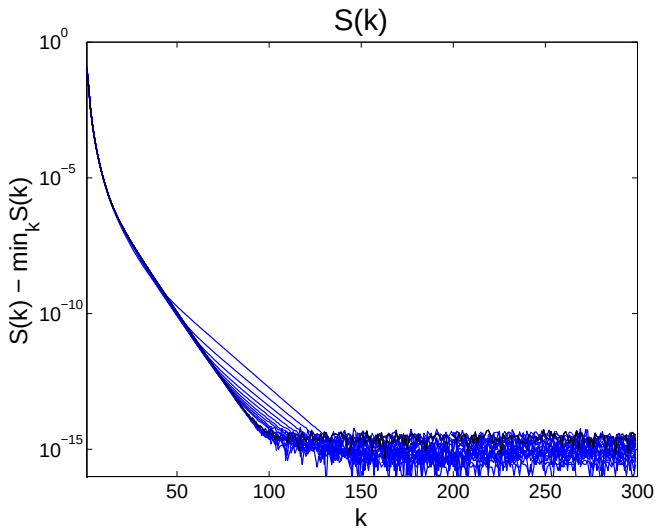

Figure 4.4: The error in the action plotted in function of the number of iterations on semilogarithmic scale, for various values of N between 100 and 10,000. The black curve corresponds to $N = 10,000$. These graphs indicate linear convergence, with a rate which is independent of N since the timestep $\Delta\tau$ is independent of N.

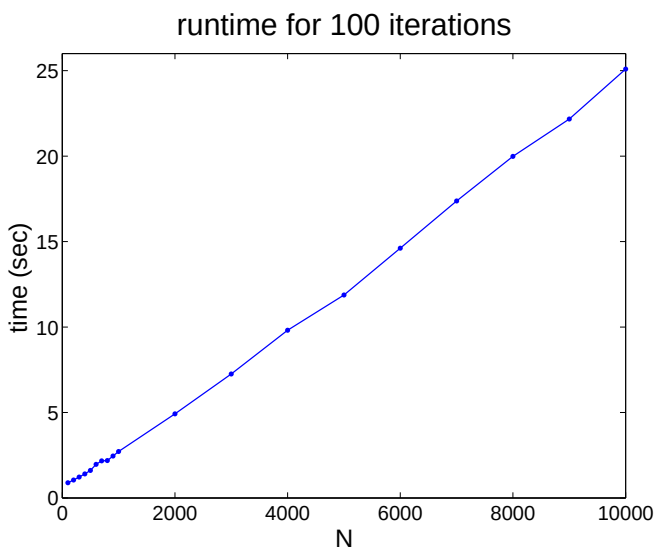

Figure 4.5: The runtime for 100 iterations at various gridsizes (using MatLab 6.5 running under Windows XP on a 1.5 GHz Pentium 4), showing linear dependency in N.

values of N in a semilogarithmic plot. The runtime for 100 iterations for various gridsizes is plotted in Figure 4.5. It shows linear dependency on N, which is due to the fact that all of our operations have a cost of order $O(N)$, including solving the linear system in Step 2 and the linear interpolation in Step 3. These observations are consistent with the estimate (3.9) for the cost.

4.2 SPDE: An SPDE generalization of the Maier-Stein model

4.2.1 One dimension

As a natural generalization of the SDE (4.1), we consider the following SPDE analogue of this equation (here written as a standard PDE for the sake of clarity):

$$\begin{cases} u_t = \kappa u_{xx} + u - u^3 - \beta u v^2 + \sqrt{\varepsilon}\,\eta_u(x,t), \\ v_t = \kappa v_{xx} - (1+u^2)v + \sqrt{\varepsilon}\,\eta_v(x,t). \end{cases} \tag{4.5}$$

Here $x \in [0,1]$ and we assume periodic boundary conditions. $\kappa > 0$ is an additional parameter, and $\eta_u(x,t)$, $\eta_v(x,t)$ are spatio-temporal white-noises (i.e. the space-time derivatives of Brownian sheets, $W_u(x,t)$ and $W_v(x,t)$, defined on $(x,t) \in [0,1] \times [0,\infty)$). The system (4.5) is formal, but it can be shown (see [11]) by rewriting it in integral form that its solutions are well-defined and Hölder continuous and define a Markov process adapted to the filtrations of $W_u(t,x)$ and $W_v(t,x)$. In addition, it was shown in [11] that (4.5) satisfies a large deviations

principle with action functional

$$S_T(u,v) = \frac{1}{2}\int_0^T \int_0^1 \Big((u_t - \kappa u_{xx} - u + u^3 + \beta u v^2)^2 \\ + \big(v_t - \kappa v_{xx} + (1+u^2)v\big)^2 \Big)\, dx\, dt. \tag{4.6}$$

Thus, like its finite-dimensional analogue (4.1), (4.5) will display bistability in the limit as $\varepsilon \to 0$ in the sense that the invariant measure of the process defined by (4.5) is concentrated in a small neighborhood around the two stable equilibrium solutions of the deterministic equation obtained by setting $\varepsilon = 0$ in (4.5): these are $\big(u_\pm(x), v_\pm(x)\big) \equiv (\pm 1, 0)$. Here we are interested in analyzing the pathways of transition between these points which, with probability 1 as $\varepsilon \to 0$, are located in a small tube around the minimizer of the action (4.6) over both $(u(t,x), v(t,x))$ and T.

By analogy with what happens in the finite-dimensional system, we expect that when the system is not too far from gradient, i.e. when β is small enough, the minimum action path will follow the graph of a heteroclinic orbit connecting (u_-, v_-) and (u_+, v_+). The only difference with the finite-dimensional situation is then that, if the coefficient κ in (4.5) is small enough, $\kappa < \kappa_{crit} = \frac{1}{4\pi^2} \approx 0.0253$, there are many such orbits because (4.5) has many unstable equilibrium points. As a result there will be several minimum action paths (one global minimizer and several local minimizers). How to identify these unstable critical points as a way to benchmark the results from the minimum action method is explained below.

On the other hand, if the system is far from gradient, i.e. if β is large enough, then we expect that the piece of the minimum action path connecting the stable

equilibrium point (u_-, v_-) to an unstable equilibrium point will be different from the heteroclinic orbit connecting these points. Our results below confirm this intuition.

It is worth pointing out that traditional shooting methods to solve the Hamilton equations associated with the minimization of (4.6) are inapplicable here. The reason is that, unlike their finite-dimensional analogue, these equations are only well-posed as a boundary value problem in time, which prohibits the use of the shooting method. Hence, the technique used by Maier and Stein in [22] in the context of the finite-dimensional diffusion (4.1) cannot be used to obtain the minimizers of (4.6). Next we show that the gMAM is the right alternative to do this minimization.

The gMAM in infinite dimensions. We want to apply our algorithm to find the minimum action path connecting the stable states (u_-, v_-) and (u_+, v_+). In order to do so, we first recast our theoretical results to the present infinite-dimensional setting. Basically, this amounts to changing the finite-dimensional inner product $\langle \cdot, \cdot \rangle_a$ by its analogue in function space, and the only result we actually need is the equivalent of (2.2) and (2.14) which we state without proof:

$$V\big((u_-, v_-), (u_+, v_+)\big) = \inf_{\varphi} \hat{S}(\varphi) \tag{4.7}$$

where the infimum is taken over all spatially-periodic functions $\varphi(x, \alpha) : [0, 1] \times [0, 1] \to \mathbb{R}^2$ subject to $\varphi(\cdot, 0) \equiv (-1, 0)$ and $\varphi(\cdot, 1) \equiv (1, 0)$, and where

$$\hat{S}(\varphi) = \int_0^1 \left(\|\varphi'\|_{L^2} \|B(\varphi)\|_{L^2} - \langle \varphi', B(\varphi) \rangle_{L^2} \right) d\alpha. \tag{4.8}$$

Here

$$B(\varphi) := b(\varphi) + \kappa \varphi_{xx}, \tag{4.9}$$

where b is given by (4.2).

The steepest descent flow associated with (4.7) is the analogue of (4.3). It can be written in compact form as

$$\dot{\varphi} = \lambda^2 \varphi'' - \lambda(\partial B - (\partial B)^*)\varphi' - (\partial B)^* B + \lambda\lambda' \varphi' + \mu \varphi', \tag{4.10}$$

where $\lambda = \|B(\varphi)\|_{L^2} / \|\varphi'\|_{L^2}$, and ∂B is the operator

$$\partial B = \nabla b(\varphi) + \kappa \partial_x^2. \tag{4.11}$$

Explicitly, (4.10) is

$$\begin{aligned} \dot{\varphi} = \lambda^2 \varphi'' - \lambda(\nabla b - (\nabla b)^T)\varphi' - (\nabla b)^T b - \kappa(\nabla b + (\nabla b)^T)\varphi_{xx} \\ - \kappa^2 \varphi_{xxxx} - \kappa \begin{pmatrix} \langle \varphi_x, \nabla\nabla b_1 \, \varphi_x \rangle \\ \langle \varphi_x, \nabla\nabla b_2 \, \varphi_x \rangle \end{pmatrix} + \lambda\lambda' \varphi' + \mu \varphi'. \end{aligned} \tag{4.12}$$

(4.12) can be solved by discretizing $\varphi(x, \alpha, \tau)$ in x, α and τ and using a generalization of the algorithm in Section 3.2. There is, however, an additional difficulty caused by the presence of the spatial derivatives such as φ_{xxxx}. To stabilize the code with respect to those, we may use a FFT-based pseudo-spectral code in x

and Duhamel's principle to solve $\dot{\varphi} + \kappa^2 \varphi_{xxxx} =$ *(remaining terms)* explicitly. However, having tried this approach, we found that a slightly more efficient alternative was to not go pseudo-spectral but rather split each iteration step into two, the first one being the equivalent of the semi-implicit Step 2 (evaluating φ'' at the new time step) in which the term $\kappa^2 \varphi_{xxxx}$ at the right-hand side of (4.12) was excluded, and the second one being an implicit step with that term only. In this approach all spatial derivatives of φ were estimated by finite differences.

To apply the gMAM, we initialized the transition path as the linear interpolation between (u_-, v_-) and (u_+, v_+) and added a bump to break the degeneracy due to the periodicity in x, i.e. we set

$$\varphi\big(\tau = 0, x, \alpha\big) := \big(-1 + 2\alpha + 2\sin^2(\pi\alpha)\sin^2(\pi x), 0\big).$$

We chose the gridsizes $\Delta x = 1/128$ and $\Delta\alpha = 1/100$, and set the stepsize to $\Delta\tau = 0.1$. We then started the algorithm with the model parameters $\beta = 1$ and $\kappa = 0.01$, and after 40 seconds and about 120 iterations we obtained an approximate solution. Using a continuation method, we decreased κ to $\kappa = 0.003$ and then to $\kappa = 0.001$, each time running the algorithm starting from the previous solution. The whole sequence of operations took about two minutes using MatLab 6.5 running under Windows XP on a 1.5 GHz Pentium 4. Then we started the process over again, this time increasing κ until it reached $\kappa = 0.026 > \kappa_{crit}$.

Results. Figure 4.6 shows plots for $\beta = 1$ and various values of κ. Each of them shows the first components of the functions $\varphi(\cdot, \alpha)$ for equidistant values of α as

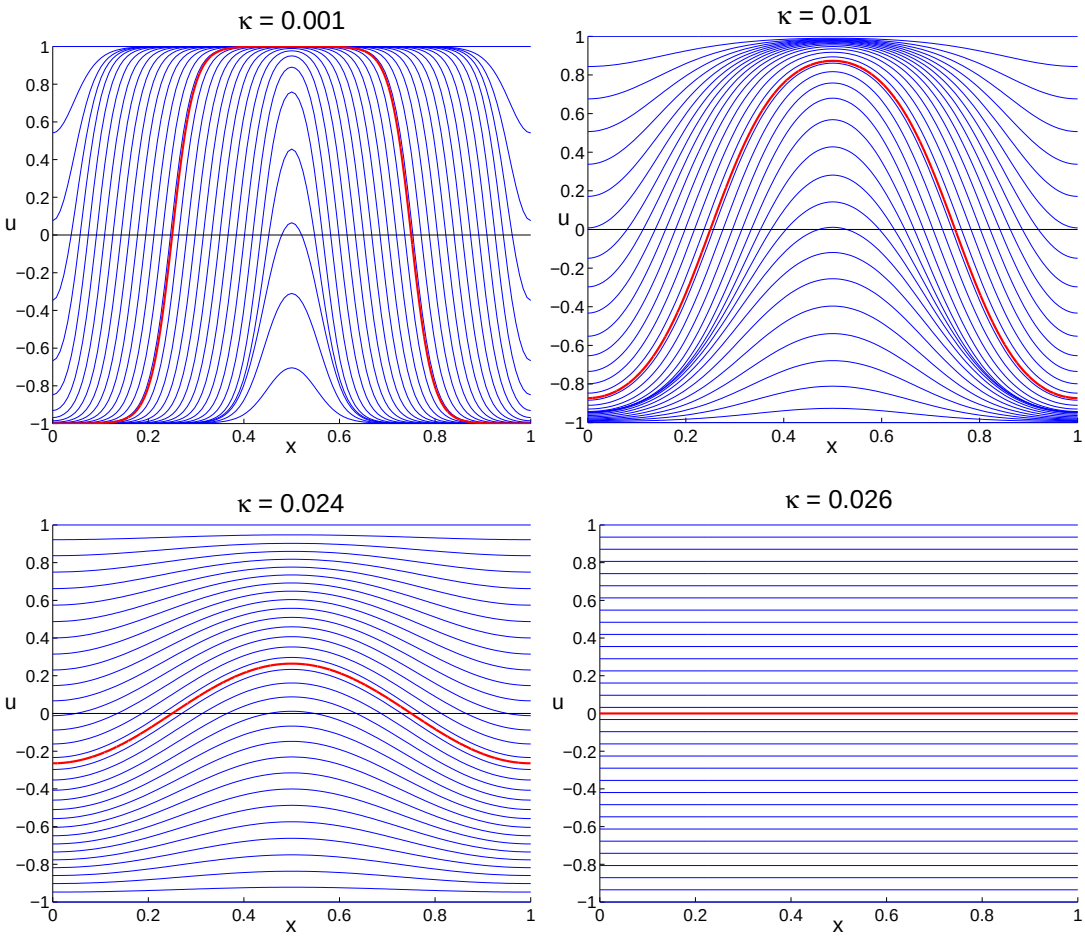

Figure 4.6: Snapshots along the minimum action path from (u_-, v_-) to (u_+, v_+) for the SPDE generalization of the Maier-Stein model. The parameters are $\beta = 1$ and $\kappa = 0.001$, 0.01, 0.024 and 0.026. The red lines are the unstable equilibrium points.

blue lines. To check our results, we added in each figure the lowest energy saddle point of the system as a red line, determined independently using the method explained below. One can see that the transition paths found by our algorithm indeed pass through these saddle points to very satisfying accuracy. Figure 4.7 shows a three-dimensional representation of the solution for $\kappa = 0.01$.

We then added a little bump also to the v-field (the second component of φ), set $\beta = 10$ and restarted the algorithm. For this value of β, by analogy with what happens in the finite-dimensional Maier-Stein model, we expected that the field v will assist in the transition during the uphill path. The gMAM confirmed this intuition, as shown in Figure 4.8. Now as u makes a transition similar to the previous one, v also increases around $x = \frac{1}{2}$ but vanishes again as the saddle point is reached.

Figure 4.9 shows the plot for $\kappa = 0.001$. As we can see, smaller values for κ lead to steeper domain walls. Finally, Figure 4.10 shows an example for a two-periodic local minimizer, obtained by starting from a two-periodic initial curve.

Notice that the minimum action paths in this example are degenerate due to the spatial periodicity (i.e. if $\varphi(\tau, x, \alpha)$ is a minimizer of the action, so is $\varphi(\tau, x + c, \alpha)$ for any $c \in \mathbb{R}$). This degeneracy, however, was broken by our choice of initial condition and does not appear to affect the convergence of the algorithm.

Finding saddle points. Since we know that the minimum action paths must go through critical points (such as saddle points), to check our results we computed these critical points using the following strategy. Any critical point of the

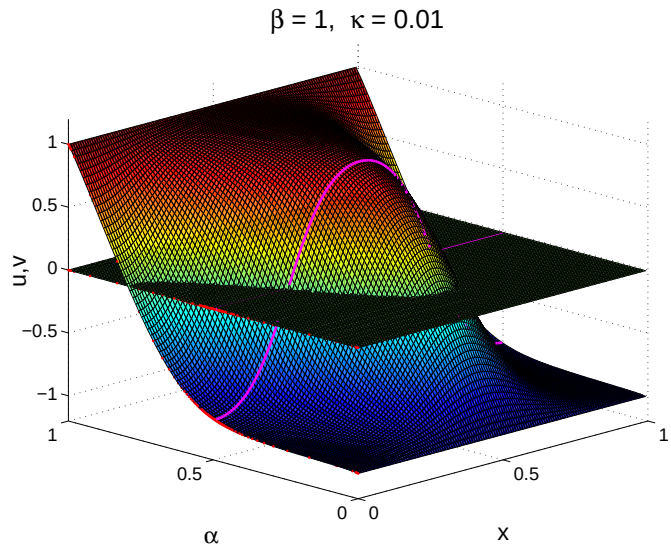

Figure 4.7: The minimum action path from (u_-, v_-) to (u_+, v_+) for $\kappa = 0.01$ and $\beta = 1$. The two surfaces represent the two components u and v of the minimizer $\varphi^\star$, the pink line shows the saddle point, and the red dots at the side mark points at equidistant times.

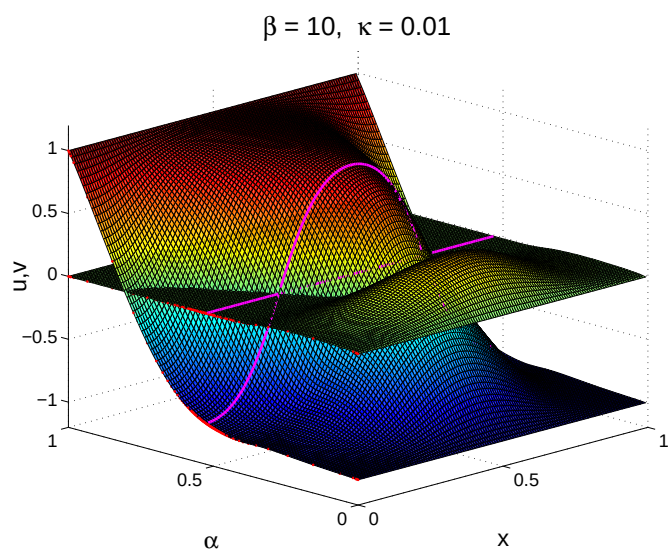

Figure 4.8: Same as in Figure 4.7, for $\kappa = 0.01$ and $\beta = 10$. For those parameters the field v assists in the transition.

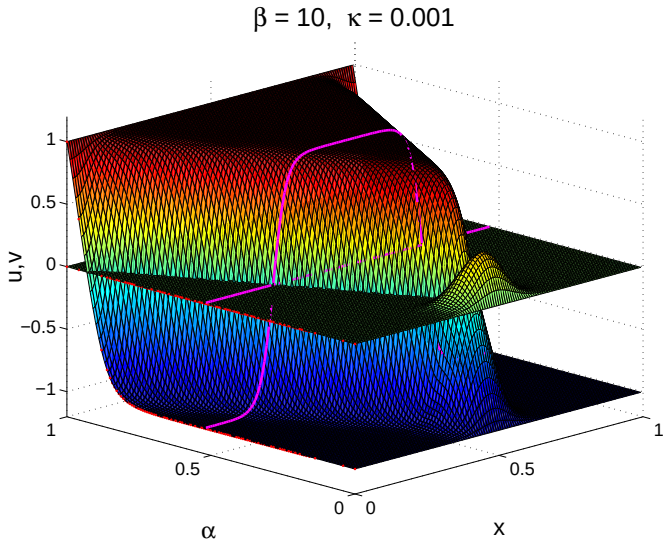

Figure 4.9: Same as in Figure 4.7, for $\beta = 10$ and $\kappa = 0.001$. In comparison with Figure 4.8 we observe that smaller values for κ lead to steeper domain walls.

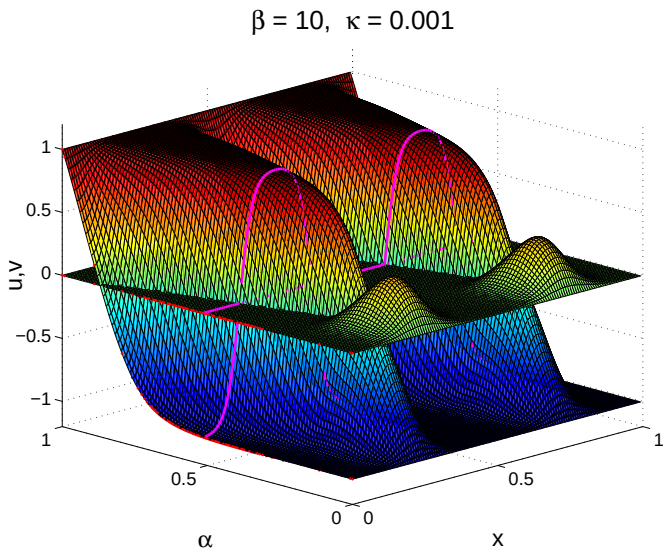

Figure 4.10: A local minimizer with period $\frac{1}{2}$ for $\beta = 10$ and $\kappa = 0.001$.

form $(u(x), 0)$ must fulfill the equation

$$0 = B_1\big((u,0)\big) = b_1\big((u,0)\big) + \kappa u_{xx} = u - u^3 + \kappa u_{xx}. \tag{4.13}$$

This equation has three constant solutions $u_k(x) \equiv (k, 0)$ for $k = -1, 0, +1$. The functions $(u_{\pm 1}, 0)$ can be shown to be stable states whereas $(u_0, 0)$ is an unstable critical point. To find nonconstant solutions of (4.13), multiply this equation by u_x and integrate, to obtain

$$-\tfrac{1}{4}(1-u^2)^2 + \tfrac{1}{2}\kappa u_x^2 = cst =: -\tfrac{1}{4}E^2,$$

$E \in [0, 1]$, or equivalently

$$|u_x(u)| = \kappa^{-1/2}\sqrt{\tfrac{1}{2}(1-u^2)^2 - \tfrac{1}{2}E^2}.$$

Additional solutions can thus be obtained by inverting the function

$$x(u) = \int_{u_-}^{u} \frac{1}{u_x(u')}\, du', \qquad u_- \le u \le u_+, \tag{4.14}$$

where $u_\pm$ are the locations at which $u_x = 0$, i.e. $u_\pm = \pm\sqrt{1-E}$, and then setting $u(x) := u(2x(u_+) - x)$ for $x(u_+) \le x \le 2x(u_+)$. For every choice of E this leads to a solution with period $p(E) = 2x(u_+)$, but we are only interested in those values for which $1/p(E) \in \mathbb{N}$ since $u(x)$ must be periodic on the original domain $x \in [0, 1]$. These values of E can be found from (4.14):

$$p(E) = 2\int_{u_-}^{u_+} \frac{1}{u_x(u)}\,du = 2\sqrt{\kappa}\int_0^1 u^{-1/2} f_E(u)\,du,$$

where

$$f_E(u) = (1-\tfrac{1}{2}u)^{-1/2}\Big(\big(\tfrac{1}{2}(1-E)u+2E\big)^{-1/2} + \big((1-E)(1-\tfrac{1}{2}u)+2E\big)^{-1/2}\Big).$$

We can compute the integral for $p(E)$ numerically for several values of $E \in [0,1]$ and then determine for which E_0 we have $p(E_0) = 1$. Finally, we can invert the solution $x(u)$ for $E = E_0$ to find the corresponding $u_{E_0}(x)$.

As a final remark, note that $p(E)$ can be shown to take its minimum at $E = 1$, and its value there is

$$\begin{aligned} p(1) &= 2\sqrt{2\kappa}\int_0^1 u^{-1/2}(1-\tfrac{u}{2})^{-1/2}\,du \\ &= 4\sqrt{\kappa}\int_0^{1/2} u^{-1/2}(1-u)^{-1/2}\,du \\ &= 2\sqrt{\kappa}\int_0^1 u^{-1/2}(1-u)^{-1/2}\,du = 2\pi\sqrt{\kappa}. \end{aligned}$$

Therefore, if $\kappa > \kappa_{crit} := \frac{1}{4\pi^2}$ then for every $E \in [0,1]$ we have $p(E) \geq p(1) = 2\pi\sqrt{\kappa} > 1$, i.e. $1/p(E) \notin \mathbb{N}$, and there is no non-constant critical point (with $v = 0$).

4.2.2 Two dimensions

Next we give a quick demonstration showing that even a two-dimensional SPDE can be treated by the gMAM within reasonable computation time. The two-

dimensional version of the SPDE (4.5) that we will treat in this section is

$$\begin{cases} u_t = \kappa\Delta u + u - u^3 - \beta u v^2 + \sqrt{\varepsilon}\,\eta_u(x,y,t), \\ v_t = \kappa\Delta v - (1+u^2)v + \sqrt{\varepsilon}\,\eta_v(x,y,t), \end{cases} \tag{4.15}$$

where $\Delta = \partial_x^2 + \partial_y^2$, $x, y \in [0,1]$ and we assume periodic boundary conditions in both x and y. This leads to the action

$$\begin{aligned} S_T(u,v) = \frac{1}{2}\int_0^T \int_0^1 \int_0^1 \Big(&(u_t - \kappa\Delta u - u + u^3 + \beta u v^2)^2 \\ &+ \big(v_t - \kappa\Delta v + (1+u^2)v\big)^2 \Big)\, dx\, dy\, dt. \end{aligned} \tag{4.16}$$

Then for

$$B(\varphi) = b(\varphi) + \kappa\Delta\varphi \qquad \text{and} \qquad \partial B = \nabla b(\varphi) + \kappa\Delta$$

the formulas (4.7) and (4.8) for the associated minimization problem and the formula (4.10) for the steepest descent flow are still valid. Explicitly, the latter now reads

$$\begin{aligned} \dot\varphi = \lambda^2\varphi'' - \lambda(\nabla b - (\nabla b)^T)\varphi' - (\nabla b)^T b - \kappa(\nabla b + (\nabla b)^T)\Delta\varphi - \kappa^2\Delta\Delta\varphi \\ - \kappa\left[\begin{pmatrix} \langle \varphi_x, \nabla\nabla b_1\, \varphi_x\rangle \\ \langle \varphi_x, \nabla\nabla b_2\, \varphi_x\rangle \end{pmatrix} + \begin{pmatrix} \langle \varphi_y, \nabla\nabla b_1\, \varphi_y\rangle \\ \langle \varphi_y, \nabla\nabla b_2\, \varphi_y\rangle \end{pmatrix}\right] + \lambda\lambda'\varphi' + \mu\varphi'. \end{aligned} \tag{4.17}$$

We again implemented both a pseudospectral solver and the two-step approach and found both to work well. We set $N = 80$, $\Delta x = \Delta y = \frac{1}{64}$ and $\Delta\tau = 0.02$, chose the initial condition

$$\varphi_0(\tau = 0, \alpha, x, y) = \begin{pmatrix} 2\alpha^2 - 1 + 2\sin^2(\pi\alpha)\sin^2(\pi x)\sin^2(\pi y) \\ 0 \end{pmatrix},$$

and started the gMAM with the parameters $\kappa = 0.01$ and $\beta = 1$. After an overnight run the algorithm converged. Using the continuation method, we then tried all combinations of values for $\kappa \in \{0.01,\ 0.003,\ 0.001\}$ and $\beta \in \{1,\ 3,\ 6,\ 10\}$, where for $\beta \geq 6$ we introduced a bump in the second component of φ by setting it to $0.1\alpha(1-\alpha)$.

Results. Figures 4.11 - 4.15 show some local minimizers that we found using the gMAM. Figure 4.11 shows the u-field of a local minimizer for $(\beta, \kappa) = (1,\ 0.001)$ (the v-field remains zero). The figures are to be read starting from the top left and then proceeding clockwise. The larger square on the right is the saddle point. Figures 4.12 and 4.13 show a local minimizer for $(\beta, \kappa) = (10,\ 0.001)$; as one can see, the v-field assists during the nucleation at the beginning of the transition, but it stays inactive throughout the propagation of the domain walls. The transition of the u-field on the other hand appears to be almost identical to the one for $\beta = 1$ in Figure 4.11. Figure 4.14 shows a local minimizer for $(\beta, \kappa) = (1,\ 0.01)$, illustrating that higher values of κ lead to less steep domain walls (i.e. to blurrier pictures).

Finally, note that global minimizers are constant with respect to either x or y; the problem then reduces to the one-dimensional one from the previous section. Figure 4.15 shows an interesting local minimizer that is a combination of two global minimizers that are constant in the x- and in the y-direction, respectively.

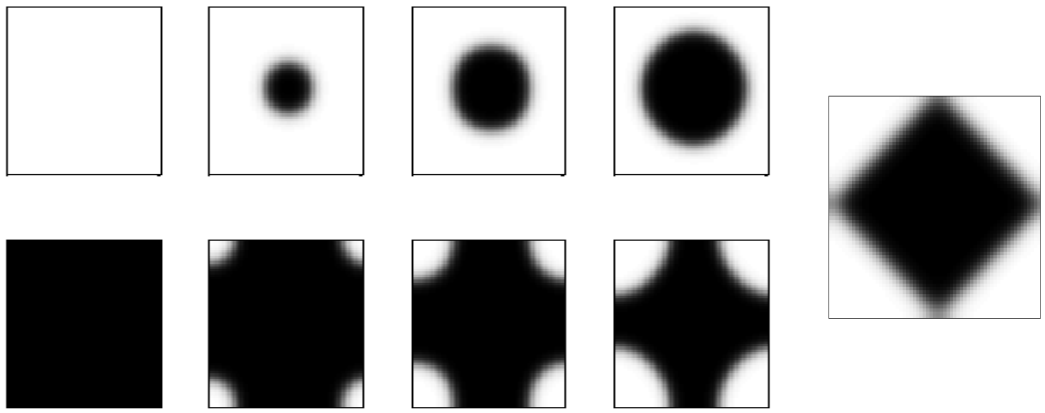

Figure 4.11: The u-field of a local minimizer for the two-dimensional SPDE, with $\beta = 1$ and $\kappa = 0.001$. The corresponding v-field remains zero throughout the transition.

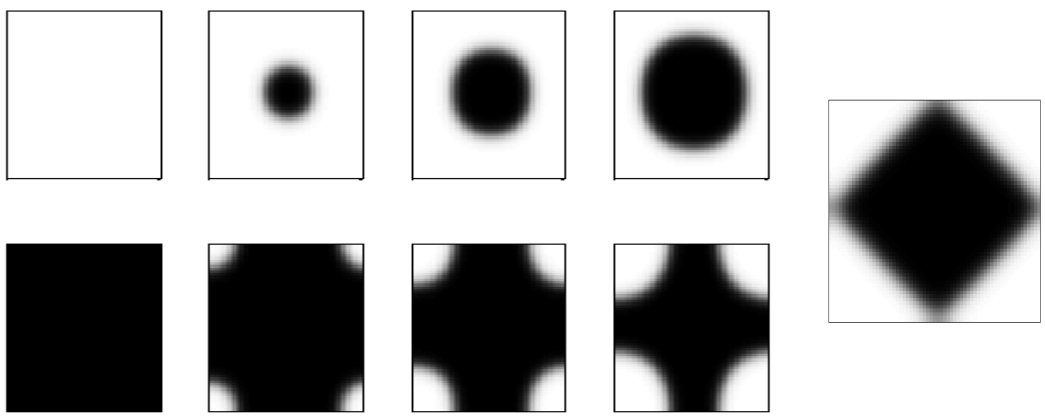

Figure 4.12: The u-field of a local minimizer for $\beta = 10$ and $\kappa = 0.001$. There is almost no difference to the one of local minimizer in Figure 4.11 for $\beta = 1$.

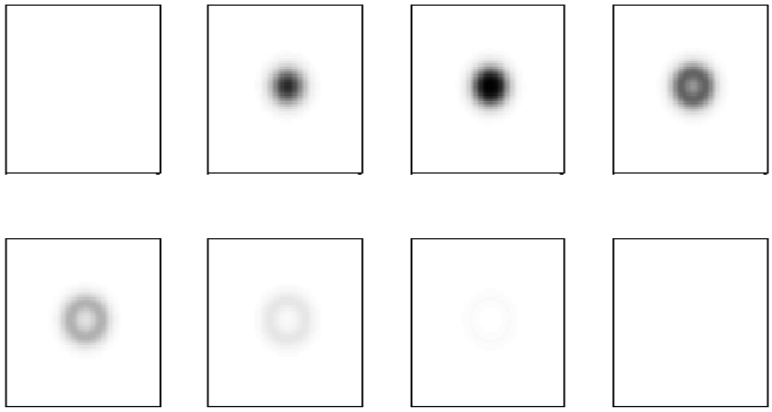

Figure 4.13: The v-field of the local minimizer of Figure 4.12 for $\beta = 10$. We observe that for this value of β the v-field assists at the beginning of the transition. The snapshots in this Figure correspond only to the time between the first two squares for the u-field; during the rest of the transition the v-field is inactive.

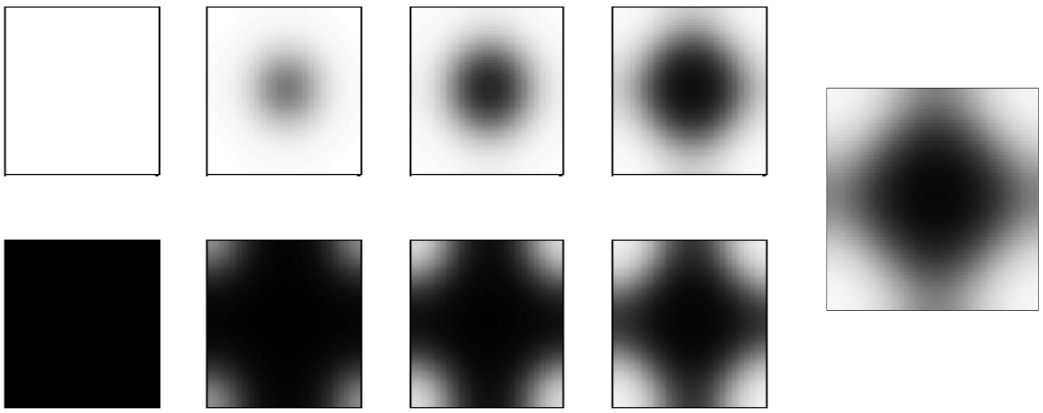

Figure 4.14: A local minimizer for $\beta = 1$ and $\kappa = 0.01$ (u-field only).

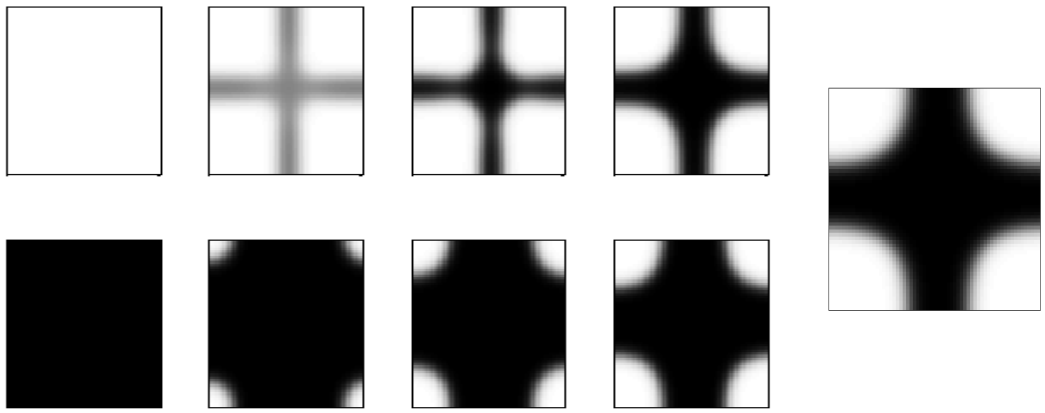

Figure 4.15: Another local minimizer for $\beta = 10$ and $\kappa = 0.001$ (u-field only).

4.3 Continuous-time Markov chain: The genetic switch

As a last example, we apply our technique to a birth-death process with a positive feedback loop that results in two stable states. The model was first defined by Roma et al [28] and describes a mechanism in molecular biology called the genetic switch, illustrated in Figure 4.16. (To illustrate the numerics, in this section we will only treat the simplifying model of [28] as is; a more precise treatment and a more detailed introduction to the topic will be done later in Chapter 6.)

A bacterial cell contains plasmids (high-copy DNA strings) with two gene sites a and b that can be transcribed and translated into proteins A and B. Those in turn can form polymers that bind to the operator site of the respective other gene, preventing further production of the corresponding protein. Other reactions are

to reverse polymer formation or protein binding, and the degradation of proteins.

This setup leads to bistable behavior: If the cell is in a state with many proteins A and only few proteins B then there are likely also many polymers A_l, the operator site of gene b will be blocked most of the time, and thus only few new proteins B are being produced. Since there are only few proteins B, it is unlikely that a polymer B_m will bind to the operator site of gene a, and the production rate of new proteins A will stay high. Therefore there will be a stable state with many proteins A and few proteins B, and by symmetry of the mechanism there is another stable state with few proteins A and many proteins B. Bistability arises because the fluctuations leading to a switch from one stable state to the other are rare events.

A simplifying description of this process keeps track only of the numbers (X_a, X_b) of the proteins of the two types and models polymer formation and binding to the DNA only by defining the production rate as a function of the number of proteins of the respective other type, the precise form of which is motivated by the Hill equation [34]. It thus consists of only four reactions, as listed in Table 4.1.

Here, Ω is the system size parameter (such as the total number of proteins in the cell), $(x_a, x_b) := (X_a/\Omega, X_b/\Omega)$ is the protein density, a_1 and a_2 are rate parameters that combine the rates for transcription into RNA and their translation into proteins, and l and m are the cooperativity parameters which represent the numbers of proteins per polymer. In the simulations below, we use the same model parameters as Roma et al. [28], namely $a_1 = 156$, $a_2 = 30$, $\mu_1 = \mu_2 = 1$, $l = 1$, $m = 3$.

Under this kind of scaling the stochastic system for $x := (x_a, x_b)$ satisfies a large deviations principle as $\varepsilon = \frac{1}{\Omega} \to 0$, with Hamiltonian

$$\begin{aligned} H(x,\theta) = &\frac{a_1}{1+x_b^m}(e^{\theta_a}-1) + \mu_1 x_a(e^{-\theta_a}-1) \\ &+ \frac{a_2}{1+x_a^l}(e^{\theta_b}-1) + \mu_2 x_b(e^{-\theta_b}-1) \end{aligned} \tag{4.18}$$

(see [29]) which we used in the gMAM algorithm presented in Chapter 3.

Figure 4.17 shows the transition path obtained with gMAM, for the full path and only for the uphill path. They match those found in [28] using a shooting method based on the Hamilton equations associated with the Hamiltonian (4.18). While in general for continuous-time Markov chains no explicit expression for $\hat{\vartheta}(x,y)$ exists, for the simple Hamiltonian (4.18) we could indeed find one. Comparing the curve obtained from gMAM using the explicit formula for $\hat{\vartheta}(x,y)$ with the one using the algorithm from Section 3.4 in an inner loop, we found that both curves matched exactly. When we applied the technique to obtain second-order accuracy as described in Section 3.3, the corner of the path at the saddle point was sharp, and λ at that point vanished up to machine precision.

Table 4.2 shows the results of our performance tests on this model. For various gridsizes N it lists the optimal stepsize $\Delta\tau$, the number of steps necessary until the change in the action $\hat{S}(\varphi)$ per iteration drops below 10^{-7} and the corresponding runtime. We observe again that the maximum stepsize $\Delta\tau$ is roughly independent of the gridsize N, and that the runtime is close to linear in N.

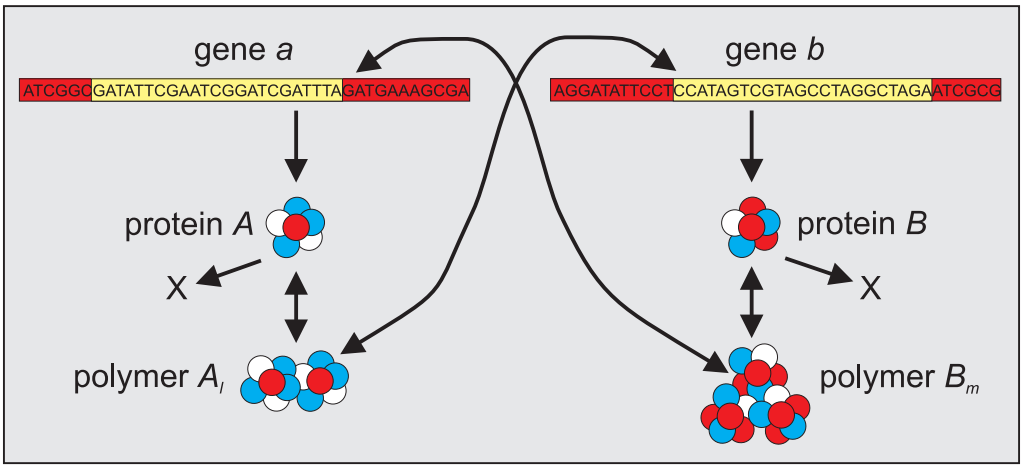

Figure 4.16: The mechanism of the genetic switch. The X symbolizes protein degradation.

reaction type	rate	state change of (X_a, X_b)
protein production	$\Omega a_1(1+x_b^m)^{-1}$ $\Omega a_2(1+x_a^l)^{-1}$	$(1,0)$ $(0,1)$
protein degradation	$\Omega\mu_1 x_a$ $\Omega\mu_2 x_b$	$(-1,0)$ $(0,-1)$

Table 4.1: The reactions of the Roma model.

gridsize N	stepsize $\Delta\tau$	# iterations	runtime
100	0.22	20	0.5 sec
300	0.27	15	0.6 sec
1,000	0.26	18	1.3 sec
3,000	0.26	18	3.1 sec
10,000	0.27	19	11.6 sec

Table 4.2: Algorithm performance for the Roma model.

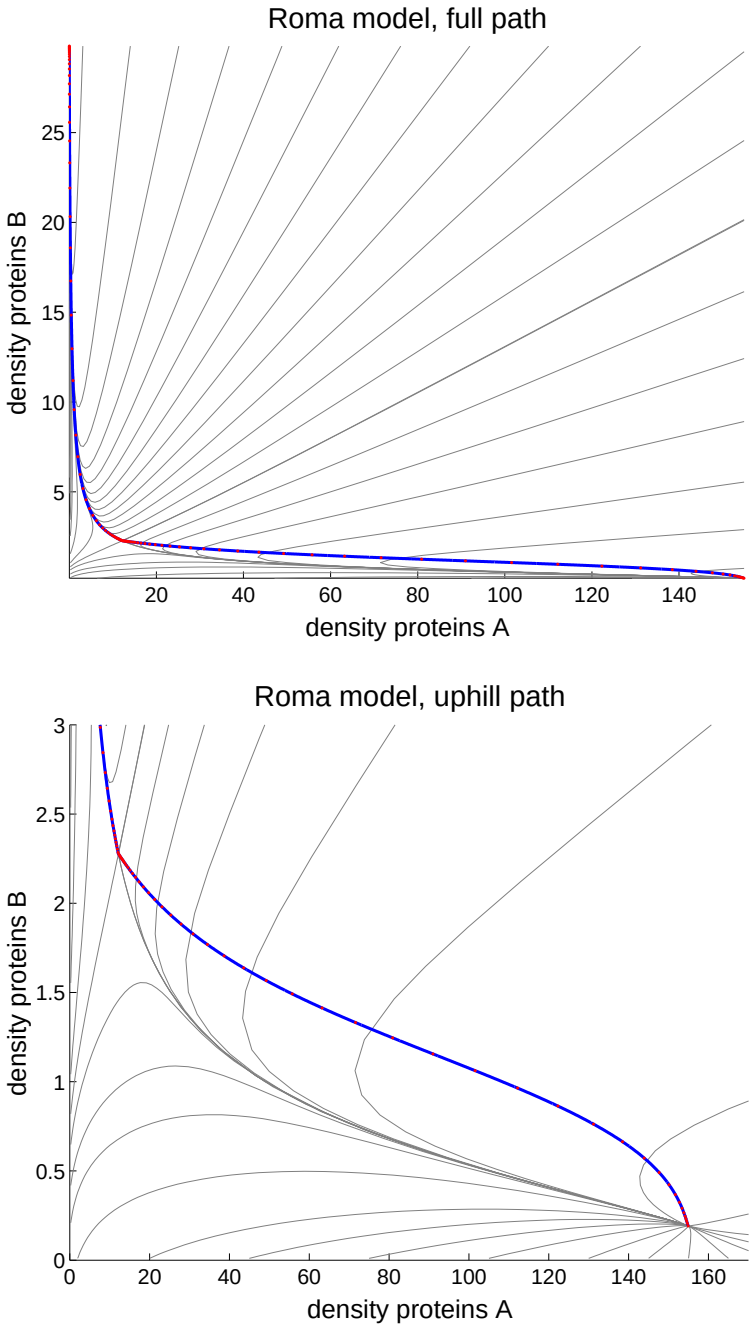

Figure 4.17: The minimum action path for the Roma model. x- and y-axis denote the densities of the proteins of type A and B, respectively. The first panel shows the full path from the right stable state to the left stable state; the second panel shows the piece of the path from the right stable state to the saddle point.

Chapter 5

Minimization with variable endpoints

5.1 Endpoint constraints

Another use of the geometric action $\hat{S}$ is the computation of the invariant measure: Let the family of diffusion processes $\{(X_t^\varepsilon)_{t\geq 0},\, \varepsilon > 0\}$ on $D \subset \mathbb{R}^n$ fulfill a large deviations principle with action $S_T(\psi)$, denote by $x_1, \ldots, x_K$ its stable states (we assume that there are no other stable structures such as limit cyles), let μ_ε denote its invariant measure (with $\mu_\varepsilon(D) = 1$), and let $B \subset D$ be a Borel set. Then the limiting behavior of $\mu_\varepsilon(B)$ is given by

$$\lim_{\varepsilon\to 0} -\varepsilon \log \mu_\varepsilon(B) = \inf_{k=1,\ldots,K} \inf_{T>0} \inf_{\psi\in\bar{C}^B_{x_k}(0,T)} S_T(\psi), \tag{5.1}$$

where we denote by $\bar{C}^B_{x_k}(0,T)$ the space of all functions $\psi \in \bar{C}(0,T)$ such that $\psi(0) = x_k$ and $\psi(T) \in B$. Note that $\mu_\varepsilon(B)$ is the probability that after reaching

equilibrium the process (X_t^ε) is in the set B.

While the minimization over the finitely many points $x_1, \ldots, x_K$ is not a problem, we see that again we have a double minimization problem over both $T > 0$ and a set of paths $\psi \in \bar{C}(0,T)$. In contrast to the minimization problem (3.1), in this case the end point x_2 of the path ψ is allowed to vary within the set B. But using our technique from Chapter 2, we can easily reformulate the problem in terms of $\hat{S}$:

Lemma 7. *The minimization problem in* (5.1) *can be reformulated as*

$$\lim_{\varepsilon \to 0} -\varepsilon \log \mu_\varepsilon(B) = \inf_{k=1,\ldots,K} \inf_{\varphi \in \bar{C}^B_{x_k}(0,1)} \hat{S}(\varphi). \tag{5.2}$$

Proof.

$$\begin{aligned}
\inf_{T>0} \inf_{\psi \in \bar{C}^B_{x_1}(0,T)} S_T(\psi) &= \inf_{T>0} \inf_{\varphi \in \bar{C}^B_{x_1}(0,1)} \inf_{\psi \in \bar{C}_\varphi(0,T)} S_T(\varphi) \\
&= \inf_{\varphi \in \bar{C}^B_{x_1}(0,1)} \inf_{T>0} \inf_{\psi \in \bar{C}_\varphi(0,T)} S_T(\varphi) \\
&= \inf_{\varphi \in \bar{C}^B_{x_1}(0,1)} \hat{S}(\varphi),
\end{aligned}$$

where in the last step we used the representation (2.3) of $\hat{S}$. □

This frees us from the minimization over T, and we can again make use of the invariance of $\hat{S}(\varphi)$ under reparametrization of φ by reparametrizing the path after each iteration of our steepest descent scheme. Denoting the local action of $\hat{S}$ by $\ell(x,y)$, the variation of $\hat{S}$ applied to any test function $\eta \in C^\infty(0,1)$ with

$\eta(0) = 0$ (but not necessarily $\eta(1) = 0$) can be computed as follows:

$$\begin{aligned}
\big(\delta\hat{S}(\varphi)\big)(\eta) &= \text{first-order approximation in } \eta \text{ of } \hat{S}(\varphi+\eta) - \hat{S}(\varphi) \\
&= \int_0^1 \big(\langle \ell_x, \eta\rangle + \langle \ell_y, \eta'\rangle\big)\, d\alpha \\
&= \int_0^1 \langle \ell_x - \partial_\alpha \ell_y, \eta\rangle\, d\alpha + \big\langle \ell_y(\varphi(1), \varphi'(1)), \eta(1)\big\rangle \\
&= \langle D\hat{S}(\varphi), \eta\rangle_{L^2([0,1],\,\mathbb{R}^n)} + \big\langle \hat{\vartheta}(\varphi(1), \varphi'(1)), \eta(1)\big\rangle,
\end{aligned}$$

where ℓ_x and ℓ_y are evaluated at $(\varphi(\alpha), \varphi'(\alpha))$ unless specified otherwise, and where $D\hat{S}(\varphi)$ is the variation of $\hat{S}(\varphi)$ taken in the L^2-norm, as computed in Section 3.1, with both endpoints fixed. In the last step we used that

$$\ell_y(x,y) = \partial_y \langle y, \hat{\vartheta}(x,y)\rangle = \hat{\vartheta}(x,y) + \hat{\vartheta}_y(x,y)^T y = \hat{\vartheta}(x,y)$$

by representation (2.5) and (E.3) in Appendix E. Assume now that the boundary of B can be written as $\partial B = \{x \in D \,|\, g(x) = 0\}$ for some differentiable function $g : D \to \mathbb{R}$. When looking for the minimizer in (5.2), we can clearly restrict ourselves to curves φ with endpoint $\varphi(1) \in \partial B$ since any piece of the curve beyond the first hitting point of B can only increase the action and may thus be deleted at no cost (unless some x_k is in B, in which case (5.2) is equal to zero). Therefore, to adjust the gMAM to the present situation, we can introduce a Lagrange multiplier term ν to enforce the endpoint constraint $g(\varphi(\tau, 1)) = 0$

throughout the steepest descent flow:

$$\begin{cases} \dot{\varphi}(\tau,\alpha) = -\lambda H_{\theta\theta} D\hat{S}(\varphi), \\ \\ \dot{\varphi}(\tau,1) = -\hat{\vartheta}\big(\varphi(\tau,1),\varphi'(\tau,1)\big) + \nu\nabla g(\varphi(\tau,1)), \\ \\ \varphi(\tau,0) = x_1, \quad \varphi(0,\alpha) = \varphi^0(\alpha) \end{cases} \tag{5.3}$$

for $\alpha \in [0,1]$ and $\tau \geq 0$. Here, $\nu(\tau) \in \mathbb{R}$ is chosen such that $g(\varphi(\tau,1)) \equiv 0$, i.e.

$$\begin{aligned} 0 &\equiv \partial_\tau g(\varphi) = \big\langle \nabla g(\varphi), \dot{\varphi} \big\rangle \\ &= \big\langle \nabla g(\varphi), -\hat{\vartheta}(\varphi,\varphi') + \nu\nabla g(\varphi) \big\rangle \\ \Rightarrow \qquad \nu &= \frac{\big\langle \nabla g(\varphi), \hat{\vartheta}(\varphi,\varphi') \big\rangle}{|\nabla g(\varphi)|^2}, \end{aligned}$$

with all occurrences of φ, φ' and $\dot{\varphi}$ evaluated at $(\tau,1)$. This shows the following proposition:

Proposition 5. *The minimizer $\varphi^\star$ of the minimization problem in Lemma 7 is the steady state solution of the steepest descent flow*

$$\begin{cases} \dot{\varphi}(\tau,\alpha) = P_{\varphi'}\big(\lambda^2\varphi'' - \lambda H_{\theta x}\varphi' + H_{\theta\theta}H_x\big) + \mu\varphi' \\ \qquad\qquad = \lambda^2\varphi'' - \lambda H_{\theta x}\varphi' + H_{\theta\theta}H_x + \lambda\lambda'\varphi' + \mu\varphi' \quad \textit{for} \quad \alpha \in (0,1), \\ \\ \dot{\varphi}(\tau,1) = -P_{\nabla g(\tau,1)}\,\hat{\vartheta}\big(\varphi(\tau,1),\varphi'(\tau,1)\big), \\ \\ \varphi(\tau,0) = x_1, \quad \varphi(0,\alpha) = \varphi^0(\alpha), \end{cases} \tag{5.4}$$

for $\alpha \in [0,1]$ and $\tau \geq 0$. Here $\mu\varphi'$ is the Lagrange multiplier term introduced to enforce that $|\varphi'(\tau,\cdot)| \equiv cst(\tau)$ along the curve, we abbreviate $P_c := I - \frac{c\otimes c}{|c|^2}$ for every $c \in \mathbb{R}^n$, and the initial curve φ^0 must be chosen such that it fulfills the constraints $|\varphi^{0\prime}| \equiv cst$ and $g(\varphi^0(1)) = 0$.

Note that for $\alpha \in (0,1)$ we are again working with $-\lambda H_{\theta\theta} D\hat{S}(\varphi)$ to improve the stability of the code. Each iteration of the gMAM will now consist of three steps: (i) follow the flow (5.4) (without the term $\mu\varphi'$) for one discrete time step, (ii) redistribute the points along the curve to make them equidistant, and (iii) update the endpoint once more using Newton-Raphson's method to make $g(\varphi(1)) = 0$, i.e. replace φ_N^k by

$$\varphi_N^k - \frac{g(\varphi_N^k)}{|\nabla g(\varphi_N^k)|^2} \nabla g(\varphi_N^k),$$

to get rid of second-order errors in $g(\varphi(1))$.

5.2 Endpoint penalties

One more situation in which the use of $\hat{S}$ is helpful is the following: Let μ_ε be the invariant measure defined in the previous section, and let $f : D \to \mathbb{R}$ be a bounded continuous function. Then we have the limit

$$\lim_{\varepsilon\to 0} -\varepsilon \log \int_D e^{-\frac{1}{\varepsilon} f(x)}\, d\mu_\varepsilon(x) = \inf_{k=1,\dots,K} \inf_{T>0} \inf_{\psi\in \bar{C}_{x_k}(0,T)} \big(S_T(\psi) + f(\psi(T))\big), \quad (5.5)$$

where the integral is the expectation with respect to the invariant measure (this can be considered as a generalization of Varadhan's Lemma). In this case there is no constraint on the endpoint $\psi(T)$, but because of the additive penalty term $f(\psi(T))$ a tradeoff between the length and location of the curve and the position of its endpoint may have to be made. Again, we can use $\hat{S}$ to carry out the minimization over $T > 0$ analytically:

Lemma 8. *The minimization problem in* (5.5) *can be reformulated as*

$$\lim_{\varepsilon \to 0} -\varepsilon \log \int_D e^{-\frac{1}{\varepsilon} f(x)} \, d\mu_\varepsilon(x) = \inf_{k=1,\ldots,K} \inf_{\varphi \in \bar{C}_{x_k}(0,1)} \hat{S}_f(\varphi), \tag{5.6}$$

where $\hat{S}_f(\varphi) := \hat{S}(\varphi) + f(\varphi(1))$.

Proof.

$$\begin{aligned}
&\inf_{T>0} \inf_{\psi \in \bar{C}_{x_k}(0,T)} \big(S_T(\psi) + f(\psi(T))\big) \\
&\quad = \inf_{T>0} \inf_{\varphi \in \bar{C}_{x_k}(0,1)} \inf_{\psi \in \bar{C}_\varphi(0,T)} \big(S_T(\psi) + f(\psi(T))\big) \\
&\quad = \inf_{\varphi \in \bar{C}_{x_k}(0,1)} \inf_{T>0} \inf_{\psi \in \bar{C}_\varphi(0,T)} \big(S_T(\psi) + f(\varphi(1))\big) \\
&\quad = \inf_{\varphi \in \bar{C}_{x_k}(0,1)} \Big(\Big(\inf_{T>0} \inf_{\psi \in \bar{C}_\varphi(0,T)} S_T(\psi)\Big) + f(\varphi(1))\Big) \\
&\quad = \inf_{\varphi \in \bar{C}_{x_k}(0,1)} \big(\hat{S}(\varphi) + f(\varphi(1))\big) \\
&\quad = \inf_{\varphi \in \bar{C}_{x_k}(0,1)} \hat{S}_f(\varphi).
\end{aligned} \tag{5.7}$$

□

The variation of $\hat{S}_f$ applied to any test function $\eta \in C^\infty(0,1)$ with $\eta(0) = 0$ is

$$\big(\delta \hat{S}_f(\varphi)\big)(\eta) = \langle D\hat{S}(\varphi), \eta\rangle_{L^2([0,1],\,\mathbb{R}^n)} + \big\langle \hat{\vartheta}(\varphi(1), \varphi'(1)) + \nabla f(\varphi(1)), \eta(1)\big\rangle, \quad (5.8)$$

and we arrive at the following proposition:

Proposition 6. *The minimizer $\varphi^\star$ of the minimization problem in Lemma 8 is the steady state solution of the steepest descent flow*

$$\begin{cases} \dot{\varphi}(\tau,\alpha) = P_{\varphi'}\big(\lambda^2\varphi'' - \lambda H_{\theta x}\varphi' + H_{\theta\theta}H_x\big) + \mu\varphi' \\ \qquad\qquad = \lambda^2\varphi'' - \lambda H_{\theta x}\varphi' + H_{\theta\theta}H_x + \lambda\lambda'\varphi' + \mu\varphi' \quad \textit{for} \quad \alpha \in (0,1), \\ \\ \dot{\varphi}(\tau,1) = -\hat{\vartheta}(\varphi(\tau,1), \varphi'(\tau,1)) - \nabla f(\varphi(\tau,1)), \\ \\ \varphi(\tau,0) = x_1, \quad \varphi(0,\alpha) = \varphi^0(\alpha), \end{cases} \quad (5.9)$$

for $\alpha \in [0,1]$ and $\tau \geq 0$. Here $\mu\varphi'$ is again the Lagrange multiplier term introduced to enforce the constraint that $|\varphi'(\tau,\cdot)| \equiv cst(\tau)$ along the curve, and the initial curve φ^0 must be chosen such that it fulfills the constraint $|\varphi^{0\prime}| \equiv cst$.

Proposition 6 allows for a straight-forward modification of the gMAM algorithm introduced in Chapter 3 by modifying the linear system (3.6) so that

$$\tilde{\varphi}_N = \varphi_N - \Delta\tau\big(\hat{\vartheta}_N^k + \nabla f(\varphi_N^k)\big).$$

We remind the reader that in the case of a diffusion process an analytic formula for $\hat{\vartheta}(x,y)$ is given by (2.12).

As a final remark, we want to point out that it is easily possible to generalize the use of the gMAM to combinations of the minimization problems (5.1) and (5.5), i.e. to situations in which one has both an endpoint constraint and a penalty term.

5.3 Example: SDE

Consider the diffusion process

$$dX = -M\nabla V(X)\,dt + \sqrt{\varepsilon}\,dW, \tag{5.10}$$

where the potential $V(x) \geq 0$ achieves its minimal value 0 at the only critical point $x_1 = 0$, say, and where M is of the form $M = I + J$, with $J^T = -J$.

The specific form of the SDE allows us to solve the minimization problems (5.1) and (5.5) analytically (see also [12, Theorem 3.1]), which enables us to benchmark the results obtained by the gMAM. We start by rewriting the corresponding action:

$$\begin{aligned}
S_T(\psi) &= \frac{1}{2}\int_0^T |\dot{\psi} + M\nabla V(\psi)|^2\,dt \\
&= \frac{1}{2}\int_0^T \left|(\dot{\psi} - M^T\nabla V(\psi)) + (\underbrace{M + M^T}_{=2I})\nabla V(\psi)\right|^2 dt \\
&= \frac{1}{2}\int_0^T \Big(\left|\dot{\psi} - M^T\nabla V(\psi)\right|^2 + 4\left|\nabla V(\psi)\right|^2 \\
&\qquad\qquad + 4\big\langle \dot{\psi} - M^T\nabla V(\psi), \nabla V(\psi)\big\rangle\Big)\,dt
\end{aligned}$$

$$= \frac{1}{2}\int_0^T \Big(\big|\dot{\psi} - M^T\nabla V(\psi)\big|^2 + 4\partial_t V(\psi) \\ + 4\Big(\underbrace{\big|\nabla V(\psi)\big|^2 - \big\langle M^T\nabla V(\psi), \nabla V(\psi)\big\rangle}_{=0}\Big)\Big)\, dt.$$

Since $\langle J^T x, x\rangle = \langle x, Jx\rangle = \langle x, -J^T x\rangle$ and thus $\langle J^T x, x\rangle = 0$ for every x, the third term in the last line vanishes. Therefore, if $\psi(0) = x_1$ then

$$S_T(\psi) = \frac{1}{2}\int_0^T |\dot{\psi} - M^T\nabla V(\psi)|^2\, dt + 2V(\psi(T)), \tag{5.11}$$

since $V(\psi(0)) = V(x_1) = 0$.

Analytic solution: Endpoint constraints. To solve the minimization problem (5.1), we minimize $S_T(\psi)$ as given in (5.11) as follows: First we determine the optimal endpoint $\psi(T)$, say $x_2^\star$, that minimizes $V(\psi(T))$ on the set ∂B by using a Lagrange multiplier. Then we make the integral term vanish by setting $\psi(t) := \chi(T-t)$, with $\chi(t)$ starting at $x_2^\star$ and following the ODE $\dot{\chi} = -M^T\nabla V(\chi)$. Since $\lim_{t\to\infty}\chi(t) = x_1$, ψ will connect x_1 to $x_2^\star$ as $T \to \infty$. This shows that

$$\inf_{T>0}\inf_{\psi\in\bar{C}^B_{x_1}(0,T)} S_T(\psi) = 2V(x_2^\star). \tag{5.12}$$

As an example in $\mathbb{R}^2$, we consider $V(x,y) = \frac{1}{2}x^2 + y^2$ and

$$B := \{(x,y) \in \mathbb{R}^2_+ \,|\, xy \geq 2\}.$$

To find $x_2^\star =: (x, y)$, we define

$$F(x, y, \lambda) := 2V(x, y) - \lambda(xy - 2) = x^2 + 2y^2 + \lambda(xy - 2)$$

and set $\nabla_{x,y,\lambda} F(x, y, \lambda) = 0$. We obtain

$$\begin{aligned} & 2x + \lambda y = 0 \quad \wedge \quad 4y + \lambda x = 0 \quad \wedge \quad xy - 2 = 0 \\ \Rightarrow \quad & 2x + \Big(-\frac{4y}{x}\Big) y = 0 \quad \wedge \quad y = \frac{2}{x} \\ \Rightarrow \quad & 2x - \frac{16}{x^3} = 0 \quad \wedge \quad y = \frac{2}{x} \end{aligned}$$

and thus $x_2^\star = (2^{3/4}, 2^{1/4})$. This shows that the minimum value (5.12) and thus the limit in (5.1) is $2V(x_2^\star) = (2^{3/4})^2 + 2(2^{1/4})^2 = 4\sqrt{2}$, independently of the choice of M.

Analytic solution: Endpoint penalties. To minimize

$$S_T(\psi) + f(\psi(T)) = \frac{1}{2} \int_0^T |\dot{\psi} - M^T \nabla V(\psi)|^2 \, dt + \big[2V(\psi(T)) + f(\psi(T))\big], \quad (5.13)$$

and thus solve the minimization problem (5.5) we proceed similarly: We first determine the optimal endpoint $x_2^\star$ by setting the derivative of $2V + f$ equal to zero, then we make the integral part vanish by having ψ follow the flow $-M^T \nabla V(x)$ backwards in time. This shows that

$$\inf_{T>0} \inf_{\psi \in \bar{C}_{x_1}(0,T)} \big(S_T(\psi) + f(\psi(T))\big) = 2V(x_2^\star) + f(x_2^\star). \quad (5.14)$$

As an example, we take $V(x)$ as before and $f(x,y) = \frac{1}{2}(x-1)^2 + \frac{1}{2}(y-1)^2$. Then $x_2^\star = (x,y)$ fulfills

$$\begin{aligned} 0 &= 2\nabla V(x,y) + \nabla f(x,y) = \begin{pmatrix} 2x + (x-1) \\ 4y + (y-1) \end{pmatrix} \\ \Rightarrow \qquad x_2^\star &= (\tfrac{1}{3}, \tfrac{1}{5}), \end{aligned}$$

and the right-hand side of (5.14) and thus the limit in (5.5) equals

$$2V(x_2^\star) + f(x_2^\star) = \left((\tfrac{1}{3})^2 + 2(\tfrac{1}{5})^2\right) + \left(\tfrac{1}{2}(\tfrac{1}{3}-1)^2 + \tfrac{1}{2}(\tfrac{1}{5}-1)^2\right) = \tfrac{11}{15}.$$

Numerical results using gMAM. To compare the output of the gMAM with our analytical solutions, we chose $M = \left(\begin{smallmatrix} 1 & -1 \\ 1 & 1 \end{smallmatrix}\right)$. Figure 5.1 shows the minimizing path for the minimization problem (5.2) (blue curve with red dots that are equidistant in time), obtained after 60 iterations of the gMAM using the gridsize $N = 1,000$ and the stepsize $\Delta\tau = 0.2$, on top of the analytic solution (black). The green line in Figure 5.1 is the boundary of the set B. As we can see, the curves match as much as one can expect from the chosen grid size (the black curve is hardly visible on this plot).

Figure 5.2 shows the accuracy of the curve as a function of N (for values of N between 50 and 50,000), on doubly-logarithmic scale. The upper curve is the error of the endpoint: it is a straight line with slope -1.01, indicating an error of the order $O(1/N)$. The lower curve is the error of the action; this slightly noisier curve fits a straight line with slope -2.00 and thus indicates an error in the action of the order $O(1/N^2)$. The runtime needed for the 60 iterations (no figure)

increased linearly from 0.34 seconds for $N = 50$ to 138 seconds for $N = 50,000$ (using MatLab 6.5 running under Windows XP on a 1.5 GHz Pentium 4).

Figures 5.3 and 5.4 show the corresponding graphs for the minimization problem (5.6). This time we used 40 iterations at $\Delta\tau = 0.3$, the choices for N stayed the same. As one can see, the minimizing curve found by gMAM matches our analytical solution again, and the graphs displaying the accuracy very much resemble the ones displayed in Figure 5.2 (here with slopes -1.02 and -2.02), indicating again first-order accuracy in N for the endpoint and second-order accuracy for the action (including the penalty $f(\varphi(1))$). The runtime increased linearly from 0.25 to 48 seconds. Notice that here the time for one iteration is shorter than in the previous example because if one imposes a boundary constraint (which depends on both components of the endpoint) then the linear system (3.6) cannot be solved separately for each component of the path anymore.

5.4 Finance application: Valuation of equity index options

The basic strategy of the gMAM – minimizing a parametrization-free functional by a two-step procedure consisting of one steepest-descent step and one reparametrization step – can also be useful to address problems other than the double minimization over T and ψ, as long as the problem at hand can be rephrased in terms of a parametrization-free functional $\tilde{S}$. One situation in which this is

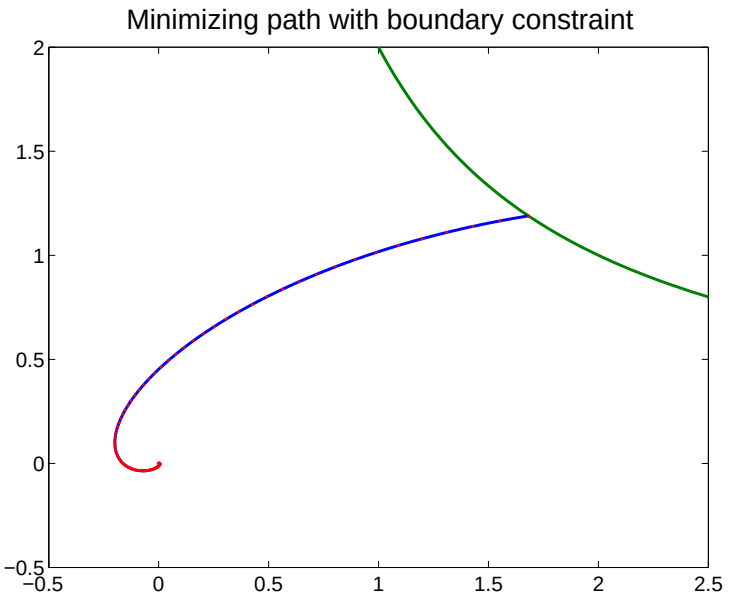

Figure 5.1: Minimization with boundary constraint: the minimizing path.

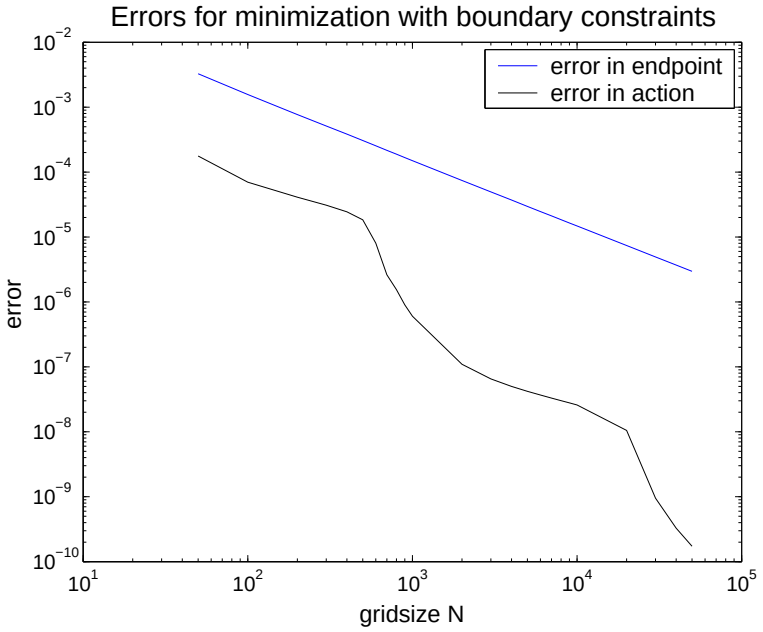

Figure 5.2: Minimization with boundary constraint: accuracy measurements.

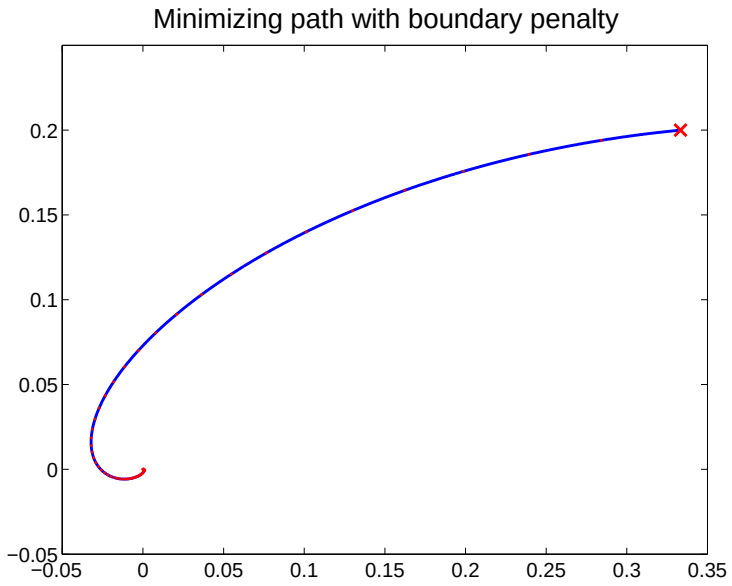

Figure 5.3: Minimization with penalty: the minimizing path.

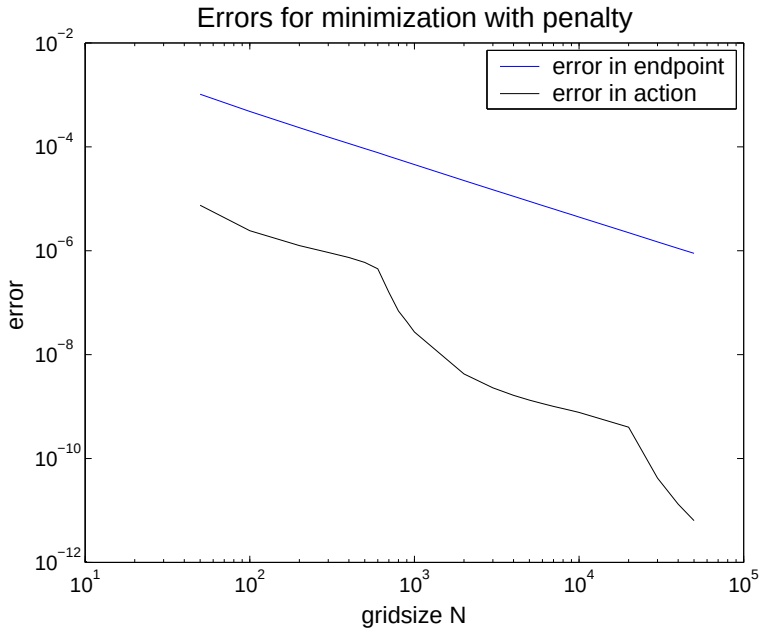

Figure 5.4: Minimization with penalty: accuracy measurements.

possible is the computation of the short-time aymptotics of SDEs, or equivalently, fixed-T action minimization without the presence of a drift term.

As an illustrative example we choose a typical question arising in mathematical finance: the valuation of a European option.

Finance introduction: European options. A European option is a contract that gives its holder the right (but not the obligation) to buy (call option) or to sell (put option) an equity from the seller of the option at a fixed future date (the expiration date) at a set price (the strike price). In the case of a call option, if at the expiration time the value of the underlying equity is above the strike price, then the holder will exercise his option to buy the equity at the strike price, and he has then the possibility of selling it again on the market at the higher market price. If at the expiration time the value of the equity is below the strike price, then he will instead choose not to exercise his right; in that case his loss is the money that he had to pay to enter the contract. Put options on the other hand are exercised if the price of the equity falls below the strike price; they are one way to take advantage of a decline of the equity's value.

The value of the option, i.e. the price that one has to pay to enter the contract, has to reflect the probability that at the expiration time the value of the equity is above (or below) the strike price. The seller of the contract will therefore typically use a mathematical model describing the market to estimate this probability and thus to compute the value of the option. If the expiration time and the variance of the market price are small compared to the difference between the strike price and the spot (current) price then the event that the price of the equity will indeed

bridge this gap is a rare event, and large deviations theory can provide us with analytic formulas for the asymptotics.

The gMAM for short-time asymptotics. Let us assume that the n-dimensional process $(X_t)_{t\geq 0}$ representing the values of the equities fulfills the SDE

$$dX_t = b(X_t)\,dt + \sigma(X_t)\,dW_t, \tag{5.15}$$

where W_t is an n-dimensional Brownian motion. We are interested in the behavior of $X_{\varepsilon T}$ for small $\varepsilon > 0$ and some fixed $T > 0$. The change of variables $\bar{X}_t := X_{\varepsilon t}$ shows that the process $(\bar{X}_t)_{t\geq 0}$ fulfills the SDE

$$d\bar{X}_t = \varepsilon b(\bar{X}_t)\,dt + \sqrt{\varepsilon}\sigma(\bar{X}_t)\,dW_t,$$

and thus the asymptotic behavior of $X_{\varepsilon T} = \bar{X}_T$ can be determined by large deviations theory: for any regular Borel set B (which will denote the set on which the option is exercised) we have

$$\begin{aligned} \lim_{\varepsilon\to 0+} \varepsilon \log \mathbb{P}(X_{\varepsilon T} \in B \,|\, X_0 = x_1) &= \lim_{\varepsilon\to 0+} \varepsilon \log \mathbb{P}(\bar{X}_T \in B \,|\, \bar{X}_0 = x_1) \\ &= -\inf_{\psi\in \bar{C}^B_{x_1}(0,T)} S_T(\psi), \quad \text{where} \\ S_T(\psi) &= \frac{1}{2}\int_0^T |\dot{\psi}|^2_{A(\psi)}\,dt \end{aligned} \tag{5.16}$$

and $A(x) = \sigma(x)\sigma(x)^T$. Notice that in contrast to the typical situation in which we would apply the gMAM, i.e. the minimization over both T and ψ, in (5.16)

T is kept fixed. But the fact that in the present case the drift b does not enter the action $S_T(\psi)$ allows us to express (5.16) in terms of a parametrization-free functional $\tilde{S}$ anyway, using a different trick: applying Hölder's inequality, we obtain the estimate

$$\big(T \cdot 2S_T(\psi)\big)^{1/2} = \Big(\int_0^T 1^2\,dt\Big)^{1/2}\Big(\int_0^T |\dot{\psi}|^2_{A(\psi)}\,dt\Big)^{1/2} \tag{5.17}$$

$$\geq \int_0^T 1 \cdot |\dot{\psi}|_{A(\psi)}\,dt \tag{5.18}$$

$$\Rightarrow \qquad S_T(\psi) \geq \frac{1}{2T}\Big(\int_0^T |\dot{\psi}|_{A(\psi)}\,dt\Big)^2 = \frac{1}{2T}\Big(\int_{\gamma(\psi)} |dz|_{A(z)}\Big)^2. \tag{5.19}$$

Equality in (5.18) and thus in (5.19) is achieved if the two integrands in (5.17) are positive multiples of each other, i.e. if ψ is parametrized such that $|\dot{\psi}|_{A(\psi)} \equiv cst$ almost everywhere, which is certainly possible for any T and any path $\gamma(\psi)$. Therefore we can rewrite (5.16) as

$$\begin{aligned}\inf_{\psi\in\bar{C}^B_{x_1}(0,T)} S_T(\psi) &= \inf_{\varphi\in\bar{C}^B_{x_1}(0,1)}\ \inf_{\psi\in\bar{C}_\varphi(0,T)} S_T(\psi) = \inf_{\varphi\in\bar{C}^B_{x_1}(0,1)} \frac{1}{2T}\Big(\int_{\gamma(\varphi)} |dz|_{A(z)}\Big)^2 \\ &= \frac{1}{2T}\Big(\inf_{\varphi\in\bar{C}^B_{x_1}(0,1)} \int_{\gamma(\varphi)} |dz|_{A(z)}\Big)^2,\end{aligned}$$

and we conclude that

$$\lim_{\varepsilon\to 0+} \varepsilon \log \mathbb{P}(X_{\varepsilon T} \in B \,|\, X_0 = x_1) = -\frac{1}{2T}\Big(\inf_{\varphi\in\bar{C}^B_{x_1}(0,1)} \tilde{S}(\varphi)\Big)^2, \tag{5.20}$$

$$\text{where} \qquad \tilde{S}(\varphi) = \int_0^1 |\varphi'|_{A(\varphi)}\,d\alpha. \tag{5.21}$$

Note that $\tilde{S}(\varphi)$ was not computed from $S(\psi)$ in the same way as $\hat{S}(\varphi)$ (since the underlying problem here is different), therefore the corresponding Euler-Lagrange equation is different from (3.2). One can however compute the variation of $\tilde{S}(\varphi)$ as given by (5.21) by hand and find that the (preconditioned) steepest-descent flow for the minimization problem (5.20) is

$$\begin{aligned} \dot{\varphi} &= -A^{-1}(\varphi)|\varphi'|_{A(\varphi)} D\tilde{S}(\varphi) \\ &= \varphi'' + A^{-1}\big[(\nabla A \star \varphi') - \tfrac{1}{2}(\nabla A \star \varphi')^T\big]\varphi' - \tfrac{\lambda'}{\lambda}\varphi', \end{aligned} \tag{5.22}$$

where $\lambda = |\varphi'|_{A(\varphi)}$, and where we introduced the notation $\nabla A(x) \star c$ for the matrix

$$(\nabla A(x) \star c)_{i,j} := \big[\big(\partial_{x_j} A(x)\big)c\big]_i = \sum_{k=1}^{n} \partial_{x_j} A(x)_{ik} c_k, \qquad 1 \le i, j \le n.$$

Example: Options on equity indices. Let us now consider the real-valued process $(C_t)_{t\ge 0}$ describing the total value of a basket of m assets with prices $f_k(t)$ and constant weights w_k (as done in [2]), i.e.

$$C_t = \sum_{k=1}^{m} w_k f_k(t),$$

where the prices f_k follow the SABR model [15]

$$\begin{aligned} df_k &= a_k f_k^{\beta}\, dW_k, \\ da_k &= \nu_k a_k\, dZ_k, \\ f_k(0) &= f_k^0, \quad a_k(0) = a_k^0 \end{aligned} \tag{5.23}$$

for all $k = 1, \ldots, m$. Here, a_k is the volatility of the asset price f_k, ν_k is the volatility of a_k, $\beta > 0$ is a parameter, $Y := (W_1, \ldots, W_m, Z_1, \ldots, Z_m)$ is a Brownian process with correlation coefficients ρ_{ij} (i.e. $\mathbb{E}(dY_i\, dY_j) = \rho_{ij}\, dt$), and the starting point $(a_1^0, \ldots, a_n^0, f_1^0, \ldots, f_n^0)$ of the process is the spot configuration. For simplicity we only consider the case $\beta = 1$.

To simplify the SDE (5.23) and thus the steepest-descent flow (5.22), we use Itô's formula by applying the change of variables $f_k' = \log(f_k/f_k^0)$ and $a_k' = \log(a_k/a_k^0)$, and we obtain

$$\begin{aligned} df_k' &= \text{drift term} + a_k'\, dW_k, \\ da_k' &= \text{drift term} + \nu_k\, dZ_k, \\ f_k'(0) &= a_k'(0) = 0 \end{aligned} \tag{5.24}$$

for all $k = 1, \ldots, m$. The $2m$-dimensional process $(X_t)_{t \geq 0}$ in this case is

$$X_t = (a_1', \ldots, a_n', f_1', \ldots, f_n')(t), \qquad X_0 = x_1 = 0,$$

and its diffusion matrix is given by $A_{ij}(x) = \rho_{ij}\sigma_i(x)\sigma_j(x)$, for

$$\sigma_k = \begin{cases} a_k' & \text{if } 1 \leq k \leq m, \\ \nu_k & \text{if } m+1 \leq k \leq 2m. \end{cases} \tag{5.25}$$

The set B in formula (5.20) is

$$B_{call} = \Big\{ (a_1', \ldots, a_n', f_1', \ldots, f_n') \in \mathbb{R}^{2m} \,\Big|\, \sum_{k=1}^{n} w_k f_k^0 e^{f_k'} \geq C^S \Big\},$$

for a call option, or

$$B_{put} = \left\{ (a'_1, \ldots, a'_n, f'_1, \ldots, f'_n) \in \mathbb{R}^{2m} \;\middle|\; \sum_{k=1}^{n} w_k f_k^0 e^{f'_k} \leq C^S \right\},$$

for a put option, where C^S is the strike price.

Results. We set $\rho_{ij} = 0$ if $1 \leq i \leq m$ and $m+1 \leq j \leq 2m$ or vice versa. In lack of real historical data to estimate the parameters, we also set $\rho_{ij} = \delta_{ij}$ for $m+1 \leq i, j \leq 2m$ and $\nu_i = 1$ for all $1 \leq i \leq m$. As initial values, we set $a_k^0 = f_k^0 = 1$ for every k, every asset was assigned equal weight $w_k \equiv \frac{1}{m}$, and as the strike price we chose twice the spot price. We discretized our curves at $N = 50$ points and used the stepsize $\Delta\tau = 0.003$. We say that the algorithm converged if the action does not change by more than 10^{-5} if we run the algorithm further.

We first tried our algorithm for a basket of size $m = 2$, with $\rho_{ij} = \delta_{ij}$ also for $1 \leq i, j \leq m$, i.e. for uncorrelated asset prices f_k. Here the gMAM converged after about 5,000 steps ($\approx$ 45 seconds). We found two local minima of the action $\tilde{S}$ (Figure 5.5 shows one of them): one minimum corresponding to the case in which the price of the first asset rises drastically while the second one stays almost stagnant, and by symmetry of the system another local minimum with the roles of the two assets interchanged. Figure 5.5 also shows that the volatily of the price that is responsible for the transition increases fast at the beginning, allowing for easier fluctuations of the asset price.

When we increased the basket size to $m = 100$, it took about 23,000 iterations ($\approx$ 127 minutes) until the gMAM had reached convergence. We observed again

that the optimal way to reach the strike price is that one asset price f_k rises whereas all other asset prices remain approximately the same (no figure).

We then chose again $m = 2$, but now we changed the correlations ρ_{ij}, $1 \leq i, j \leq m$, of the prices f_k so that the two asset prices are closer correlated: we set

$$(\rho_{ij})_{1 \leq i,j \leq m} = \begin{pmatrix} 1 & 1 \\ 1 & 1.04 \end{pmatrix},$$

corresponding to the decomposition $W_1 = \bar{W}_1$, $W_2 = \bar{W}_1 + .2\bar{W}_2$ for independent Brownian motions $\bar{W}_1$ and $\bar{W}_2$. In this case the convergence turned out to be a bit slower than in the uncorrelated case, but after 32,000 steps ($5\frac{1}{2}$ minutes) we obtained a solution, as shown in Figure 5.6. As one would expect, the minimizing path suggests that both prices have to rise simultaneously in this case.

Conclusions. Although the time until convergence for $m = 100$ seems large at first, in practice this would only play a minor role: one would run the gMAM only once until convergence and then use a continuity method by feeding the gMAM with updated estimates for the volatilities and correlation coefficients as time passes by. On the other hand, comparing our measurements to those in Sections 4.1 and 4.3 suggests that one might be able to speed up the convergence a lot more, for example by chosing another preconditioner. Also, when we did further experiments we observed that stronger correlations between the various components of $(X_t)_{t \geq 0}$ can make the PDE very stiff and thus force us to decrease the stepsize $\Delta\tau$ in order to keep the scheme stable – further work at this end may prove fruitful.

Finally, while this section demonstrates that the gMAM is a promising tool to address the problem of options valuation, tests on real-world data would be necessary to check whether the first-order large deviations approximation is justifiable in this case.

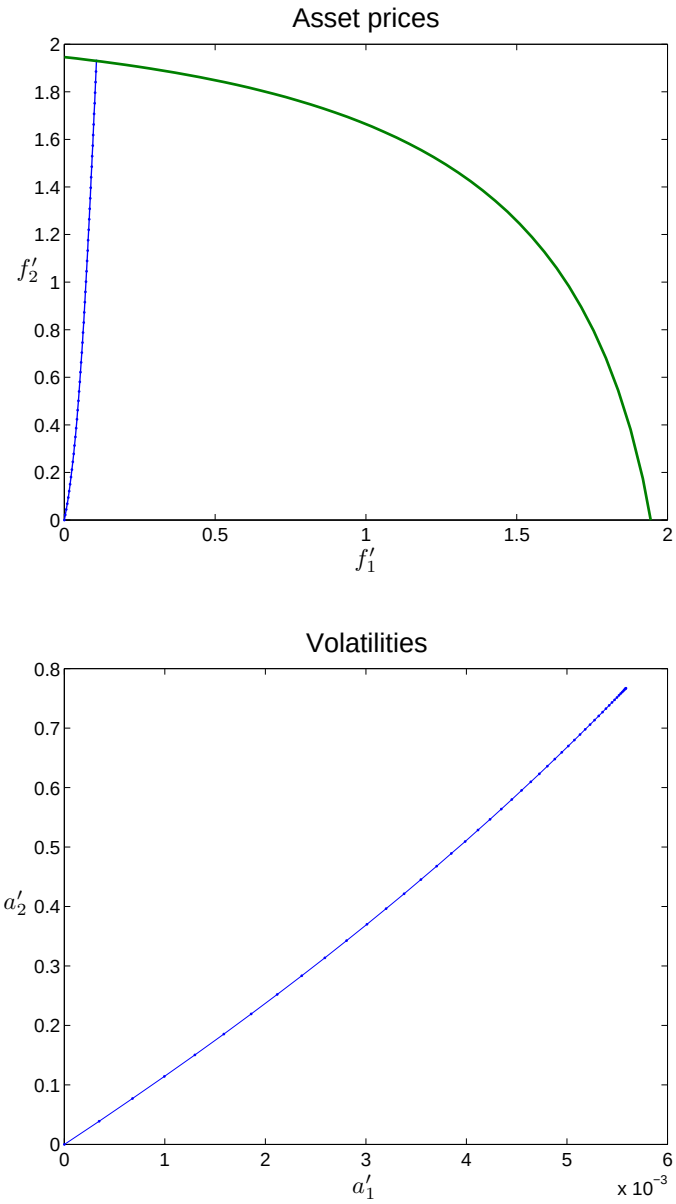

Figure 5.5: One transition path for the basket size $m = 2$ and *uncorrelated* asset prices. The first panel shows that the price f_2' increases while f_1' changes only very little. The green line is the set of points corresponding to the strike price. The volatily a_2' increases rapidly at the beginning, allowing for easier fluctuations of f_2'.

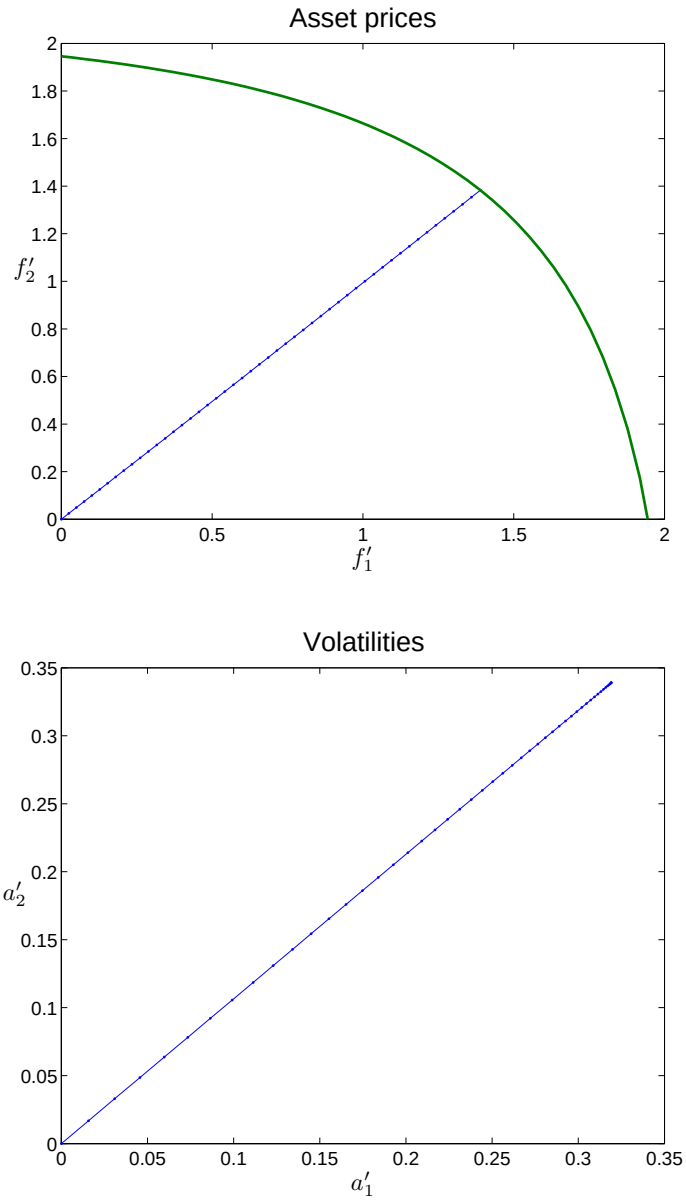

Figure 5.6: A transition path for basket size $m = 2$ and *correlated* asset prices. The prices of both assets as well as their volatilities increase simultaneously.

Chapter 6

Application: Synthetic Biology

6.1 Introduction to Synthetic Biology

Biology is currently undergoing a series of exciting innovations that not only revolutionize the field itself with novel techniques for developing and producing medications or vaccines against the major diseases, but that will ultimately change the way we think of life. With DNA sequencing and synthesis becoming both cheaper and more reliable, one is now able to read out and study DNA sequences of existing organisms, manipulate and reassemble them in the computer at will, and finally synthesize them and introduce them into living cells [10].

As error rates in DNA synthesis have been decreased to only 1 error in about 10000 bases and the maximum length for sequenced DNA has been pushed to up to 35000 bases [4], one has now enough freedom to design the first simple genetic networks that can perform basic tasks: logical building blocks for operations such as AND and XOR that can be assembled to perform binary summation or

counting, inverters that when connected properly form a one-bit-memory (the genetic switch [13]) or a ring-oscillator [9], input-sensors that allow circuits to react on the environment, e.g. on light [21] or heat, and output mechanisms that can return the results of a logical operation, e.g. by making the cell glow [9, 26] or even smell in various scents [27, 17].

The current stage of synthetic biology is comparable to the early years of electrical engineering [3]: The basic building blocks have to be identified, constructed, studied, improved, and finally made available to a new generation of "genetic engineers" who will not have to know the details of how the parts work, but only how to assemble them correctly in order to fulfill a task of interest. The long-term success of synthetic biology, going beyond initial toy projects and leading towards the routine design of sophisticated genetic networks performing complex logical tasks, will thus highly depend on the successful development of a collection of reliable standard biological parts and devices.

At currently about 1$ per base, DNA synthesis is still too expensive to solely use generous trial-and-error methods in order to improve the performance and stability of the basic building blocks. Both financially and qualitatively one will benefit from the aid of advanced mathematical and numerical tools, even if "Moore's law for synthetic biology" will remain valid for a bit longer.

But while both theory and numerics for the simulation and study of the *short*-term behavior of genetic networks have been readily developed – useful models are based on continuous-time Markov jump processes or on their approximations by ordinary differential equations – the tools for the study of *long*-term behavior, especially noise-driven system failure, require improvement in order to be used

routinely and efficiently. This is the main objective of the work presented in this chapter.

Significance of Rare Events in Biological Networks. All reactions in a cell, such as transcription, translation, protein binding or protein degradation, are stochastic: one cannot predict for sure how many reactions of a certain type will happen within a given time interval, but only with which probability each number of reactions will occur. While the probability for some typical behavior of the cell may be close to 1, there is always a small probability that the cell will behave very differently.

The probability of a whole sequence of such unlikely events to occur, ultimately leading to unwanted behavior on the macroscopic scale (system failure), is even smaller, but still nonzero. As a result, if one only waits long enough it will eventually happen in any given cell, and if one observes thousands or millions of cells then the expected time until one of them shows atypical behavior is even shorter.

A genetic engineer who designs a new genetic network would like the expected time until system failure to be much longer than the typical operation time of the network. Part designers should therefore explore and exploit methods to increase the stability of their parts. To do so, the first step is to understand which reactions of the network are most likely to be responsible for system failure. Once this is understood, one can modify this weak component, design backup mechanisms that kick in once atypical behavior is detected, etc.

6.2 The Genetic Switch

The considerations above apply to any network that stores and works with binary information, such as adders, subtracters, counters, or more complex logical networks to be designed in the future, but as a simple example we consider here the plasmid-based genetic switch which has been artificially engineered in [13].

Plasmids are typically short (1-400 kilobase pairs long) circular double-stranded DNA molecules that are mainly found in bacterial cells. They are separate from the main chromosomal DNA and capable of autonomous replication. Their most important feature in the context of this mathematical treatment is that they can exist hundreds or even thousands of times in each cell, since this allows us to apply large deviations theory (which describes the limiting behavior as the system size goes to ∞).

In the genetic switch, a plasmid contains two gene sites a and b that can be transcribed and translated into proteins A and B. Those in turn can form polymers that bind to the operator site of the respective other gene, preventing further production of the corresponding protein. Other reactions are to reverse polymer formation or protein binding, and the degradation of proteins. See Figure 4.16 for an illustration.

This setup leads to bistable behavior: If the cell is in a state with many proteins A and only few proteins B then there are likely also many polymers A_l, the operator site of gene b will be blocked most of the time, and thus only few new proteins B are being produced. Since there are only few proteins B, it is unlikely that a polymer B_m will bind to the operator site of gene a, and

the production rate of new proteins A will stay high. Therefore there will be a stable state with many proteins A and few proteins B, and by symmetry of the mechanism there is another stable state with few proteins A and many proteins B. Bistability arises because the fluctuations leading to a switch from one stable state to the other are rare events.

The cell is in one of these two stable states, and the network can thus be considered as a one-bit memory. However, if a rare sequence of unlikely reactions occurs, it is possible for the cell to spontaneously fall from one stable state into the other, thus losing its information. Part designers want to make these unwanted switches as rare as possible.

Biological Networks as Markov Jump Processes. A network such as the genetic toggle switch illustrated in Figure 4.16 can be modeled as a Markov jump process. Our state space variables are $X_{1/2}$ = number of proteins A/B, $X_{3/4}$ = number of polymers A_l/B_m, and $X_{5/6}$ = number plasmids that are blocked on gene sites a/b. The rescaled state vectors are the protein densities $(x_1, \ldots, x_6) := (\varepsilon X_1, \ldots, \varepsilon X_6)$, where $\Omega := \varepsilon^{-1}$ is the system size parameter (in our case the total number of plasmids in the cell). The dynamics is specified by the reaction list,

$$R_j = (\varepsilon^{-1}\nu_j(x), \varepsilon e_j), \qquad j = 1, \ldots, J, \tag{6.1}$$

where $\varepsilon^{-1}\nu_j(x)$ are the rates, εe_j are the change (or stoichiometric) vectors, and J is the total number of reactions (here $J = 12$): Given a state $x \in D = \mathbb{R}^n_+$ (here $n = 6$), the occurrences of the reactions on an infinitesimal time interval of length dt are independent of each other, and the probability for reaction R_j to

happen during this time interval is given by $\varepsilon^{-1}\nu_j(x)dt$. The state of the system after reaction R_j is $x + \varepsilon e_j$. The twelve reactions of the genetic switch and their rate functions $\nu_j(x)$ are listed in Table 6.1.

Markov jump processes of the type above can be simulated by Gillespie's Stochastic Simulation Algorithm (SSA) [14]. However, the SSA is very inefficient for the study of rare events because of the time scale separation inherent to those events: by definition one would have to simulate the system for a very long time until only one rare event is observed, which makes them virtually inaccessible by direct numerical simulations. In addition, the results of SSA simulations are

reaction type	rate	state change of $(X_1, \ldots, X_6)$
protein production	$\nu_1(x) = k_1^A(1 - x_6)$ $\nu_2(x) = k_1^B(1 - x_5)$	$e_1 = (1, 0, 0, 0, 0, 0)$ $e_2 = (0, 1, 0, 0, 0, 0)$
protein degradation	$\nu_3(x) = k_2^A x_1$ $\nu_4(x) = k_2^B x_2$	$e_3 = (-1, 0, 0, 0, 0, 0)$ $e_4 = (0, -1, 0, 0, 0, 0)$
polymer formation	$\nu_5(x) = k_3^A x_1^l$ $\nu_6(x) = k_3^B x_2^m$	$e_3 = (-l, 0, 1, 0, 0, 0)$ $e_4 = (0, -m, 0, 1, 0, 0)$
polymer degradation	$\nu_7(x) = k_4^A x_3$ $\nu_8(x) = k_4^B x_4$	$e_3 = (l, 0, -1, 0, 0, 0)$ $e_4 = (0, m, 0, -1, 0, 0)$
protein binding	$\nu_9(x) = k_5^A x_3(1 - x_6)$ $\nu_{10}(x) = k_5^B x_4(1 - x_5)$	$e_3 = (0, 0, -1, 0, 0, 1)$ $e_4 = (0, 0, 0, -1, 1, 0)$
protein unbinding	$\nu_{11}(x) = k_6^A x_6$ $\nu_{12}(x) = k_6^B x_5$	$e_3 = (0, 0, 1, 0, 0, -1)$ $e_4 = (0, 0, 0, 1, -1, 0)$

Table 6.1: The reactions of the genetic switch illustrated in Figure 4.16. The variables are: $X_{1/2}$ = number of proteins A/B, $X_{3/4}$ = number of polymers A_l/B_m, $X_{5/6}$ = number of plasmids that are blocked on gene sites a/b. For our example, we chose the parameters symmetrically, $k_i^A = k_i^B$, motivated by values chosen in [32]: $l = m = 2$, $k_1 = 1$, $k_2 = .8$, $k_3 = 5V_c$, $k_4 = 5$, $k_5 = 5V_c$, $k_6 = 5$, where $V_C = 2$.

typically noisy and difficult to interpret, especially if there are many species involved in many reactions.

Large Deviations Theory for Networks. Large Deviations Theory (LDT, [12, 29]) provides the right framework to bypass direct numerical simulations and compute the pathways and rates of rare events in networks in the limit as the system size Ω goes to infinity (i.e. as $\varepsilon \to 0$). LDT characterizes the maximum likelihood transition path leading from one given stable state in state space to another as the minimizer $\psi^\star$ of the action functional

$$S_T(\psi) = \int_0^T L\big(\psi(t), \dot{\psi}(t)\big)\, dt, \qquad \text{where} \tag{6.2}$$

$$L(x,y) = \sup_{\theta \in \mathbb{R}^n} \big(\langle y, \theta\rangle - H(x,\theta)\big) \qquad \text{and} \qquad H(x,\theta) = \sum_{j=1}^{J} \nu_j(x) \left(e^{\langle \theta, e_j\rangle} - 1\right)$$

are the Lagrangian and the Hamiltonian of the theory, expressed in terms of the rates $\nu_j(x)$ and the change vectors e_j. The action (6.2) must be minimized over all times $T > 0$ and over all paths ψ subject to the constraint that they connect the two metastable states of interest. The rate at which this rare transition event occurs is then to leading order given by $\exp(-\Omega S_{T^\star}(\psi^\star))$.

Applying gMAM. However, since both start and end points are critical points, we know from Lemma 3 that $T^\star = \infty$, so that no minimizer $\psi^\star$ exists (see the discussion in Section 2.4). Instead, we now make use of the theory developed in Chapter 2 and work with the geometric action $\hat{S}$ instead. Using the gMAM in Section 3.2 (with the algorithm from Section 3.4 to compute $\hat{\vartheta}(\varphi, \varphi')$ in an inner

loop), we solve the minimization problem (3.1), i.e. we minimize $\hat{S}(\varphi)$ among all curves connecting the two stable states in question, thus neglecting the time parametrization of the curve. The existance of a minimizer $\varphi^\star$ is made plausible by Proposition 1 (iii).

Proposition 2 now tells us that transitions will most likely follow the curve $\gamma(\varphi^\star)$ (notice that Remark 3 (ii) states that in our case the second condition in (2.15) is redundant). Information on the optimal time parametrization away form the critical points can be recovered using the technique presented in Section 2.4. Finally, due to the equality of the expressions in (2.1) and (2.2) the rate for such transitions is to leading order given by $\exp(-\Omega\hat{S}(\varphi^\star))$.

The transition path. Figure 6.1 shows the maximum likelihood transition path for the genetic switch model from the state with many proteins A and few proteins B to the opposite state. The two panels show different projections of the six-dimensional path onto the plane: In the first, the total densities of proteins of the two types are plotted, including the proteins that have formed polymers (i.e. $x_1 + lx_3$ over $x_2 + mx_4$). In the second, the variables x_5 and x_6 are shown, i.e. the fractions of the operator sites that are blocked. The color of the paths indicates the speed of the transition (red=slow, blue=fast).

As one can see, the most likely way for this transition to happen is that the number of proteins A decreases and simultaneously the number of bound operator sites a increases (so that fewer proteins A are can be produced), until the system is in a symmetric state. From there it will fall into the other state simply by following the flow (see Equation (1.13)).

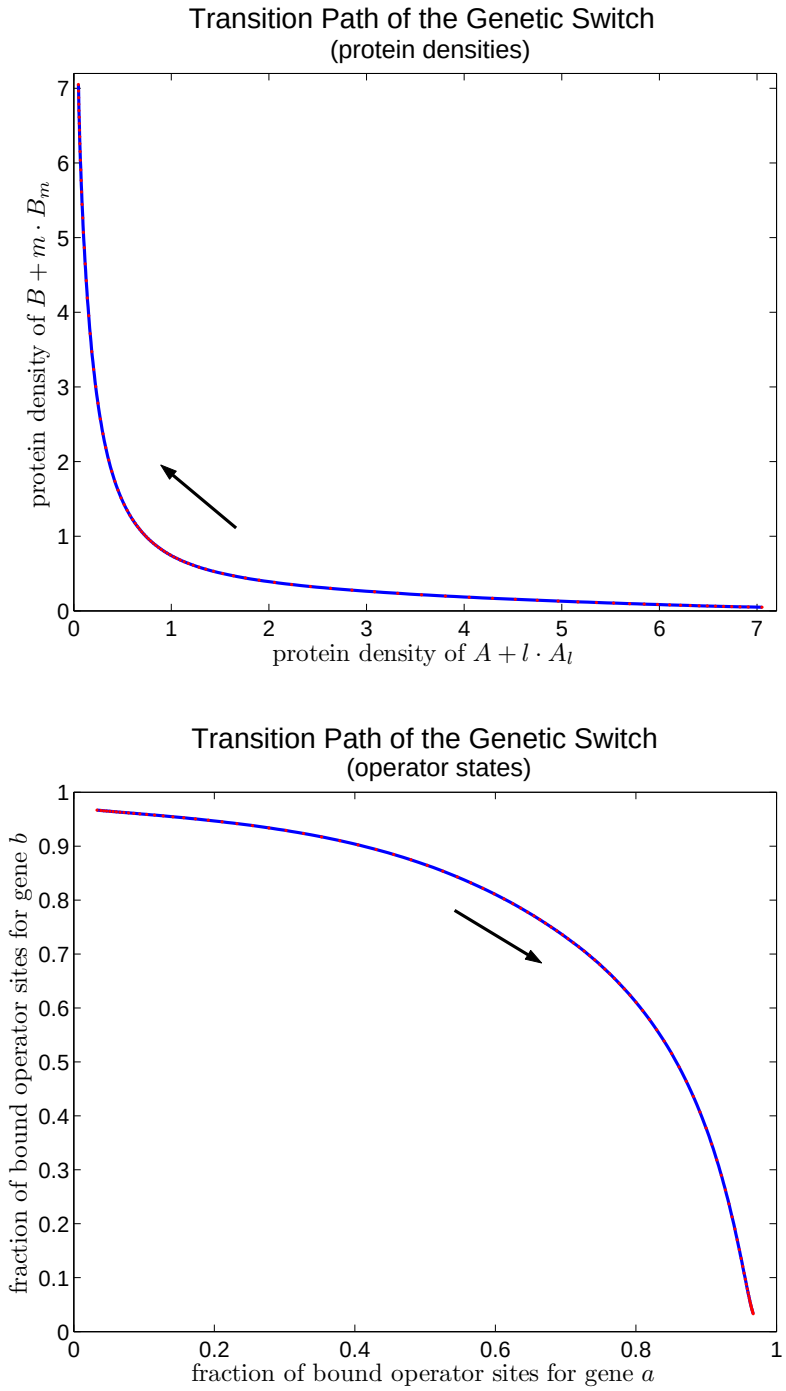

Figure 6.1: The maximum likelihood transition path of the genetic switch.

6.3 A tool to identify sources for instability in networks

While much insight can already be gained from plots of the type in Figure 6.1, it provides only incomplete information about the mechanisms that lead to the rare transition, i.e. what are the frequencies $\mu_j(t)$ at which the reactions R_j happen during the transition. Knowing these frequencies would be helpful because comparing them to the rates $\nu_j(\psi^\star(t))$ that one would expect during *typical* behavior when the process is in state $\psi^\star(t)$ can help spot those reactions that behave more atypical than others. These reactions are the ones that are most likely the reason for system malfunction (i.e. the transition) and that would have to be the target of further efforts in the network design process.

An easy way to spot the responsible reactions is to plot the function $\log\big(\mu_j(t)/\nu_j(\psi^\star(t))\big)$ for each j: the larger the amplitude of this graph, the more atypical the corresponding reaction behaves during the transition. Notice also that this resolves the problem of analyzing transition paths in higher dimensions: with this approach we can simply plot J one-dimensional graphs into one panel, one for each reaction.

Maximum likelihood reaction rates (MLRRs). Let us now prepare for Lemma 9 in which we make the notion of the MLRRs $\mu_j(t)$ precise, and which will tell us how to compute them.

We denote by $E \in \mathbb{R}^{n\times J}$ the matrix whose columns contain the reaction vectors e_j. Let x_1 be some vector in $\mathbb{R}^n$, and let $(N_t^\varepsilon)_{t\geq 0}$ be the continuous-time

Markov chain in $\mathbb{R}^J_+$ starting from 0 and with reactions $(\varepsilon^{-1}\nu_j(x_1 + E\,\cdot\,), \varepsilon u_j)$, where u_j is the j-th unit vector in $\mathbb{R}^J$. Let $(X^\varepsilon_t)_{t\geq 0}$ be the continuous-time Markov process in $\mathbb{R}^n$ defined by $X^\varepsilon_t := x_1 + EN^\varepsilon_t$. Then (X^ε_t) is a continuous-time Markov chain process with reactions $(\varepsilon^{-1}\nu_j(\cdot), \varepsilon e_j)$ as before (starting from x_1), and (N^ε_t) is the associated process that counts the number of reactions of each type (rescaled by a factor of ε).

Thus we would like to know the most likely path of (N^ε_t) that one observes when (X^ε_t) is making a rare transition. The slope of that path at each time t would tell us the most likely number of reactions of each type per time during the transition (again rescaled by ε), i.e. the MLRRs $\mu_j(t)$.

Note that although E is not an invertible matrix, for fixed $\varepsilon > 0$ we can still recover (N^ε_t) from (X^ε_t) by comparing the jumps in the piecewise constant process (X^ε_t) to the reaction vectors e_j. In the limit as $\varepsilon \to 0+$ however this information is getting lost: the optimal transition path $\psi^\star$ of (X^ε_t) (i.e. the minimizer of the action S_T associated with (X^ε_t)) is an absolutely continuous function, and thus there is no trivial way of computing from it the optimal transition path $\chi^\star$ of (N^ε_t). Instead, Lemma 9 will tell us how to find $\chi^\star$, and it gives us a formula for the MLRRs $\mu_j(t)$. Its proof is carried out later in Section 6.4.

In the following we will denote by $\mathcal{D}_d(0,T)$ the Skorokhod space of functions $f : [0,T] \to D$, i.e. the space of all such functions f that are right-continuous and have left limits, equipped with the topology induced by the Skorokhod metric

$$\begin{aligned} d_d(f,g) &:= \inf_{\lambda\in\Lambda} \max\left\{ |f(t) - g(\lambda(t))|_{[0,T]},\ \sup_{s\neq t}\left|\log\left(\frac{\lambda(t)-\lambda(s)}{t-s}\right)\right|\right\}, \\ \Lambda &:= \{\lambda : [0,T] \to [0,T] \,|\, \lambda \text{ continuous, strictly increasing, surjective}\}, \end{aligned}$$

and we will denote by $\mathcal{D}_d^J(0,T)$ the Skorokhod space of functions $f:[0,T]\to\mathbb{R}_+^J$ (with analogously defined metric $d_d^J(f,g)$). (See e.g. [6, Section 8.6] or [29, p. 480-481] for details about the Skorokhod space.)

Lemma 9. *Let $T>0$, let B be a closed subset of $\mathcal{D}_d(0,T)$ consisting only of functions starting from x_1. Assume that B is regular with respect to the action S_T associated to the process (X_t^ε), and that the minimizer*

$$\psi^\star = \operatorname*{argmin}_{\psi\in B} S_T(\psi)$$

exists and is unique. Define the MLRRs $\mu_j:[0,T]\to(0,\infty)$ and the path $\chi^\star:[0,T]\to\mathbb{R}^J$ by

$$\mu_j(t) := \nu_j(\psi^\star(t))\exp\Big(\big\langle\theta^\star\big(\psi^\star(t),\psi^{\star\prime}(t)\big),e_j\big\rangle\Big), \tag{6.3}$$

$$\chi_j^\star(t) = \int_0^t \mu_j(\tau)\,d\tau, \tag{6.4}$$

for every $j=1,\dots,J$ and every $t\in[0,T]$, where $\theta^\star(x,y)$ is defined implicitly by (1.23), i.e. by the equation $y=\sum_{j=1}^J \nu_j(x)e^{\langle\theta^\star(x,y),e_j\rangle}e_j$.

Then for every $\eta>0$ we have

$$\lim_{\varepsilon\to0+}\mathbb{P}\big(|N_t^\varepsilon-\chi^\star(t)|_{[0,T]}<\eta\,\big|\,X^\varepsilon\in B\big)=1, \tag{6.5}$$

Equations (6.5) and (6.4) tell us that if the event B occurs then $dN_j^\varepsilon\approx d\chi_j^\star\approx\mu_j(t)\,dt$ for small ε, whereas during typical behavior we would expect that $dN_j^\varepsilon\approx\nu_j(\psi^\star(t))\,dt$. Using (6.3), the logarithm of the ratio of these two terms is

given by

$$\log \frac{\mu_j}{\nu_j(\psi^\star)} = \langle \theta^\star(\psi^\star, \psi^{\star\prime}), e_j \rangle, \qquad j = 1, \ldots, J, \quad t \in [0, T]. \tag{6.6}$$

In short, our diagnostic plot explained at the beginning of this section consists of the components of the vector function

$$t \mapsto \Big(\log \frac{\mu_j}{\nu_j(\psi^\star)} \Big)_{j=1,\ldots,J} = E^T \theta^\star(\psi^\star, \psi^{\star\prime}). \tag{6.7}$$

If in addition T is the *optimal* time to traverse $\gamma(\psi^\star)$, i.e. $T = T^\star(\varphi^\star)$, where $\gamma(\varphi^\star) = \gamma(\psi^\star)$, then $\theta^\star(\psi^\star(t), \psi^{\star\prime}(t)) = \hat{\vartheta}(\varphi^\star, \varphi^{\star\prime})|_{\alpha=\alpha(t)}$, and we obtain a time-rescaled version of (6.7) by plotting instead

$$\alpha \mapsto \Big(\log \frac{\mu_j}{\nu_j(\psi^\star)}\Big|_{t=t(\alpha)} \Big)_{j=1,\ldots,J} = E^T \hat{\vartheta}(\varphi^\star, \varphi^{\star\prime}). \tag{6.8}$$

As we see, after using the gMAM to find $\varphi^\star$, the computation of the diagnostic plot comes entirely for free.

We have thus discovered another interpretation of $\hat{\vartheta}(x, y)$: it is a measure of the "unlikeliness" of moving into the direction y when the system is in state x.

Results. Figure 6.2 shows the MLRRs for the genetic switch model from Section 6.2 in two ways: In the first panel one can see a plot of the functions $\alpha \mapsto \mu_j(t(\alpha))$, in the second panel they are plotted in relation to the typical rates $\nu_j(\psi^\star(t))$, using formula (6.8). While the first one can be used to see how active the various reactions are during the transition, the second one is a measure of

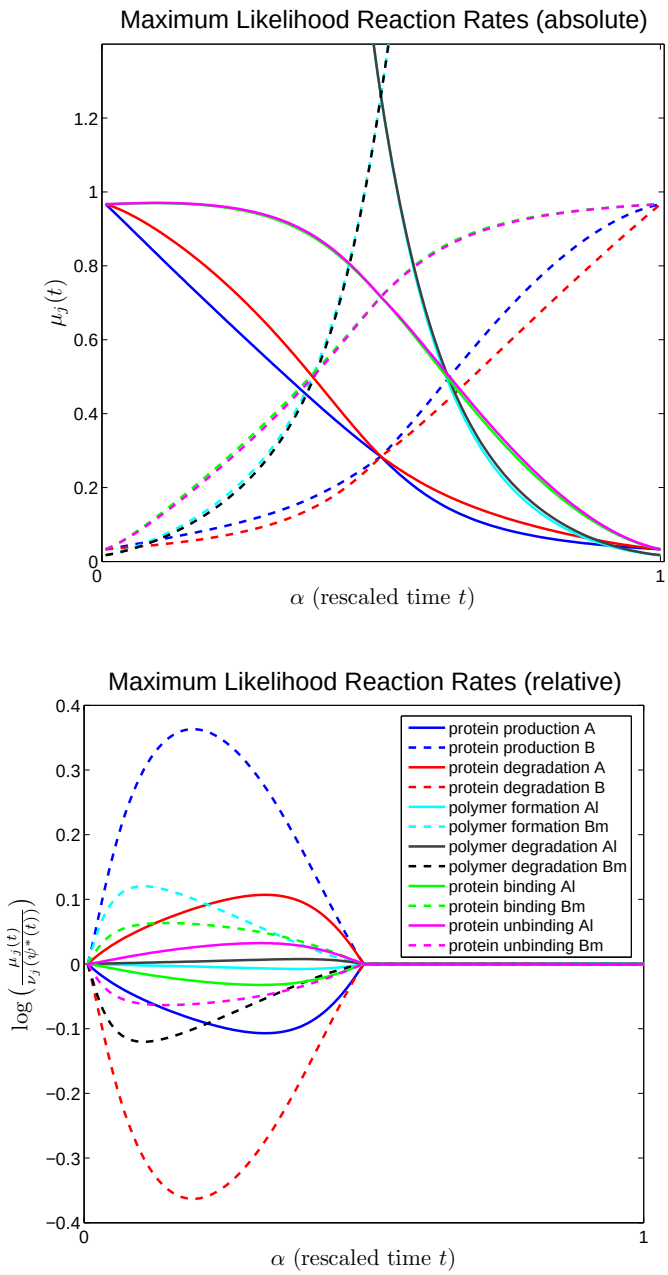

Figure 6.2: (color plot) The maximum likelihood reaction rates, as laid out in Section 6.2. The upper panel shows the values of the rates $\mu_j(t)$ for all 12 reactions (see the lower panel for the legend). The lower panel shows their relations to the rates during typical behavior, given by the logarithm of their ratios: Values greater or lower than 0 indicate abnormally high or low reaction rates, respectively.

how atypical the various reactions behave.

At this point we will not draw any conclusions from these plots since our parameters do not correspond to those of any actual biological organism. Note however that in the second panel one can see that all reactions behave normally during the second half of the transition. This is a feature that is common among all networks, caused by the fact that the system has entered the domain of attraction to the second stable state and can just follow the flow from there on.

6.4 Proof of Lemma 9

Proof of Lemma 9. Let S_T^+ denote the action associated with the process (N_t^ε), and let

$$B_+ := \{\chi \in \mathcal{D}_d^J(0,T) \,|\, \chi(0) = 0,\ x_1 + E\chi \in B\}.$$

Observe that B_+ is a closed subset of $\mathcal{D}_d^J(0,T)$ since it is the pre-image of the closed set $\{0\} \times B$ under the continuous mapping $F : \mathcal{D}_d^J(0,T) \to \mathbb{R} \times \mathcal{D}_d(0,T)$, $F(\chi) := (\chi(0), x_1 + E\chi)$. This notation allows us to rewrite (6.5) as

$$\lim_{\varepsilon \to 0+} \mathbb{P}\big(|N_t^\varepsilon - \chi^\star(t)|_{[0,T]} < \eta \,\big|\, N^\varepsilon \in B_+\big) = 1 \tag{6.9}$$

for every $\eta > 0$. The continuity of $\chi^\star$ on $[0,T]$ now enables us to replace $|N_t^\varepsilon - \chi^\star(t)|_{[0,T]}$ by $d_d(N^\varepsilon, \chi^\star)$: if η is small and $d_d(N^\varepsilon, \chi^\star) < \eta$ then there exists a rescaling $\lambda \in \Lambda$ such that $|\lambda(t) - t|_{[0,T]} < 2T\eta$ and $|N_t^\varepsilon - \chi^\star(\lambda(t))|_{[0,T]} < 2\eta$, and we find that

$$|N_t^\varepsilon - \chi^\star(t)|_{[0,T]} \leq |N_t^\varepsilon - \chi^\star(\lambda(t))|_{[0,T]} + |\chi^\star(\lambda(t)) - \chi^\star(t)|_{[0,T]} < 2\eta + \delta(\eta),$$

for some function $\delta(\eta)$ with $\lim_{\eta\to 0+}\delta(\eta) = 0$ (obtained using the uniform continuity of $\chi^\star$). We conclude that if $d_d(N^\varepsilon, \chi^\star)$ is small then $|N_t^\varepsilon - \chi^\star(t)|_{[0,T]}$ is small. Therefore, (6.9) and thus (6.5) follows if we can show that

$$\lim_{\varepsilon\to 0+} \mathbb{P}\big(d_d(N^\varepsilon, \chi^\star) < \eta \,\big|\, N^\varepsilon \in B_+\big) = 1 \tag{6.10}$$

for all $\eta > 0$. Finally, we will see at the end of this proof that B_+ is regular with respect to S_T^+, so using [12, Theorem 3.4] and the closedness of B_+, it only remains to show that $\chi^\star$ as given by (6.3) - (6.4) is the unique minimizer of S_T^+ over the set B_+.

Do to so, observe first that if we denote the local actions of S_T and S_T^+ by ℓ and ℓ^+, respectively, then by considering $\Theta = E^T\theta^\star$ for $\theta^\star = \theta^\star\big(x_1 + E\chi, (x_1 + E\chi)^{\cdot}\big)$ one can show the estimate

$$\begin{aligned}
\ell^+(\chi, \dot\chi) &= \sup_{\Theta\in\mathbb{R}^J}\Big(\langle\Theta, \dot\chi\rangle - \sum_{j=1}^J \nu_j(x_1 + E\chi)\big(e^{\langle\Theta, u_j\rangle} - 1\big)\Big) && (6.11)\\
&\geq \big\langle\theta^\star, E\dot\chi\big\rangle - \sum_{j=1}^J \nu_j(x_1 + E\chi)\big(e^{\langle\theta^\star, Eu_j\rangle} - 1\big) && (6.12)\\
&= \big\langle\theta^\star, (x_1 + E\chi)^{\cdot}\big\rangle - \sum_{j=1}^J \nu_j(x_1 + E\chi)\big(e^{\langle\theta^\star, e_j\rangle} - 1\big)\\
&= \sup_{\theta\in\mathbb{R}^n}\Big(\langle\theta, (x_1 + E\chi)^{\cdot}\rangle - \sum_{j=1}^J \nu_j(x_1 + E\chi)\big(e^{\langle\theta, e_j\rangle} - 1\big)\Big)\\
&= \ell\big(x_1 + E\chi, (x_1 + E\chi)^{\cdot}\big).
\end{aligned}$$

We have equality in (6.12) if and only if the maximizing Θ in (6.11) is $\Theta = E^T\theta^\star$, i.e. if

$$\dot{\chi} \;=\; \sum_{j=1}^{J} \nu_j(x_1 + E\chi)e^{\langle\Theta,u_j\rangle}u_j\Big|_{\Theta=E^T\theta^\star} = \sum_{j=1}^{J} \nu_j(x_1 + E\chi)e^{\langle\theta^\star,e_j\rangle}u_j = \mu^{x_1+E\chi},$$

where for every absolutely continuous function $\psi \in B$ we define the vector function $\mu^\psi(t)$ by

$$\mu_j^\psi := \nu_j(\psi)e^{\langle\theta^\star(\psi,\dot{\psi}),e_j\rangle}, \qquad j = 1, \ldots, J.$$

Integrating the estimate for the local actions above from 0 to T, we have

$$\forall\chi \in B_+ : \quad S_T^+(\chi) \geq S_T(x_1 + E\chi) \geq \inf_{\psi\in B} S_T(\psi) = S_T(\psi^\star) \tag{6.13}$$

(note that if χ is not absolutely continuous then (6.13) is trivially true since the left hand side equals ∞). Equality holds in the second and first step in (6.13), respectively, if and only if

$$x_1 + E\chi = \psi^\star \qquad \forall t \in [0,T] \text{ and} \tag{6.14}$$

$$\dot{\chi} = \mu^{x_1+E\chi} = \mu^{\psi^\star} \qquad \text{for a.e. } t \in [0,T] \tag{6.15}$$

(using also the uniqueness of the minimizer $\psi^\star$). This shows that $\chi^\star$ as defined by (6.3) - (6.4) is the unique minimizer of S_T^+ over B_+ if and only if it is the unique function in B_+ fulfilling (6.14) and (6.15). That in turn is indeed the case: Clearly, there can only be one function in B_+ fulfilling (6.15), which shows uniqueness. On the other hand, $\chi^\star$ fulfills (6.15) by definition, and it also fulfills

(6.14) since

$$\begin{aligned}(x_1 + E\chi^\star)^\cdot &= E\dot{\chi}^\star = E\mu(t) = E\sum_{j=1}^{J} \nu_j(\psi^\star) e^{\langle \theta^\star(\psi^\star, \dot{\psi}^\star), e_j \rangle} u_j \\ &= \sum_{j=1}^{J} \nu_j(\psi^\star) e^{\langle \theta^\star(\psi^\star, \dot{\psi}^\star), e_j \rangle} e_j = \dot{\psi}^\star\end{aligned}$$

for a.e. $t \in [0, T]$ by definition of $\theta^\star$, and since $(x_1 + E\chi^\star)|_{t=0} = x_1 + E0 = x_1 = \psi^\star(0)$. Finally, $\chi^\star$ is in B_+ because it is continuous and because it fulfills (6.14), where $\psi^\star \in B$. This terminates the proof that $\chi^\star$ is the unique minimizer of S_T^+ over B_+.

To see that B_+ is regular with respect to S_T^+, first repeat the argument above to show that for every absolutely continuous $\psi \in B^0$ there exists a (unique) $\chi \in B_+$, given by $\chi(t) := \int_0^t \mu^\psi(\tau)\, d\tau$, that fulfills $\psi = x_1 + E\chi$ and $S_T^+(\chi) = S_T(\psi)$. In this case χ is in fact in B_+^0: if $\delta > 0$ and $\tilde{\chi} \in \mathcal{D}_d^J(0, T)$ is such that $d_d^J(\tilde{\chi}, \chi) < \delta$, then

$$d_d(x_1 + E\tilde{\chi}, \psi) = d_d(x_1 + E\tilde{\chi}, x_1 + E\chi) = d_d(E\tilde{\chi}, E\chi) \leq \max(|E|, 1)\delta,$$

and thus for small enough δ we have $x_1 + E\tilde{\chi} \in B$, i.e. $\tilde{\chi} \in B_+$. Therefore we can conclude that

$$\inf_{\chi \in B_+^0} S_T^+(\chi) \leq \inf_{\psi \in B^0} S_T(\psi) = S_T(\psi^\star)$$

(using the regularity of B). Since we also know from (6.13) that

$$\inf_{\chi \in B_+} S_T^+(\chi) \geq S_T(\psi^\star), \tag{6.16}$$

this shows that $\inf_{\chi \in B_+^0} S_T^+(\chi) \leq \inf_{\chi \in B_+} S_T^+(\chi)$, and therefore (since B_+ is closed) that B_+ is regular with respect to S_T^+. $\square$

Chapter 7

Conclusions

Summary. We have proposed a variant of the MAM, the geometric minimum action method (gMAM), which is tailored to the double minimization problem required to compute the quasipotential $V(x_1, x_2)$ in Freidlin-Wentzell theory of large deviations. The key idea behind the gMAM is to reformulate of the Freidlin-Wentzell action functional on the space of curves. With this reformulation, we guarantee that the new action will have minimizers (that is, curves) in a broader class of situations, in particular when the points x_1 and x_2 in $V(x_1, x_2)$ are stable equilibrium points of the deterministic dynamics (in contrast, the original action minimized both over the paths and their length in time fails to have a minimizer in this case). The corresponding minimizer of the action is the curve of maximum likelihood by which the transitions between these stable equilibrium points occur due to the presence of the small noise.

We demonstrated on several examples that the gMAM adjusts easily to a variety of situations, such as different types of dynamics (SDE, SPDE, continuous-

time Markov chains) or problems with endpoint constraints or penalties, and we showed how the gMAM can be useful also for the determination of short-term asymptotics of an SDE (Section 5.4). We finally dedicated one chapter to potential applications in the field of synthetic biology, and we built a tool on top of the gMAM that can help detect the sources of instabilities in (genetic) networks.

Future work. *Theoretical aspects.* We believe that the potential uses of our geometric action $\hat{S}(\varphi)$ have not yet been fully exploited in this thesis. In particular, it may be possible to prove a large deviations principle on the space of curves. While the associated action would much likely be $\hat{S}$, the appropriate function space and the precise form of the metric defined on it (which will be closely related to the Fréchet distance) are yet to be found. Questions that one may address with such a large deviations principle are those about rare events that do not depend on the time-parametrization of the random path, such as the rare event that a process hits a given set A before it hits another set B to which the process is attracted. The search for a large deviations principle based on $\hat{S}$ will certainly be one of our points of focus in the close future.

Algorithmical aspects. While our rather simple approach to minimize $\hat{S}(\varphi)$ via preconditioned steepest descent and reparametrization has already carried us quite far, there will be ways to improve upon our scheme. In particular, we have not yet investigated the use of quasi-Newton methods such as BFGS, or of multiscale methods that adjust the number of grid-points as the algorithm progresses. Also, in Sections 5.4 and 6.2 we noticed that if the parameters of

the model single out a submanifold of the state space that the process is very unlikely to leave then this can lead to a very stiff steepest-descent PDE and force us to drastically reduce the stepsize $\Delta\tau$. Since this situation is likely to arise in higher-dimensional continuous-time Markov chains whose reaction rates have different orders of magnitude, resolving this issue will be important for the use of the gMAM in the context of Chapter 6.

Reaction-diffusion systems. In this thesis our work on the Maier-Stein model mainly emphasized algorithmical aspects, i.e. we wanted to demonstrate that the gMAM can indeed be applied even to two-dimensional SPDEs using only MatLab on a regular PC. It will now be a natural next step to actually investigate several interesting aspects of this model, especially the role that the v-field plays in the transition. We already saw that the v-field assists in the nucleation part of the process, i.e. it is nonzero during the initial phase of the transition when the u-field leaves the stable point. Our expectation that the v-field would also assist in the domain wall propagation in the two-dimensional SPDE however turned out to be wrong as it seems: Figure 4.13 illustrates that the v-field vanishes long before the process has reached the saddle point. Further questions to investigate are the analytical computation of the sharp-interface-limit as $\kappa \to 0+$.

Synthetic biology. In this thesis a lot of work has been done to open the door to potentially useful applications of the gMAM to the newly arising field of synthetic biology. In lack of time however so far we could not apply the tools that we have developed to actual biological systems from the literature, with real data. Our goal for the close future is to get the attention of scientists from other fields and to explore possible ways of collaboration with genetic engineers.

Other applications. Finally, the potential use of the gMAM is not confined to the realm of large deviations theory. In particular, all our calculations (including the proof of Proposition 1) rely only on the Assumptions 1–3 for the Hamiltonian $H(x,\theta)$. One examplary problem outside of large deviations theory that still fits into the framework of this thesis is the determination of the instanton by which quantum tunneling arises (for background on this problem see e.g. [20, 30]). The relevant minimization problem in this case is

$$V(x_1,x_2) = \inf_{T>0} \inf_{\psi\in\bar{C}^{x_2}_{x_1}(0,T)} \int_0^T \left(\tfrac{1}{2}|\dot{\psi}(t)|^2 + U(\psi(t))\right) dt. \tag{7.1}$$

Here x_1 and x_2 are minima of the potential $U \geq 0$, and it is assumed that $U(x_1) = U(x_2) = 0$. Hence x_1 and x_2 are critical points according to our definition (1.24). It is well-known [30] that this minimization problem can be recast into a geodesic problem in terms of the Agmon distance, i.e. $V(x_1,x_2)$ can be expressed as

$$V(x_1,x_2) = \inf_{\varphi\in\bar{C}^{x_2}_{x_1}(0,1)} 2\int_0^1 \sqrt{U(\varphi(\alpha))}\,|\varphi'(\alpha)|\,d\alpha. \tag{7.2}$$

The instanton is the minimizer of this action. The gMAM can be straightforwardly applied to (7.2) since the corresponding Hamiltonian $H(x,\theta) = \frac{1}{2}|\theta|^2 - U(x)$ fulfills the Assumptions 1–3 of our paper.

Other uses of the gMAM include similar problems involving finding geodesics in high-dimensional space with Riemannian metric.

Appendix A

Three technical lemmas

The goal in Appendix A is to prove Lemmas 10 and 12 which are needed in the proof of Proposition 1 (i).

Lemma 10. *Let $\psi_1 \in \bar{C}(0, T_1)$ and $\psi_2 \in \bar{C}(0, T_2)$ with $\gamma(\psi_1) = \gamma(\psi_2)$, and let the local action $\ell : D \times \mathbb{R}^n \to [0, \infty)$ have the property that for all $x \in D$, $y \in \mathbb{R}^n$ and $c \geq 0$ we have $\ell(x, cy) = c\ell(x, y)$. Then*

$$\int_0^{T_1} \ell(\psi_1, \psi_1')\, dt = \int_0^{T_2} \ell(\psi_2, \psi_2')\, dt.$$

Proof. Let $\psi_1(t) = \varphi(\alpha(t))$ for all $t \in [0, T_1]$, some $\varphi \in \bar{C}(0, 1)$ with $|\varphi'(\alpha)| \equiv cst$ a.e., and for some absolutely continuous rescaling $\alpha : [0, 1] \to [0, T_1]$ with $\alpha' \geq 0$ almost everywhere. Then for all $t \in [0, T_1]$ we have $\dot{\psi}_1(t) = \varphi'(\alpha(t))\alpha'(t)$, and we can compute

$$\int_0^{T_1} \ell(\psi_1, \dot{\psi}_1)\, dt \;=\; \int_0^{T_1} \ell\Big(\varphi\big(\alpha(t)\big), \varphi'\big(\alpha(t)\big)\alpha'(t)\Big)\, dt$$

$$
\begin{aligned}
&= \int_0^{T_1} \ell\Big(\varphi\big(\alpha(t)\big), \varphi'\big(\alpha(t)\big)\Big)\alpha'(t)\,dt \\
&= \int_0^1 \ell\big(\varphi(\alpha), \varphi'(\alpha)\big)\,d\alpha.
\end{aligned}
$$

Since the same calculation can be made for ψ_2, we are done. □

To prepare for the proof of Lemma 12, we need to show some technical properties of H and $\theta^\star$ first.

Lemma 11. *(i) The following equalities hold:*

$$L_y(x,y) = \theta^\star(x,y), \tag{A.1}$$

$$\theta^\star_y(x,y) = H^{-1}_{\theta\theta}(x,\theta^\star(x,y)) = L_{yy}(x,y). \tag{A.2}$$

(ii) Assumption 3 implies the limits

$$\lim_{\theta\to\infty} H(x,\theta) = \infty \qquad \textit{and} \tag{A.3}$$

$$\lim_{y\to\infty} \theta^\star(x,y) = \infty, \tag{A.4}$$

uniformly in x on compact sets.

Proof. (i) By differentiating (1.22) with respect to y and using (1.23), we obtain

$$L_y(x,y) = \theta^\star(x,y) + \big(\theta^\star_y(x,y)\big)^T\big(y - H_\theta(x,\theta^\star(x,y))\big) = \theta^\star(x,y),$$

Differentiating (1.23) with respect to y leads us to

$$H_{\theta\theta}(x, \theta^\star(x, y))\, \theta^\star_y(x, y) = I$$

and thus

$$H^{-1}_{\theta\theta}(x, \theta^\star(x, y)) = \theta^\star_y(x, y) = L_{yy}(x, y).$$

(ii) Denoting $\hat{e}_\theta := \theta/|\theta|$, for any compact set $K \subset D$ Assumption 3 implies that

$$\begin{aligned} \langle H_\theta(x, \theta), \hat{e}_\theta \rangle &= \int_0^{|\theta|} \langle \hat{e}_\theta, H_{\theta\theta}(x, \tau\hat{e}_\theta)\hat{e}_\theta \rangle \, d\tau + \langle H_\theta(x, 0), \hat{e}_\theta \rangle \\ &\geq m(x)|\theta| - \sup_{x \in K} |H_\theta(x, 0)| \geq m_K|\theta| - C_K. \end{aligned} \tag{A.5}$$

Performing one more integration, we find

$$\begin{aligned} H(x, \theta) &= \int_0^{|\theta|} \langle H_\theta(x, \tau\hat{e}_\theta), \hat{e}_\theta \rangle \, d\tau + H(x, 0) \\ &\geq \int_0^{|\theta|} (m_K\tau - C_K)\, d\tau - C'_K = \frac{1}{2} m_K|\theta|^2 - C_K|\theta| - C'_K, \end{aligned}$$

proving (A.3).

To prove (A.4), assume that it is not true. Then there exists a sequence (y_k) with $y_k \to \infty$ and a bounded sequence (x_k) such that the sequence $(\theta^\star(x_k, y_k))$ is bounded. But continuity of $H_\theta(\cdot, \cdot)$ then implies that also $y_k = H_\theta(x_k, \theta^\star(x_k, y_k))$ stays bounded, and we have a contradiction. □

Now we are ready to prove Lemma 12.

Lemma 12. *For the functions λ_k defined in the proof of Proposition 1 we have*

$|\lambda_k|_\infty < \infty$ *and*

$$\sup_{k\in\mathbb{N}} |L(\varphi, \varphi'\lambda_k)/\lambda_k|_\infty < \infty, \tag{A.6}$$

where $|\cdot|_\infty$ *denotes the* L^∞*-norm on* $[0,1]$.

Proof. First let us show that $M := |\lambda(\varphi, \varphi')|_\infty < \infty$. To do so, suppose $M = \infty$. Then for every $l \in \mathbb{N}$ the set $\{\alpha \in [0,1] \,|\, \lambda(\varphi, \varphi') > l\}$ would have non-zero measure, and we could construct a sequence (α_l) such that $\lim_{l\to\infty} \lambda(\varphi(\alpha_l), \varphi'(\alpha_l)) = \infty$ and $|\varphi'(\alpha_l)| = L_\varphi$ for every $l \in \mathbb{N}$, where $L_\varphi > 0$ is the constant such that $|\varphi'| \equiv L_\varphi$ almost everywhere. Now comparing (1.23) with (2.7), we see that $\hat{\vartheta}(\varphi, \varphi') = \theta^\star(\varphi, \varphi'\lambda)$, so that

$$0 \equiv H\big(\varphi, \hat{\vartheta}(\varphi, \varphi')\big)\big|_{\alpha=\alpha_l} = H\big(\varphi, \theta^\star(\varphi, \varphi'\lambda)\big)\big|_{\alpha=\alpha_l} \to \infty \quad \text{as } l \to \infty$$

by Lemma 11 (ii), and we would have a contradiction. This shows that $M < \infty$, and thus also that $|\lambda_k|_\infty \le \max\{M, \frac{1}{k}\} < \infty$. Now we can begin our estimate by showing that

$$\begin{aligned}
|H(\varphi, \theta^\star(\varphi, \varphi'\lambda_k))| &\le \big|H(\varphi, \theta^\star(\varphi, \varphi'\lambda))\big| \\
&\qquad + \int_\lambda^{\lambda_k} \big|\partial_\tau H(\varphi, \theta^\star(\varphi, \tau\varphi'))\big|\, d\tau \\
&= 0 + \int_\lambda^{\lambda_k} \Big|\big\langle H_\theta\big(\varphi, \theta^\star(\varphi, \tau\varphi')\big), \theta^\star_y(\varphi, \tau\varphi')\varphi'\big\rangle\Big|\, d\tau \\
&= \int_\lambda^{\lambda_k} \tau \Big|\big\langle \varphi', H^{-1}_{\theta\theta}\big(\varphi, \theta^\star(\varphi, \tau\varphi')\big)\varphi'\big\rangle\Big|\, d\tau \\
&\le \frac{L_\varphi^2}{m_K} \int_\lambda^{\lambda_k} \tau\, d\tau = \frac{L_\varphi^2}{2m_K}\Big(\lambda_k^2 - \lambda^2\Big) \le \frac{L_\varphi^2 \lambda_k^2}{2m_K},
\end{aligned}$$

where in the third and fourth step we used (1.23), (A.2), and Assumption 3 with $K := \gamma(\varphi)$ and $\xi = H^{-1/2}\varphi'$. Thus

$$\left|H\big(\varphi, \theta^\star(\varphi, \varphi'\lambda_k)\big)/\lambda_k\right| \leq \frac{L_\varphi^2 \lambda_k}{2m_K} \leq \frac{L_\varphi^2 \max\{1, M\}}{2m_K} =: C < \infty,$$

and we obtain the bound

$$\begin{aligned} |L(\varphi, \varphi'\lambda_k)/\lambda_k| &\leq \left|\langle \theta^\star(\varphi, \varphi'\lambda_k), \varphi'\rangle\right| + \left|H(\varphi, \theta^\star(\varphi, \varphi'\lambda_k))/\lambda_k\right| \\ &\leq L_\varphi \max\left\{|\theta^\star(x, y)| \,\middle|\, x \in \gamma(\varphi),\ |y| \leq L_\varphi \max\{1, M\}\right\} + C \\ &< \infty. \end{aligned}$$

□

Appendix B

Proofs of Lemmas 1 and 3

Proof of Lemma 1. (i) and (ii): If $\lambda(x, y) = 0$ then the second equation in (2.7) tells us that $\hat{\vartheta}(x, y)$ is a minimizer of $H(x, \cdot)$ (which by Assumption 3 is unique). Because of the first equation in (2.7) we thus have $\inf_{\theta \in \mathbb{R}^n} H(x, \theta) = 0$, and together with Assumption 1 this implies that $H(x, 0) = 0$. But this means that 0 minimizes $H(x, \cdot)$, so we must have $\hat{\vartheta}(x, y) = 0$. Now the second equation in (2.7) finally says that $H_\theta(x, 0) = 0$, and 0 must be a critical point.

To show the reverse direction, observe that if x is a critical point then $(\lambda, \hat{\vartheta}) = (0, 0)$ solves (2.7).

(iii): Let $x \in D$ be a critical point. By definition (1.23) of $\theta^\star$, $H_\theta(x, 0) = 0$ tells us that $\theta^\star(x, 0) = 0$. Therefore x fulfills $L(x, 0) = \langle \theta^\star(x, 0), 0 \rangle - H(x, \theta^\star(x, 0)) = 0 - H(x, 0) = 0$, and we can apply l'Hospital's rule and use (A.1) to find the limit

$$\lim_{\lambda \to 0+} L(x, \lambda y)/\lambda = \langle y, L_y(x, 0) \rangle = \langle y, \theta^\star(x, 0) \rangle = 0.$$

(iv): For the representation (2.5) this follows from part (ii), for the representation

(2.6) it follows from parts (i) and (iii) combined. For the representation (2.4), observe that if φ is a critical point then $\vartheta = 0$ is the minimum of $H(\varphi, \cdot)$, and that thus we have $H(\varphi, \vartheta) = 0$ only for $\vartheta = 0$. $\square$

Proof of Lemma 3. (i) Let L_φ be the constant such that $|\varphi'| \equiv L_\varphi$ a.e., and let $\alpha \in [0, 1]$ such that $|\varphi'(\alpha)| = L_\varphi$. Denote $\varphi := \varphi(\alpha)$, $\varphi' := \varphi'(\alpha)$, $\hat{\vartheta} := \hat{\vartheta}(\varphi, \varphi')$, $\lambda := \lambda(\varphi, \varphi')$, and finally $\varphi_c := \varphi(\alpha_c)$.

Since φ_c is a critical point, we have $H_\theta(\varphi_c, 0) = 0$ and $H(\varphi_c, 0) = 0$, and the latter equality together with Assumption 1 tells us that also $H_x(\varphi_c, 0) = 0$. Thus if we expand $H(\varphi, \hat{\vartheta})$, which is zero by definition of $\hat{\vartheta}$, around the point $(x, \theta) = (\varphi_c, 0)$, the zeroth- and first-order terms vanish, and we obtain

$$\begin{aligned} 0 &= H(\varphi, \hat{\vartheta}) \\ &= \tfrac{1}{2}\langle \varphi - \varphi_c, H_{xx}(\tilde{x}, \tilde{\theta})(\varphi - \varphi_c)\rangle + \langle \hat{\vartheta}, H_{\theta x}(\tilde{x}, \tilde{\theta})(\varphi - \varphi_c)\rangle + \tfrac{1}{2}\langle \hat{\vartheta}, H_{\theta\theta}(\tilde{x}, \tilde{\theta})\hat{\vartheta}\rangle \end{aligned}$$

for some point $(\tilde{x}, \tilde{\theta})$ on the straight line between $(\varphi_c, 0)$ and $(\varphi, \hat{\vartheta})$. Note that φ is in the compact set $\gamma(\varphi)$, and so Equation (A.3) of Lemma 11 and the first equation in (2.7) tell us that also $\hat{\vartheta}$ must lie within some compact set independent of α. Since $|\tilde{x} - \varphi_c| \leq |\varphi - \varphi_c|$ and $|\tilde{\theta}| \leq |\hat{\vartheta}|$, this means that also $(\tilde{x}, \tilde{\theta})$ is within some compact set $K \times K' \subset D \times \mathbb{R}^n$ independent of α, and applying Assumptions 2 and 3 we find

$$\begin{aligned} \tfrac{1}{2} m_K |\hat{\vartheta}|^2 &\leq \tfrac{1}{2}\langle \hat{\vartheta}, H_{\theta\theta}(\tilde{x}, \tilde{\theta})\hat{\vartheta}\rangle \\ &= -\tfrac{1}{2}\langle \varphi - \varphi_c, H_{xx}(\tilde{x}, \tilde{\theta})(\varphi - \varphi_c)\rangle - \langle \hat{\vartheta}, H_{\theta x}(\tilde{x}, \tilde{\theta})(\varphi - \varphi_c)\rangle \end{aligned}$$

$$
\begin{aligned}
&\leq C\big(|\varphi-\varphi_c|^2+|\varphi-\varphi_c|\,|\hat{\vartheta}|\big)\\
&\leq C\big(|\varphi-\varphi_c|^2+\tfrac{C}{m_K}|\varphi-\varphi_c|^2+\tfrac{m_K}{4C}|\hat{\vartheta}|^2\big)
\end{aligned}
$$

by Cauchy's inequality. Therefore we have $|\hat{\vartheta}| \leq C'|\varphi-\varphi_c|$ for some constant $C'>0$, and expanding around $(\varphi_c, 0)$ again we conclude that

$$
\begin{aligned}
\lambda &= \tfrac{1}{L_\varphi}|\lambda\varphi'| = \tfrac{1}{L_\varphi}|H_\theta(\varphi,\hat{\vartheta})|\\
&= \tfrac{1}{L_\varphi}\big|\underbrace{H_\theta(\varphi_c,0)}_{=0}+H_{\theta x}(\tilde{x}',\tilde{\theta}')(\varphi-\varphi_c)+H_{\theta\theta}(\tilde{x}',\tilde{\theta}')\hat{\vartheta}\big|\\
&\leq C''(|\varphi-\varphi_c|+|\hat{\vartheta}|)\\
&\leq C''(1+C')|\varphi-\varphi_c| \leq C''(1+C')L_\varphi|\alpha-\alpha_c|.
\end{aligned}
$$

(ii) This is now a direct consequence of (i) since $\lambda \leq C|\alpha-\alpha_c|$ implies for arbitrarily small $\epsilon>0$ that

$$
\int_{|\alpha-\alpha_c|\leq\epsilon}\frac{d\alpha}{\lambda} \geq \frac{1}{C}\int_{|\alpha-\alpha_c|\leq\epsilon}\frac{d\alpha}{|\alpha-\alpha_c|}=\infty.
$$

□

Appendix C

Proof of Lemma 2

Proof of Lemma 2. (i) In order to apply Arzèla-Ascoli's Theorem, we quickly check that the set $C_{X,M}$ is uniformly bounded and uniformly equicontinuous: For all $\varphi \in C_{X,M}$ and all $\alpha \in [0,1]$ we have

$$|\varphi(\alpha)| = \Big|\varphi(0) + \int_0^\alpha \varphi'(a)\,da\Big| \le |\varphi(0)| + \int_0^\alpha |\varphi'(a)|\,da \le \sup_{x\in X}|x| + M$$

and

$$|\varphi(\alpha+h) - \varphi(\alpha)| = \Big|\int_\alpha^{\alpha+h} \varphi'(a)\,da\Big| \le \int_\alpha^{\alpha+h} |\varphi'(a)|\,da \le Mh.$$

This proves precompactness of the set $C_{X,M}$. To prove that $C_{X,M}$ is also closed, take any sequence (φ_n) in $C_{X,M}$ that converges uniformly to some $\varphi \in C(0,1)$. We have to show that $\varphi \in C_{X,M}$. Clearly, $\varphi(0) \in X$. Furthermore, for every $a, b \in [0,1]$, $a < b$, we have

$$\sup_{a\le\alpha_0<\cdots<\alpha_N\le b} \sum_{i=1}^N |\varphi(\alpha_i) - \varphi(\alpha_{i-1})| = \sup_{a\le\alpha_0<\cdots<\alpha_N\le b} \lim_{n\to\infty} \sum_{i=1}^N |\varphi_n(\alpha_i) - \varphi_n(\alpha_{i-1})|$$

$$\leq \liminf_{n\to\infty} \sup_{a\leq\alpha_0<\cdots<\alpha_N\leq b} \sum_{i=1}^{N} |\varphi_n(\alpha_i) - \varphi_n(\alpha_{i-1})|$$
$$= \liminf_{n\to\infty} \int_a^b |\varphi_n'(\alpha)|\, d\alpha \leq M(b-a) < \infty.$$

This shows that φ is absolutely continuous and that $|\varphi'| \leq M$ almost everywhere. Therefore $C_{X,M}$ is closed, and since it is also precompact, it must be compact.

(ii) now follows directly from (i) by observing that $C_M^{x,y}$ is a closed subset of $C_{X,M}$, for $X := \{x\}$.

(iii) We want to show that for any sequence (φ_n) in $C_{X,M}$ converging to some $\varphi \in C_{X,M}$ we have $\hat{S}(\varphi) \leq \liminf_{n\to\infty} \hat{S}(\varphi_n)$. We can follow exactly the lines of the proof of [29, Lemma 5.42], applied to convergence in $|\cdot|_{[0,1]}$ (which in [29] is denoted by d_c), except for two modifications: First, since we do not have the equivalent of [29, Lemmas 5.17 and 5.18] (the integrand of $\hat{S}$ increases only linearly in $|\varphi'|$), we have to restrict $\hat{S}$ to $C_{X,M}$, which guarantees uniform equicontinuity of the sequence (φ_n), as shown in part (i). Second, we have to adjust the definition of the lower bound $\ell^\delta(x,y)$ for the local action $\ell(x,y)$ to our case and show that our function ℓ^δ still fulfills the required properties, i.e. weak convexity in y and lower semicontinuity in the limit $(x,y,\delta) \to (x_0,y_0,0+)$, although the technique in [29, Lemma 5.40] to prove the latter fails in our case (since our equivalent of $g^\delta(x,\theta)$ is not continuous).

For every $x \in D$, $y \in \mathbb{R}^n$ and $\delta > 0$ we define

$$\ell^\delta(x,y) := \sup\big\{\langle y,\theta\rangle \,|\, \theta \in \mathbb{R}^n \text{ s.t. } \forall z \in D:\ |z-x| \leq \delta \Rightarrow H(z,\theta) \leq 0\big\}.$$

Then for every $\bar{x} \in D$ with $|\bar{x} - x| \leq \delta$ we have

$$\ell^\delta(x, y) \leq \sup_{\substack{\theta \in \mathbb{R}^n \\ H(\bar{x},\theta) \leq 0}} \langle y, \theta \rangle = \sup_{\substack{\theta \in \mathbb{R}^n \\ H(\bar{x},\theta) = 0}} \langle y, \theta \rangle =: \ell(\bar{x}, y), \tag{C.1}$$

where ℓ is our local action from representation (2.4). Clearly, $\ell^\delta(x, y)$ is convex in y as the supremum of linear functions. To show lower semicontinuity of $\ell^\delta(x, y)$, consider first the cases when either x_0 is a critical point or when $y_0 = 0$. If x_0 is a critical point then the local action ℓ vanishes at (x_0, y_0) by Lemma 1; if $y_0 = 0$ then the local action ℓ vanishes by its definition in (C.1). Thus by Assumption 1 we have

$$\ell^\delta(x, y) \geq \langle y, 0 \rangle = 0 = \ell(x_0, y_0) \qquad \forall x, y, \delta,$$

so that

$$\liminf_{(x,y,\delta) \to (x_0, y_0, 0+)} \ell^\delta(x, y) \geq \ell(x_0, y_0). \tag{C.2}$$

In all other cases (i.e. x_0 is not a critical point and $y_0 \neq 0$) we have $H_\theta(x_0, \hat{\vartheta}) = \lambda(x_0, y_0) y_0 \neq 0$ (for $\hat{\vartheta} := \hat{\vartheta}(x_0, y_0)$) by Lemma 1 (i). Since by definition of $\hat{\vartheta}$ we have $H(x_0, \hat{\vartheta}) = 0$, for every $\varepsilon > 0$ there exists a $\tilde{\theta} \in \mathbb{R}^n$ with $|\tilde{\theta} - \hat{\vartheta}| < \varepsilon$ such that $H(x_0, \tilde{\theta}) < 0$. By continuity of $H(\cdot, \tilde{\theta})$, there exists an $\eta > 0$ such that for all $z \in D$ with $|z - x_0| \leq \eta$ we have $H(z, \tilde{\theta}) \leq 0$. Let $\delta + |x - x_0| \leq \eta$. Since

$$|z - x| \leq \delta \;\Rightarrow\; |z - x_0| \leq \delta + |x - x_0| \leq \eta,$$

we then have

$$
\begin{aligned}
\ell^{\delta}(x,y) &\geq \sup\{\langle y,\theta\rangle \,|\, \theta\in\mathbb{R}^n \text{ s.t. } \forall z\in D:\ |z-x_0|\leq\eta \Rightarrow H(z,\theta)\leq 0\} \\
&\geq \langle y,\tilde{\theta}\rangle \\
&= \langle y_0,\hat{\vartheta}\rangle + \langle y-y_0,\hat{\vartheta}\rangle + \langle y,\tilde{\theta}-\hat{\vartheta}\rangle \\
&\geq \ell(x_0,y_0) - |y-y_0||\hat{\vartheta}| - |y|\varepsilon,
\end{aligned}
$$

where in the last step we used representation (2.5) of the local action $\ell(x_0,y_0)$. Taking the $\liminf$ as $(x,y,\delta)\to(x_0,y_0,0+)$ and then letting $\varepsilon\to 0$, we see that (C.2) holds also in this case, terminating the proof of part (iii).

(iv) Let K be a non-empty closed subset of $C_{X,M}$, and let (φ_n) be a sequence in K such that $\lim_{n\to\infty}\hat{S}(\varphi_n)=\inf_{\varphi\in K}\hat{S}(\varphi)$. Since $C_{X,M}$ is compact, K is compact as well, and thus there exists a subsequence (φ_{n_k}) which converges uniformly to some $\varphi^\star\in K$ as $k\to\infty$. Because of the lower-semicontinuity of $\hat{S}$ we have

$$
\hat{S}(\varphi^\star)\leq\liminf_{k\to\infty}\hat{S}(\varphi_{n_k})=\inf_{\varphi\in K}\hat{S}(\varphi)
$$

and thus $\hat{S}(\varphi^\star)=\inf_{\varphi\in K}\hat{S}(\varphi)$. $\square$

Appendix D

Proof of Proposition 2

In this Appendix we will prove the technical details of Steps 1-3 that we omitted in the proof of Proposition 2 in Section 2.2. Let us begin with

Step 1: There exists an $\tilde{\eta} > 0$ such that for small enough δ

$$\rho(X^\varepsilon|_{[0,T]}, \varphi^\star) \leq \tilde{\eta} \quad \text{and} \quad \tau_\delta(X^\varepsilon) \leq T \quad \Rightarrow \quad \rho(X^\varepsilon|_{[0,\tau_\delta(X^\varepsilon)]}, \varphi^\star) \leq \eta. \tag{D.1}$$

Proof. Without loss of generality we may assume that $|\varphi^{\star\prime}| \equiv cst$ a.e. on $[0,1]$. Let $\alpha_0 \in (0,1)$ be large enough so that $\int_{\alpha_0}^1 |\varphi^{\star\prime}|\, d\alpha \leq \frac{1}{2}\eta$, and let

$$\begin{aligned} 0 < \tilde{\eta} \;&:=\; \frac{1}{4} \inf_{0 \leq \alpha \leq \alpha_0} |\varphi^\star(\alpha) - x_2| \\ &\leq\; \frac{1}{4}|\varphi^\star(\alpha_0) - x_2| = \frac{1}{4}\left|\int_{\alpha_0}^1 \varphi^{\star\prime}\, d\alpha\right| \leq \frac{1}{4}\int_{\alpha_0}^1 |\varphi^{\star\prime}|\, d\alpha \leq \frac{\eta}{8}. \end{aligned} \tag{D.2}$$

Let $\delta \leq \tilde{\eta}$. From the definition of the Fréchet distance, there are two weakly increasing surjective continuous functions $t(s) : [0,1] \to [0,T]$ and $\alpha(s) : [0,1] \to$

$[0,1]$ such that

$$\left|X^\varepsilon(t(s)) - \varphi^\star(\alpha(s))\right|_{[0,1]} \le 2\tilde{\eta}. \tag{D.3}$$

Let $s_0 \in [0,1]$ be such that $t(s_0) = \tau_\delta := \tau_\delta(X^\varepsilon)$, and define $\tilde{t}(s) := t(s) \wedge \tau_\delta$. (We write $a \vee b$ and $a \wedge b$ to denote the maximum and the minimum of two numbers a and b, respectively.) Then

$$\begin{aligned}
\rho(X^\varepsilon|_{[0,\tau_\delta]}, \varphi^\star) &\le |X^\varepsilon(\tilde{t}(s)) - \varphi^\star(\alpha(s))|_{[0,1]} \\
&= |X^\varepsilon(\tilde{t}(s)) - \varphi^\star(\alpha(s))|_{[0,s_0]} \vee |X^\varepsilon(\tilde{t}(s)) - \varphi^\star(\alpha(s))|_{[s_0,1]} \\
&= |X^\varepsilon(t(s)) - \varphi^\star(\alpha(s))|_{[0,s_0]} \vee |X^\varepsilon(\tau_\delta) - \varphi^\star(\alpha(s))|_{[s_0,1]} \\
&\le 2\tilde{\eta} \vee \Big(|X^\varepsilon(\tau_\delta) - x_2| + |\varphi^\star(\alpha(s)) - x_2|_{[s_0,1]}\Big).
\end{aligned} \tag{D.4}$$

To estimate the second norm in the last expression, observe that (using $t(s_0) = \tau_\delta$ and (D.3)) we have

$$\begin{aligned}
|\varphi^\star(\alpha(s_0)) - x_2| &\le |X^\varepsilon(t(s_0)) - x_2| + |X^\varepsilon(t(s_0)) - \varphi^\star(\alpha(s_0))| \\
&\le \delta + 2\tilde{\eta} \le 3\tilde{\eta} < 4\tilde{\eta} = \inf_{0\le\alpha\le\alpha_0} |\varphi^\star(\alpha) - x_2|,
\end{aligned}$$

so that $\alpha(s_0) > \alpha_0$ necessarily. Therefore by monotonicity of $\alpha(\cdot)$, for all $s \in [s_0, 1]$ we have $\alpha(s) \ge \alpha(s_0) > \alpha_0$ and thus

$$|\varphi^\star(\alpha(s)) - x_2| = \Big|\int_{\alpha(s)}^1 \varphi^{\star\prime}\, d\alpha\Big| \le \int_{\alpha_0}^1 |\varphi^{\star\prime}|\, d\alpha \le \frac{\eta}{2}.$$

We can now continue our estimate (D.4) and conclude that

$$\rho(X^\varepsilon|_{[0,\tau_\delta]}, \varphi^\star) \leq 2\tilde{\eta} \vee (\delta + \tfrac{\eta}{2}) \leq 2\tilde{\eta} \vee (\tilde{\eta} + \tfrac{\eta}{2}) \leq \tfrac{\eta}{4} \vee (\tfrac{\eta}{8} + \tfrac{\eta}{2}) < \eta$$

by (D.2), proving Step 1. □

Step 2: $\lim_{(T,\delta)\to(T^\star,0+)} \rho(\psi_{\delta,T}, \varphi^\star) = 0$.

Proof. It suffices to show that any sequence $(\psi_k)_{k\in\mathbb{N}} := (\psi_{\delta_k,T_k})_{k\in\mathbb{N}}$, where $\delta_k \to 0+$ and $T_k \to T^\star$, has a subsequence $(\psi_{k_l})_{l\in\mathbb{N}}$ such that $\lim_{l\to\infty} \rho(\psi_{k_l}, \varphi^\star) = 0$.

To show this, let (ψ_k) be such a sequence, and let us denote by φ_k the reparametrization of ψ_k (i.e. $\gamma(\varphi_k) = \gamma(\psi_k)$) such that $|\varphi_k'(\alpha)| \equiv cst_k$ for almost every $\alpha \in [0,1]$. The curve $\gamma(\varphi_k)$ starts from x_1, and by the assumption of Proposition 2 it has bounded length in the limit,

$$M := \limsup_{k\to\infty} \int_0^1 |\varphi_k'|\, d\alpha \leq \limsup_{(T,\delta)\to(T^\star,0+)} \int_0^T |\dot{\psi}_{\delta,T}|\, dt < \infty,$$

so φ_k is in the compact set $C_{\{x_1\},2M}$ defined in Lemma 2 (i) if k is sufficiently large. Thus there exists a subsequence $(\varphi_{k_l})_{l\in\mathbb{N}}$ that converges uniformly to some limit $\tilde{\varphi}^\star \in C_{\{x_1\},2M}$. In particular, we have $\rho(\psi_{k_l}, \tilde{\varphi}^\star) = \rho(\varphi_{k_l}, \tilde{\varphi}^\star) \leq |\varphi_{k_l} - \tilde{\varphi}^\star|_{[0,1]} \to 0$ as $l \to \infty$.

In the remaining part of the proof we will show that $\gamma(\tilde{\varphi}^\star) = \gamma(\varphi^\star)$, so that $\rho(\psi_{k_l}, \varphi^\star)$
$= \rho(\psi_{k_l}, \tilde{\varphi}^\star) \to 0$ as $l \to \infty$. By lower semi-continuity of $\hat{S}$ we have

$$\hat{S}(\tilde{\varphi}^\star) \quad \leq \quad \liminf_{l\to\infty} \hat{S}(\varphi_{k_l}) = \liminf_{l\to\infty} \inf_{T>0} \inf_{\psi\in\bar{C}_{\varphi_{k_l}}(0,T)} S_T(\psi)$$

$$\leq \liminf_{l\to\infty} S_{T_{k_l}}(\psi_{k_l}) = \liminf_{l\to\infty} \inf_{\substack{\psi\in\bar{C}_{x_1}(0,T_{k_l})\\ \tau_{\delta_{k_l}}(\psi)\leq T_{k_l}}} S_{T_{k_l}}(\psi). \tag{D.5}$$

We want to show that the right-hand side is less than or equal to $\hat{S}(\varphi^\star)$. Consider first the case when $T_{k_l} < T^\star$ for all l. Let $(\tilde{T}_r, \tilde{\psi}_r)_{r\in(0,\infty)}$ be the approximating sequence defined the proof of Proposition 1 (i) (only here with $r \in (0,\infty)$), i.e. such that $\forall r > 0 : \gamma(\tilde{\psi}_r) = \gamma(\varphi^\star)$, $\lim_{r\to\infty} S_{\tilde{T}_r}(\tilde{\psi}_r) = \hat{S}(\varphi^\star)$ and $\lim_{r\to\infty} \tilde{T}_r = T^\star$. Using the notation $\lambda = \lambda(\varphi^\star, \varphi^{\star\prime})$ and letting $\lambda_r = \lambda \vee \frac{1}{r}$, we find that for $\forall r_1, r_2 > 0$:

$$\begin{aligned} |\tilde{T}_{r_1} - \tilde{T}_{r_2}| &\leq \int_0^1 \Big|\frac{1}{\lambda_{r_1}} - \frac{1}{\lambda_{r_2}}\Big| \, d\alpha \leq \Big|\frac{1}{\lambda_{r_1}} - \frac{1}{\lambda_{r_2}}\Big|_{[0,1]} \\ &= \big|(\tfrac{1}{\lambda} \wedge r_1) - (\tfrac{1}{\lambda} \wedge r_2)\big|_{[0,1]} \leq |r_1 - r_2|. \end{aligned}$$

As a result, the function $r \mapsto \tilde{T}_r$ is continuous, and we can choose a sequence (r_l) such that $\tilde{T}_{r_l} = T_{k_l}$ for $\forall l \in \mathbb{N}$. Now since $\tilde{\psi}_{r_l}(\tilde{T}_{r_l}) = \varphi^\star(1) = x_2 \in B_{\delta_k}(x_2)$, we can complete the estimate (D.5) as follows:

$$\begin{aligned} \hat{S}(\tilde{\varphi}^\star) &\leq \liminf_{l\to\infty} \inf_{\substack{\psi\in\bar{C}_{x_1}(0,T_{k_l})\\ \tau_{\delta_{k_l}}(\psi)\leq T_{k_l}}} S_{T_{k_l}}(\psi) \leq \liminf_{l\to\infty} S_{T_{k_l}}(\tilde{\psi}_{r_l}) \\ &= \liminf_{l\to\infty} S_{\tilde{T}_{r_l}}(\tilde{\psi}_{r_l}) = \hat{S}(\varphi^\star) = \inf_{\varphi\in\bar{C}_{x_1}^{x_2}(0,1)} \hat{S}(\varphi). \end{aligned} \tag{D.6}$$

If $T_{k_l} \geq T^\star$ for some l, then we can define the path $\tilde{\psi}_{r_l}$ by first following the path $\psi^\star = \tilde{\psi}_{r=\infty}$ in time $T^\star$ (which in this case is well-defined since $T^\star < \infty$) and then staying at x_2 for the remaining time $\Delta_l := T_{k_l} - T^\star$. As $l \to \infty$, we have $\Delta_l \to 0$, and thus the additional action on the second part of the path goes to 0 as well,

so that $S_{\tilde{T}_{r_l}}(\tilde{\psi}_{r_l}) \to \hat{S}(\varphi^\star)$ still, and (D.6) remains valid also in this case.

Since by assumption of Proposition 2 we have

$$\tilde{\varphi}^\star(1) = \lim_{l\to\infty} \varphi_{k_l}(1) = \lim_{l\to\infty} \psi_{\delta_{k_l},T_{k_l}}(T_{k_l}) = x_2,$$

$\tilde{\varphi}^\star$ is in $\bar{C}^{x_2}_{x_1}(0,1)$, and so from (D.6) we can conclude that

$$\hat{S}(\tilde{\varphi}^\star) = \inf_{\varphi\in\bar{C}^{x_2}_{x_1}(0,1)} \hat{S}(\varphi) = \hat{S}(\varphi^\star).$$

The uniqueness of the minimizing curve $\varphi^\star$ now implies that $\gamma(\tilde{\varphi}^\star) = \gamma(\varphi^\star)$, terminating the proof. □

Step 3: For all $T, \delta > 0$ the set of paths $\{\psi \in \bar{C}_{x_1}(0,T) \,|\, \tau_\delta(\psi) \le T\}$ is regular with respect to S_T.

Proof. First, note that the set $\{\tau_\delta \le T\} := \{\psi \in \bar{C}_{x_1}(0,T) \,|\, \tau_\delta(\psi) \le T\}$ is closed since its complement $\{\tau_\delta \le T\}^c = \{\tau_\delta > T\} = \{\psi \,|\, \gamma(\psi) \cap B_\delta(x_2) = \emptyset\}$ is open. Since closing the latter set amounts to replacing the closed ball $B_\delta(x_2)$ by the corresponding open ball, we find that the interior of the set $\{\tau_\delta \le T\}$ is $\{\tau_\delta \le T\}^0 = \{\tau^0_\delta \le T\}$, where τ^0_δ denotes the infimum of all times at which the path is inside the *open* ball with radius δ around x_2. Thus we must show

$$\inf_{\substack{\psi\in\bar{C}_{x_1}(0,T)\\ \tau_\delta(\psi)\le T}} S_T(\psi) \ge \inf_{\substack{\psi\in\bar{C}_{x_1}(0,T)\\ \tau^0_\delta(\psi)\le T}} S_T(\psi),$$

since the relation $\le$ is clear. We will show that for every $\psi \in \{\tau_\delta \le T\}$ we can construct functions $\tilde{\psi}_r \in \{\tau^0_\delta \le T\}$, $r > 0$, such that $|S_T(\tilde{\psi}_r) - S_T(\psi)|$ becomes

arbitrarily small as $r \to 0+$. The function $\tilde{\psi}_r$ will be constructed in such a way that it traverses $\gamma(\psi)$ at a slightly higher speed than ψ, and we will use the time we saved to make a small excursion from the point where ψ touches the outside of the ball $B_\delta(x_2)$ into its interior and back, so that in fact $\tau_\delta^0(\tilde{\psi}) \leq T$. In order to show that the action $S_T(\tilde{\psi}_r)$ differs only slightly from $S_T(\psi)$, it turns out that one can speed up the path ψ only at places where $|\dot{\psi}|$ is bounded away from ∞. To do so, we pick some $M > \operatorname{essinf}_{0\leq t\leq T} |\dot{\psi}(t)|$, for every $r \in (0, \frac{1}{2})$ define the time-rescaling G_r via its inverse by

$$G_r^{-1}(t) := \int_0^t (1 - r\mathbf{1}_{|\dot{\psi}|\leq M})\, d\tau,\ G_r^{-1} : [0,T] \to [0,T_r],\ T_r := G_r^{-1}(T) < T,$$

where $\mathbf{1}_{|\dot{\psi}|\leq M}$ denotes the indicator function on the set $\{t \in [0,T];\ |\dot{\psi}(t)| \leq M\}$, and set $\psi_r(s) := \psi(G_r(s))$. Using

$$G_r'(s) = 1/(G_r^{-1})'(G_r(s)) = \left[1 - r\mathbf{1}_{|\dot{\psi}|\leq M}(G_r(s))\right]^{-1},$$

we find that

$$\begin{aligned} S_{T_r}(\psi_r) &= \int_0^{T_r} L(\psi_r(s), \dot{\psi}_r(s))\, ds \\ &= \int_0^{T_r} L\Big(\psi(G_r(s)), \dot{\psi}(G_r(s))/\left[1 - r\mathbf{1}_{|\dot{\psi}|\leq M}(G_r(s))\right]\Big)\, ds \\ &= \int_0^{T} L\Big(\psi(t), \dot{\psi}(t)/\left[1 - r\mathbf{1}_{|\dot{\psi}|\leq M}(t)\right]\Big)\left[1 - r\mathbf{1}_{|\dot{\psi}|\leq M}(t)\right] dt. \end{aligned}$$

As $r \to 0+$, the integrand in the last integral converges pointwise to $L(\psi(t), \dot{\psi}(t))$. To show that one can exchange limit and integral, observe that on $\{t \in [0,T] :$

$|\dot{\psi}(t)| > M\}$ the integrand equals $L(\psi, \dot{\psi})$ before taking the limit, and that on $\{t \in [0,T] : |\dot{\psi}(t)| \leq M\}$ we can use the bounded convergence theorem since (i) $\frac{1}{2} \leq (1 - r\mathbf{1}_{|\dot{\psi}|\leq M}) \leq 1$, (ii) $\psi(t)$ traverses the compact set $\gamma(\psi)$, and (iii) $L(x,y)$ is continuous. Thus $\lim_{r\to 0+} S_{T_r}(\psi_r) = S_T(\psi)$.

Now let $\nu_r := \frac{1}{2}(T - T_r) = \frac{r}{2}\int_0^T \mathbf{1}_{|\dot{\psi}|\leq M}\, dt > 0$, and pick a point $x_\delta \in \gamma(\psi) \cap \partial B_\delta(x_2)$ at which ψ touches the boundary of $B_\delta(x_2)$. Consider

$$\chi^+_{\nu_r}(t) := x_\delta + t(x_2 - x_\delta) \qquad \text{and} \qquad \chi^-_{\nu_r}(t) := x_\delta + (\nu_r - t)(x_2 - x_\delta)$$

(for $0 \leq t \leq \nu_r$): $\chi^+_{\nu_r}$ starts at x_δ and enters the ball $B_\delta(x_2)$ in the direction of its center x_2, $\chi^-_{\nu_r}$ then goes back the opposite way. The corresponding actions

$$\begin{aligned} S_{\nu_r}(\chi^+_{\nu_r}) &= \int_0^{\nu_r} L\big(x_\delta + t(x_2 - x_\delta), x_2 - x_\delta\big)\, dt \qquad \text{and} \\ S_{\nu_r}(\chi^-_{\nu_r}) &= \int_0^{\nu_r} L\big(x_\delta + (\nu_r - t)(x_2 - x_\delta), x_\delta - x_2\big)\, dt \\ &= \int_0^{\nu_r} L\big(x_\delta + s(x_2 - x_\delta), x_\delta - x_2\big)\, ds \end{aligned}$$

converge to zero as $r \to 0+$ (and thus $\nu_r \to 0+$). We can now define $\tilde{\psi}_r$ by piecing ψ_r, $\chi^+_{\nu_r}$ and $\chi^-_{\nu_r}$ together in such a way that $\tilde{\psi}_r$ moves from x_1 to x_δ along ψ_r, briefly enters and exits the interior of $B_\delta(x_2)$ via $\chi^+_{\nu_r}$ and $\chi^-_{\nu_r}$, and continues along the remaining part of ψ_r. The total time for this path is $T_r + 2\nu_r = T_r + (T - T_r) = T$, and the total action is

$$S_T(\tilde{\psi}_r) = S_{T_r}(\psi_r) + S_{\nu_r}(\chi^+_{\nu_r}) + S_{\nu_r}(\chi^-_{\nu_r}) \to S_T(\psi) + 0 + 0$$

as $r \to 0$. Since $\tau^0_\delta(\tilde{\psi}_r) \leq T$ for every $r \in (0, \frac{1}{2})$, this completes the proof. $\square$

Appendix E

Proof of Proposition 4

In the following Lemma we will compute the derivatives of $\hat{\vartheta}(x, y)$ and $\lambda(x, y)$ which we will need later in the proof of Proposition 4 to compute the Euler-Lagrange equation for $\hat{S}$.

Lemma 13 (Derivatives of $\hat{\vartheta}$ and λ)**.** *For all $x \in D$ and $y \in \mathbb{R}^n \setminus \{0\}$ we have*

$$\hat{\vartheta}_x(x, y) = -H_{\theta\theta}^{-1}\left(P_y H_{\theta x} + \lambda^{-1}\frac{y \otimes H_x}{\langle y, H_{\theta\theta}^{-1} y\rangle}\right), \tag{E.1}$$

$$\hat{\vartheta}_y(x, y) = \lambda H_{\theta\theta}^{-1} P_y, \tag{E.2}$$

$$\hat{\vartheta}_y(x, y)^T y = 0, \tag{E.3}$$

$$\lambda_x(x, y) = \frac{\lambda H_{x\theta} H_{\theta\theta}^{-1} y - H_x}{\lambda\langle y, H_{\theta\theta}^{-1} y\rangle}, \tag{E.4}$$

$$\lambda_y(x, y) = -\frac{\lambda H_{\theta\theta}^{-1} y}{\langle y, H_{\theta\theta}^{-1} y\rangle}, \tag{E.5}$$

where we abbreviate

$$P_y := I - \frac{y \otimes H_{\theta\theta}^{-1} y}{\langle y, H_{\theta\theta}^{-1} y\rangle}, \tag{E.6}$$

and where H_x, $H_{\theta x}$, $H_{x\theta}$ and $H_{\theta\theta}$ are evaluated at the point $(x, \hat{\vartheta}(x, y))$.

Proof. All formulae can be obtained by implicit differentiation of the Equations (2.7), where $\lambda = \lambda(x, y)$ and $\hat{\vartheta} = \hat{\vartheta}(x, y)$.

First we differentiate $H(x, \hat{\vartheta}(x, y)) = 0$ both with respect to x and y to obtain

$$H_x^T + H_\theta^T \hat{\vartheta}_x = H_x^T + \lambda y^T \hat{\vartheta}_x = 0, \qquad H_\theta^T \hat{\vartheta}_y = \lambda y^T \hat{\vartheta}_y = 0. \tag{E.7}$$

From the second equation we see that (E.3) holds since $\lambda = 0$ only if x is a critical point and since $\hat{\vartheta}_y$ is continuous.

Differentiating the second equation in (2.7), $H_\theta(x, \hat{\vartheta}(x, y)) = \lambda y$, with respect to x and y, we obtain

$$H_{\theta x} + H_{\theta\theta} \hat{\vartheta}_x = y \lambda_x^T, \qquad H_{\theta\theta} \hat{\vartheta}_y = \lambda I + y \lambda_y^T. \tag{E.8}$$

Left-multiplying both equations by $\lambda y^T H_{\theta\theta}^{-1}$ and using Equations (E.7), we conclude

$$\begin{aligned} \lambda y^T H_{\theta\theta}^{-1} H_{\theta x} - H_x^T &= \lambda \langle y, H_{\theta\theta}^{-1} y \rangle \lambda_x^T, \\ 0 &= \lambda^2 y^T H_{\theta\theta}^{-1} + \lambda \langle y, H_{\theta\theta}^{-1} y \rangle \lambda_y^T, \end{aligned}$$

which we can solve for λ_x^T and λ_y^T:

$$\lambda_x^T = \frac{\lambda y^T H_{\theta\theta}^{-1} H_{\theta x} - H_x^T}{\lambda \langle y, H_{\theta\theta}^{-1} y \rangle}, \qquad \lambda_y^T = -\frac{\lambda y^T H_{\theta\theta}^{-1}}{\langle y, H_{\theta\theta}^{-1} y \rangle}, \tag{E.9}$$

proving (E.4) and (E.5). We can now solve Equations (E.8) for $\hat{\vartheta}_x$ and $\hat{\vartheta}_y$ and

plug in Equations (E.9) to obtain

$$
\begin{aligned}
\hat{\vartheta}_x &= H_{\theta\theta}^{-1}(y\lambda_x^T - H_{\theta x}) \\
&= H_{\theta\theta}^{-1}\Big(\frac{\lambda y y^T H_{\theta\theta}^{-1} H_{\theta x} - y H_x^T}{\lambda\langle y, H_{\theta\theta}^{-1} y\rangle} - H_{\theta x}\Big) \\
&= -H_{\theta\theta}^{-1}\Big(P_y H_{\theta x} + \lambda^{-1}\frac{y H_x^T}{\langle y, H_{\theta\theta}^{-1} y\rangle}\Big), \\
\hat{\vartheta}_y &= H_{\theta\theta}^{-1}(\lambda I + y\lambda_y^T) \\
&= H_{\theta\theta}^{-1}\lambda\Big(I - \frac{y y^T H_{\theta\theta}^{-1}}{\langle y, H_{\theta\theta}^{-1} y\rangle}\Big) = \lambda H_{\theta\theta}^{-1} P_y,
\end{aligned}
$$

where

$$
P_y = I - \frac{y y^T H_{\theta\theta}^{-1}}{\langle y, H_{\theta\theta}^{-1} y\rangle} = I - \frac{y \otimes H_{\theta\theta}^{-1} y}{\langle y, H_{\theta\theta}^{-1} y\rangle}.
$$

This proves (E.1) and (E.2) and we are done. □

Proof of Proposition 4. Starting from the representation (2.5) of the action $\hat{S}$, we obtain

$$
\begin{aligned}
D\hat{S}(\varphi) &= \hat{\vartheta}_x^T \varphi' - \partial_\alpha(\hat{\vartheta} + \hat{\vartheta}_y^T \varphi') \\
&= \hat{\vartheta}_x^T \varphi' - \hat{\vartheta}_x \varphi' - \hat{\vartheta}_y \varphi'' - \partial_\alpha(\hat{\vartheta}_y^T \varphi') \\
&= (\hat{\vartheta}_x^T - \hat{\vartheta}_x)\varphi' - \hat{\vartheta}_y \varphi'' - \partial_\alpha(\hat{\vartheta}_y^T \varphi'). \qquad \text{(E.10)}
\end{aligned}
$$

(E.3) in Lemma 13 says that the last term in (E.10) vanishes. We can then apply the formulae (E.1) and (E.2) for the derivatives of $\hat{\vartheta}$ to obtain

$$
\lambda H_{\theta\theta} D\hat{S}(\varphi) = \lambda H_{\theta\theta}\big((\hat{\vartheta}_x^T - \hat{\vartheta}_x)\varphi' - \hat{\vartheta}_y \varphi''\big)
$$

$$
\begin{aligned}
&= \Big(P_{\varphi'}\lambda H_{\theta x} + \frac{\varphi' H_x^T}{\langle \varphi', H_{\theta\theta}^{-1}\varphi'\rangle} - \lambda H_{\theta\theta} H_{\theta x}^T P_{\varphi'}^T H_{\theta\theta}^{-1} \\
&\qquad\qquad - \frac{H_{\theta\theta} H_x \varphi'^T H_{\theta\theta}^{-1}}{\langle \varphi', H_{\theta\theta}^{-1}\varphi'\rangle}\Big)\varphi' - \lambda^2 P_{\varphi'}\varphi'' \\
&= \Big(P_{\varphi'}\lambda H_{\theta x} + \frac{\varphi' H_x^T}{\langle \varphi', H_{\theta\theta}^{-1}\varphi'\rangle}\Big)\varphi' - 0 - H_{\theta\theta} H_x - \lambda^2 P_{\varphi'}\varphi'' \\
&= P_{\varphi'}\big(-\lambda^2\varphi'' + \lambda H_{\theta x}\varphi' - H_{\theta\theta} H_x\big).
\end{aligned}
$$

The relation $\lambda = \langle H_\theta, \varphi'\rangle / |\varphi'|^2$ follows directly from (2.7). To show the second representation of this term, we use (E.4) and (E.5) to compute:

$$
\begin{aligned}
\lambda\lambda'\varphi' &= \lambda\big(\partial_\alpha \lambda(\varphi, \varphi')\big)\varphi' \\
&= \big(\langle \lambda\lambda_x, \varphi'\rangle + \langle \lambda\lambda_y, \varphi''\rangle\big)\varphi' \\
&= \frac{\langle \lambda H_{x\theta} H_{\theta\theta}^{-1}\varphi' - H_x, \varphi'\rangle - \langle \lambda^2 H_{\theta\theta}^{-1}\varphi', \varphi''\rangle}{\langle \varphi', H_{\theta\theta}^{-1}\varphi'\rangle}\,\varphi' \\
&= \frac{\langle \lambda H_{\theta x}\varphi' - H_{\theta\theta} H_x - \lambda^2\varphi'', H_{\theta\theta}^{-1}\varphi'\rangle}{\langle \varphi', H_{\theta\theta}^{-1}\varphi'\rangle}\,\varphi' \\
&= \frac{\varphi' \otimes H_{\theta\theta}^{-1}\varphi'}{\langle \varphi', H_{\theta\theta}^{-1}\varphi'\rangle}\big(-\lambda^2\varphi'' + \lambda H_{\theta x}\varphi' - H_{\theta\theta} H_x\big) \\
&= \big(I - P_{\varphi'}\big)\big(-\lambda^2\varphi'' + \lambda H_{\theta x}\varphi' - H_{\theta\theta} H_x\big).
\end{aligned}
$$

□

Appendix F

The update formula for the inner loop

Computing Step 2: Given a vector $\hat{\vartheta}^p$ and

$$h = h(\hat{\vartheta}^p), \qquad h_\theta = h_\theta(\hat{\vartheta}^p) \qquad \text{and} \qquad h_{\theta\theta} = h_{\theta\theta}(\hat{\vartheta}^p),$$

we must find θ_0, A and c such that the quadratic function

$$f(\theta) := \tfrac{1}{2}\langle \theta - \theta_0, A(\theta - \theta_0)\rangle + c$$

fulfills

$$f(\hat{\vartheta}^p) = h, \qquad f_\theta(\hat{\vartheta}^p) = h_\theta \qquad \text{and} \qquad f_{\theta\theta}(\hat{\vartheta}^p) = h_{\theta\theta}. \tag{F.1}$$

Clearly, the last equation in (F.1) implies $A = h_{\theta\theta}$. From the second equation in (F.1) we obtain

$$A(\hat{\vartheta}^p - \theta_0) = h_\theta \quad \Leftrightarrow \quad \hat{\vartheta}^p - \theta_0 = A^{-1}h_\theta = h_{\theta\theta}^{-1}h_\theta.$$

Finally, the first equation in (F.1) tells us that

$$\begin{aligned} h &= \tfrac{1}{2}\langle \hat{\vartheta}^p - \theta_0, A(\hat{\vartheta}^p - \theta_0)\rangle + c = \tfrac{1}{2}\langle A^{-1}h_\theta, AA^{-1}h_\theta\rangle + c \\ &= \tfrac{1}{2}\langle h_\theta, A^{-1}h_\theta\rangle + c \\ \Leftrightarrow \qquad c &= h - \tfrac{1}{2}\langle h_\theta, A^{-1}h_\theta\rangle = h - \tfrac{1}{2}\langle h_\theta, h_{\theta\theta}^{-1}h_\theta\rangle. \end{aligned}$$

Summarizing, f is given by

$$f(\theta) = \tfrac{1}{2}\langle \theta - \theta_0, h_{\theta\theta}(\theta - \theta_0)\rangle + [h - \tfrac{1}{2}\langle h_\theta, h_{\theta\theta}^{-1}h_\theta\rangle],$$

where $\theta_0 = \hat{\vartheta}^p - h_{\theta\theta}^{-1}h_\theta$. Thus, if $f(\theta_0) = h - \frac{1}{2}\langle h_\theta, h_{\theta\theta}^{-1}h_\theta\rangle \geq 0$ then we return $\hat{\vartheta}^{p+1} = \theta_0$.

Computing Step 3: Suppose now that $f(\theta_0) < 0$ (i.e. the region $\{f < 0\}$ is non-empty), and let some direction φ' be given. We must find the point $\hat{\vartheta}^{p+1}$ such that

$$f(\hat{\vartheta}^{p+1}) = 0 \qquad \text{and} \qquad f_\theta(\hat{\vartheta}^{p+1}) = \lambda\varphi' \quad \text{for some} \quad \lambda \geq 0. \tag{F.2}$$

The second equation in (F.2) is equivalent to

$$h_{\theta\theta}(\hat{\vartheta}^{p+1} - \theta_0) = \lambda\varphi' \qquad \Leftrightarrow \qquad \hat{\vartheta}^{p+1} = \theta_0 + \lambda h_{\theta\theta}^{-1}\varphi'.$$

To obtain λ, we then use the first equation in (F.2):

$$
\begin{aligned}
0 &= f(\hat{\vartheta}^{p+1}) \\
&= \tfrac{1}{2}\langle \hat{\vartheta}^{p+1} - \theta_0, h_{\theta\theta}(\hat{\vartheta}^{p+1} - \theta_0)\rangle + [h - \tfrac{1}{2}\langle h_\theta, h_{\theta\theta}^{-1} h_\theta\rangle] \\
&= \tfrac{1}{2}\lambda^2 \langle h_{\theta\theta}^{-1}\varphi', h_{\theta\theta} h_{\theta\theta}^{-1}\varphi'\rangle + [h - \tfrac{1}{2}\langle h_\theta, h_{\theta\theta}^{-1} h_\theta\rangle] \\
\Rightarrow \qquad \lambda &= +\left(\frac{\langle h_\theta, h_{\theta\theta}^{-1} h_\theta\rangle - 2h}{\langle \varphi', h_{\theta\theta}^{-1}\varphi'\rangle}\right)^{1/2}
\end{aligned}
$$

since we are interested in the non-negative solution λ. The point we are looking for is thus

$$
\begin{aligned}
\hat{\vartheta}^{p+1} &= \theta_0 + \left(\frac{\langle h_\theta, h_{\theta\theta}^{-1} h_\theta\rangle - 2h}{\langle \varphi', h_{\theta\theta}^{-1}\varphi'\rangle}\right)_+^{1/2} h_{\theta\theta}^{-1}\varphi' \\
&= \hat{\vartheta}^p + h_{\theta\theta}^{-1}\big(\tilde{\lambda}(\hat{\vartheta}^p)\varphi' - h_\theta\big) \quad \text{with} \quad \tilde{\lambda}(\hat{\vartheta}^p) := \left(\frac{\langle h_\theta, h_{\theta\theta}^{-1} h_\theta\rangle - 2h}{\langle \varphi', h_{\theta\theta}^{-1}\varphi'\rangle}\right)_+^{1/2}.
\end{aligned}
$$

Bibliography

[1] Allen, R. J., Warren, P. B., ten Wolde, P. R., *Sampling rare switching events in biochemical networks*, Physical Review Letters **94**, 018104 (2005)

[2] Anderson, M., Kimn, J.-H., Basket implied volatility from geodesics, http://relativity.phys.lsu.edu/postdocs/matt/publications.php (2007)

[3] Baker, D., Church, G., Collins, J., Endy, D., Jacobsen, J., Keasling, J., Modrich, P., Smolke, C., Weiss, R., *Engineering life: Building a FAB for biology*, Scientific American, June 2006, 34–41

[4] Codon Devices website, `http://www.codondevices.com/news.aspx?id=170`

[5] Da Prato, G., Zabczyk, J., *Stochastic differential equations in infinite dimensions*, Cambridge University Press, 1992

[6] Durrett, R., *Stochastic calculus - a practical introduction*, CRC Press, 1996

[7] E, W., Ren, W., Vanden-Eijnden, E., *String method for the study of rare events*, Physical Review B **66**, 052301 (2002)

[8] E, W., Ren, W., Vanden-Eijnden, E., *A minimum action method for the study of rare events*, Comm. Pure App. Math. **52**, 637–656 (2004)

[9] Elowitz, M. B., Leibler, S., *A synthetic oscillatory network of transcriptional regulators*, Nature **403**, 335–338 (2000)

[10] Endy, D., *Foundations for engineering biology*, Nature **438**, 04342 (2005)

[11] Freidlin, M. I., *Random perturbations of reaction-diffusion equations: the quasideterministic approximation*, Trans. Amer. Math. Soc. **305**.2, 665–697 (1988)

[12] Freidlin, M. I., Wentzell, A. D., *Random perturbations of dynamical systems*, 2nd Edition, Springer, 1998

[13] Gardner, T. S., Cantor, C. R., Collins, J. J., *Construction of a genetic toggle switch in Escherichia coli*, Nature **403**, 339–342 (2000)

[14] Gillespie, D. T., *A general method for numerically simulating the stochastic time evolution of coupled chemical reactions*, Journal of Computational Physics **22**, 403–434 (1976)

[15] Hagan, P. S., Kumar, D., Lesniewski, A. S., Woodward, D. E., *Managing smile risk*, Wilmott magazine, September 2002, 84–108

[16] Heymann, M., Vanden-Eijnden, E., *The geometric minimum action method: a least action principle on the space of curves*, to appear in Comm. Pure App. Math. (2007)

[17] IGEM website of the MIT team 2006, `http://openwetware.org/wiki/IGEM:MIT/ 2006/Blurb`

[18] Jónsson, H., Mills, G., Jacobsen, K. W., *Nudged elastic band method for finding minimum energy paths of transitions*, in: *Classical and quantum dynamics in condensed phase simulations*, ed. Berne, B. J., Ciccotti, G., Coker, D. F., World Scientific, 1998

[19] Keller, H. B., *Numerical methods for two-point boundary value problems*, Dover, 1992

[20] Kleinert, H., *Path integrals in quantum mechanics, statistics, polymer physics, and financial markets*, 3rd Edition, World Scientific, 2005

[21] Levskaya, A., Chevalier, A. A.; Tabor, J. J., Simpson, Z. B., Lavery, L. A., Levy M., Davidson, E. A., Scouras, A., Ellington, A. D., Marcotte, E. M., Voigt, C. A., *Engineering Escherichia coli to see light*, Nature **438**, 441–442 (2005)

[22] Maier, R. S., Stein, D. L., *A scaling theory of bifurcations in the symmetric weak-noise escape problem*, Journal of Statistical Physics **83**.3/4, 291–357 (1996)

[23] Morton, K. W., Mayers, D. F., *Numerical solution of partial differential equations - an introduction*, 2nd Edition, Cambridge University Press, 2005

[24] Nocedal, J., Wright, S. J., *Numerical optimization*, Springer Series in Operations Research, 1999

[25] Olender, R., Elber, R., *Calculation of classical trajectories with a very large time step: Formalism and numerical examples*, Journal of Chemical Physics **105**, 9299 (1996)

[26] Registry of Standard Biological Parts, `http://parts.mit.edu`, part `BBa_E0040`

[27] Registry of Standard Biological Parts, `http://parts.mit.edu`, devices `BBa_J45100` and `BBa_J45200`

[28] Roma, D. M., O'Flanagan, A., Ruckenstein, A. E., Sengupta, A. M., Mukhopadhyay, R., *Optimal path to epigenetic switching*, Physical Review E **71**, 011902 (2005)

[29] Shwartz, A., Weiss, A., *Large deviations for performance analysis - queues, communication, and computing*, Chapman & Hall (1995)

[30] Simon, B., *Semiclassical analysis of low lying eigenvalues, II. Tunneling*, The Annals of Mathematics, 2nd Series, **120**.1, 89–118 (1984)

[31] Varadhan, S. R. S., *Large deviations and applications*, SIAM, 1984

[32] Warren, P. B., ten Wolde, P. R., *Chemical models of genetic toggle switches*, Journal of Physical Chemistry B **109**, 6812–6823 (2005)

[33] Warren, P. B., ten Wolde, P. R., *Enhancement of the stability of genetic switches by overlapping upstream regulatory domains*, Physical Review Letters **92**.12, 128101 (2004)

[34] Weiss, J. N., *The Hill equation revisited: uses and misuses*, FASEB **11**, 835–841 (1997)