Minimum Action Method for the Study of Rare Events

WEINAN E
Princeton University
Institute for Advanced Study
WEIQING REN
Institute for Advanced Study
AND
ERIC VANDEN-EIJNDEN
Courant Institute
Institute for Advanced Study

Abstract

The least-action principle from the Wentzell-Freidlin theory of large deviations is exploited as a numerical tool for finding the optimal dynamical paths in spatially extended systems driven by a small noise. The action is discretized and a preconditioned BFGS method is used to optimize the discrete action. Applications are presented for thermally activated reversal in the Ginzburg-Landau model in one and two dimensions, and for noise-induced excursion events in the Brusselator taken as an example of a nongradient system arising in chemistry. In the Ginzburg-Landau model, the reversal proceeds via interesting nucleation events, followed by propagation of domain walls. The issue of nucleation versus propagation is discussed, and the scaling for the number of nucleation events as a function of the reversal time and other material parameters is computed. Good agreement is found with the numerical results. In the Brusselator, whose deterministic dynamics has a single stable equilibrium state, the presence of noise is shown to induce large excursions by which the system cycles out of this equilibrium state.

1 Introduction

Consider a stochastic ODE or PDE perturbed by a small noise,

$$\dot{x} = b(x) + \sqrt{\varepsilon}\,\dot{W}\,, \quad u_t = L(u) + \sqrt{\varepsilon}\,\eta(x,t)\,.$$

Here $\varepsilon \ll 1$, and $\dot{W}$ and η are, respectively, temporal and spatiotemporal white noises. In the absence of noise, the system evolves to its equilibrium states and stays there indefinitely. The presence of noise changes that picture over long time scales. The system may hop between metastable states, make excursions out of these states, etc.

At first sight one might think that the influence of noise happens at such a long time scale that it is rarely of practical importance. That this view is incorrect may

Communications on Pure and Applied Mathematics, Vol. LVII, 0637–0656 (2004)

be understood by noticing that most physical processes will not happen at zero temperature when thermal noise is absent. In fact, nature presents naturally a very wide range of temporal scales. The atomistic vibration that gives rise to thermal noise happens on the time scale of femtoseconds (10^{-15} seconds), and the processes of interest to our daily lives are mostly on the order of seconds or longer, leaving plenty of room for rare events caused by thermal noise to make their appearance. Chemical reactions, nucleations, and conformational changes of biomolecules are all examples of such rare events.

Traditionally the method of choice for a quantitative understanding of the effect of noise has been the Monte Carlo method or direct simulation of the Langevin equation. When the noise is small, which is the case of interest here, these methods become prohibitively expensive, due to the presence of two disparate time scales: the time scale of the deterministic dynamics and the time scale between the rare events caused by the noise.

Noticing this difficulty, alternative theories and numerical methods have been proposed. The most notable analytical work is the theory of Wentzell and Freidlin [11], which gives an estimate on the probability of the paths in terms of an action functional. The most probable path is given by the one that minimizes the action functional. The Wentzell-Freidlin action is an analogue of the Onsager-Machlup action [7, 19], and it can be used as a numerical tool in which optimal trajectories between the initial and final states in the system are computed by optimizing the action functional. We shall call methods of this type *minimum action methods*, and the optimal path the *minimal action path* (MAP).

In this paper we develop a minimum action method for spatially extended systems described by PDEs. Although the idea of using Wentzell-Freidlin action or a variation thereof as numerical tools is not new in the context of ODEs (see, e.g., [1, 19]), the main challenge in spatially extended systems is that the action functional and its derivative typically involve high-order spatial derivatives, giving rise to highly ill-conditioned numerical problems. Therefore a good choice of preconditioner is required in order to achieve an acceptable convergence rate.

It should be pointed out that if the system happens to be a gradient system, then over an infinite time interval the minimal action path becomes a minimal energy path (MEP), which is a heteroclinic orbit that connects two local minima of the potential. Efficient numerical methods have been developed for finding the MEP as well as the transition rates. Notable examples include the nudged elastic band method (NEB) [13] and the string method [3, 4, 5, 21]. These methods are typically more efficient than methods based on minimizing the action functional. Yet the minimum action method has the advantage of being applicable to both gradient and nongradient systems. It can also be applied to studying events on finite-time intervals that have exponentially small probability in small domains but become ubiquitous when the size of the system is large.

This paper is organized as follows. In Section 2 we briefly review the Wentzell-Freidlin theory. In Section 3 we present the numerical method for the example

of the one-dimensional Ginzburg-Landau model. Results on the two-dimensional Ginzburg-Landau model are presented in Section 4. Nongradient systems are discussed in Section 5 with the example of the Brusselator. Some conclusions are drawn in Section 6. The appendix contains a discussion on the limited-memory BFGS method, which is used in the minimum action method.

2 Wentzell-Freidlin Theory and the Least-Action Principle

We consider a random process $X^\varepsilon(t)$ in $\mathbb{R}^n$ defined by the stochastic differential equation (SDE)

$$\dot{X}^\varepsilon = b(X^\varepsilon) + \sqrt{\varepsilon}\,\sigma\,\dot{W}\,, \quad X^\varepsilon(0) = x_0\,. \tag{2.1}$$

Here $W(t)$ is a Wiener process in $\mathbb{R}^n$. As $\varepsilon \to 0$, the trajectory $X^\varepsilon(t)$ converges in probability to the solution $x(t)$ of the unperturbed equation

$$\dot{x} = b(x)\,, \quad x(0) = x_0\,, \tag{2.2}$$

on every finite time interval. In the path space, the probability distribution of $X^\varepsilon(t)$ is concentrated in a neighborhood of $x(t)$. Any other event that does not include $x(t)$ and its neighborhood has very small probability.

The Wentzell-Freidlin theory gives an estimate of the probability distribution of the random process in the path space. We state the estimate by first introducing an action functional in the path space. We denote by $\mathbf{C}_{[0,T]}$ the set of continuous functions on the interval $[0, T]$ with values in $\mathbb{R}^n$. In this space, we define the metric

$$\rho_T(\varphi, \psi) = \sup_{0\le t\le T} |\varphi(t) - \psi(t)|\,.$$

For $\varphi \in \mathbf{C}_{[0,T]}$, we define an action functional associated with (2.1) as

$$S_T[\varphi] = \frac{1}{2}\int_0^T \left|\sigma^{-1}(\dot{\varphi} - b(\varphi))\right|^2 dt \tag{2.3}$$

if the integral is finite. Otherwise, we set $S_T[\varphi]$ to be $+\infty$.

The Wentzell-Freidlin estimates are:

(1) For any $\delta > 0$, $\gamma > 0$, and $K > 0$ there exists an $\varepsilon_0 > 0$ such that

$$\mathsf{P}\{\rho_T(X^\varepsilon, \varphi) < \delta\} \ge \exp\left\{-\frac{1}{\varepsilon}(S_T[\varphi] + \gamma)\right\} \quad \text{for } \varepsilon < \varepsilon_0, \tag{2.4}$$

where $T > 0$ and $\varphi \in \mathbf{C}_{[0,T]}$ are such that $\varphi(0) = x_0$ and $T + S_T[\varphi] \le K$.

(2) For any $\delta > 0$, $\gamma > 0$, and $s_0 > 0$ there exists an $\varepsilon_0 > 0$ such that for $0 < \varepsilon \le \varepsilon_0$ and $s < s_0$ we have

$$\mathsf{P}\{\rho_T(X^\varepsilon, \Phi_T(s)) \ge \delta\} \le \exp\left\{-\frac{1}{\varepsilon}(s - \gamma)\right\} \tag{2.5}$$

where

$$\Phi_T(s) = \{\varphi \in \mathbf{C}_{[0,T]}, \varphi(0) = x_0, S_T[\varphi] \le s\} \quad \text{for } s > 0$$

and

$$\rho_T\big(X^\varepsilon, \Phi_T(s)\big) = \sup_{\varphi\in\Phi_T(s)} \rho_T(X^\varepsilon, \varphi)\,.$$

For details of the proofs of these estimates, we refer to [11]. Roughly speaking, the estimates (2.4) and (2.5) tell us that the probability that X^ε stays in a δ-neighborhood of a path φ is

$$\mathsf{P}\big\{\rho_T(X^\varepsilon, \varphi) < \delta\big\} \approx \exp\left\{-\frac{1}{\varepsilon} S_T[\varphi]\right\}. \tag{2.6}$$

The above estimates can be used to calculate the probability of various events associated with (2.1) by constrained minimization of the action functional. For instance, let a and b be two states of the system and fix a time T. Then the probability P_T that the system moves from a to a δ-neighborhood of b within time T can be estimated using the Wentzell-Freidlin theory:

$$\lim_{\varepsilon\to 0} \varepsilon \ln P_T = -\min_{\varphi} S_T[\varphi]\,, \tag{2.7}$$

where the minimization in (2.7) is constrained by

$$\varphi(0) = a\,, \quad \varphi(T) = b\,. \tag{2.8}$$

The minimizer of (2.7), or the MAP $\varphi^\star$, gives the most probable path for the transition from a to b in the sense that the probability that the system moves by another path is exponentially smaller in ε.

So far we have only discussed finite-dimensional systems. In the following, we will be interested in the application of Wentzell-Freidlin theory to infinite dimensional systems described by stochastic PDEs. Wentzell-Freidlin theory can be extended to PDEs (for some results, see, e.g., [2, 6, 9, 10]), with an action functional similar to (2.3). Yet this extension leads to several subtleties. One major difficulty is associated with the fact that the noise we are most interested in, the spatiotemporal white noise, is too singular for the stochastic PDE to be well-defined in dimensions higher than 1 [12]. In this case, the noise must be regularized in space, for instance, by introducing a small but finite correlation length λ. It is still an open question as to what the physical effect of such regularization is. Discussing this issue is beyond the scope of the present paper. Nevertheless, we note that the action associated with the regularized problem formally has a limit when $\lambda \to 0$. This limiting action is the one we will always use in the applications below.

3 Minimum Action Method: Application to the One-Dimensional Ginzburg-Landau Model

We will discuss the minimum action method by way of an example: the thermally activated switching of a bistable system modeled by the Ginzburg-Landau equation

$$u_t = \delta u_{xx} - \delta^{-1} V'(u)\,, \quad x \in [0, 1]\,, \tag{3.1}$$

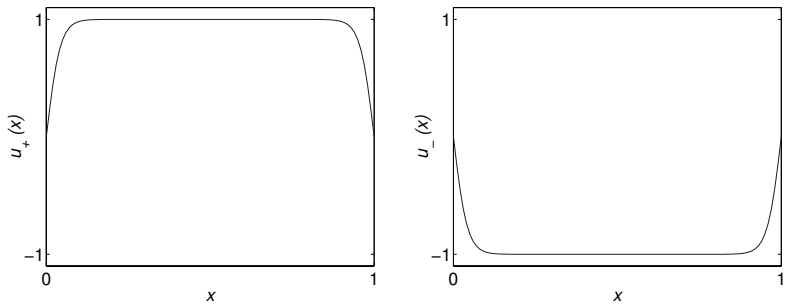

FIGURE 3.1. Two equilibrium states of (3.1) for $\delta = 0.03$.

with Dirichlet boundary conditions

$$u(0,t) = 0\,, \quad u(1,t) = 0\,. \tag{3.2}$$

We take $V(u)$ to be the standard double-well potential

$$V(u) = \frac{1}{4}(1-u^2)^2\,. \tag{3.3}$$

In equation (3.1) δ is a small parameter, which indicates that the dynamics corresponding to the reaction term, $\delta^{-1}V'(u)$, is fast while the diffusion is slow. The system can be considered as the gradient flow associated with the energy

$$E[u] = \frac{1}{2}\int_0^1 \big(\delta u_x^2 + 2\delta^{-1}V(u)\big)dx\,. \tag{3.4}$$

The dynamics in (3.1) has two stable equilibrium states, $u_+(x)$ and $u_-(x)$, which minimize the energy (3.4). When δ is small, $u_\pm(x) = \pm 1$ except in two thin layers of width δ at $x = 0$ and $x = 1$, as shown in Figure 3.1.

Now we add to (3.1) a small noise term modeling thermal effects:

$$u_t = \delta u_{xx} - \delta^{-1}V'(u) + \sqrt{\varepsilon}\,\eta\,, \tag{3.5}$$

where ε is proportional to the temperature of the system and η is a spatiotemporal white noise with covariance

$$\mathsf{E}\,(\eta(x,t)\eta(y,s)) = \delta(x-y)\delta(t-s)\,. \tag{3.6}$$

It is shown in [6] (see also [15]) that (3.5) makes sense at least in dimension 1.

The presence of the noise in (3.5) destroys the long-time stability of the equilibria $u_\pm$. For instance, if the initial state is u_+, there is a finite probability P_T that the system switches to u_- in any time interval $[0, T]$. From large-deviation theory [6], the probability P_T satisfies

$$\lim_{\varepsilon\to 0} \varepsilon \ln P_T = -\min_u S_T[u]\,, \tag{3.7}$$

where the action functional is given by

$$S_T[u] = \frac{1}{2}\int_0^T \int_0^1 \left(u_t - \delta u_{xx} + \delta^{-1} V'(u)\right)^2 dx\, dt\,. \tag{3.8}$$

The minimization problem in (3.7) is constrained by the conditions

$$u(0,t) = u(1,t) = 0\,, \quad u(x,0) = u_+(x)\,, \quad u(x,T) = u_-(x)\,. \tag{3.9}$$

The minimizer of (3.8) constrained by (3.9) defines the optimal switching path between u_+ and u_- during the time interval $[0, T]$.

3.1 Minimization of the Action Functional

To find the optimal switching path, we minimize the action (3.8) in a simple two-step numerical procedure. In the first step we discretize the action functional using finite differences. In the second step we minimize the discretized action functional using the limited memory BFGS (L-BFGS) method (see the appendix). Other methods can also be used for either step; for example, finite elements can be used in the first step. We made these choices for their simplicity.

We discretize the space-time domain $[0, 1] \times [0, T]$ with a mesh with sizes $\Delta x = 1/I$ and $\Delta t = t/J$, and we define the grid point (x_i, t_j) by

$$\begin{aligned} x_i &= i\,\Delta x\,, \quad i = 0, 1, \ldots, I,\\ t_j &= j\,\Delta t\,, \quad j = 0, 1, \ldots, J. \end{aligned}$$

We also define

$$x_{i+1/2} = \left(i + \frac{1}{2}\right)\Delta x\,, \quad t_{j+1/2} = \left(j + \frac{1}{2}\right)\Delta t\,.$$

The numerical approximation to $u(x_i, t_j)$ is denoted by $U_{i,j}$. In order to simplify the expression, we introduce the force

$$p(x,t) = u_t - \delta u_{xx} + \delta^{-1} V'(u)\,, \tag{3.10}$$

so the action can be written as

$$S_T[u] = \frac{1}{2}\int_0^T \int_0^1 p^2(x,t) dx\, dt\,. \tag{3.11}$$

We use the trapezoidal rule to discretize the spatial integral and the midpoint rule to compute the temporal integral, thus obtaining

$$S_T(U) = \frac{1}{2}\Delta x \Delta t \sum_{i=1}^{I-1}\sum_{j=0}^{J-1} P_{i,j+1/2}^2\,, \tag{3.12}$$

where $P_{i,j+1/2}$ is the numerical approximation to $p(x_i, t_{j+1/2})$ by central finite difference:

$$P_{i,j+1/2} = \frac{U_{i,j+1} - U_{i,j}}{\Delta t} + \delta^{-1} V'\left(\frac{U_{i,j+1} + U_{i,j}}{2}\right) - \frac{\delta}{2}\left(\frac{U_{i+1,j+1} - 2U_{i,j+1} + U_{i-1,j+1}}{\Delta x^2} + \frac{U_{i+1,j} - 2U_{i,j} + U_{i-1,j}}{\Delta x^2}\right). \tag{3.13}$$

The discretized boundary conditions corresponding to (3.9) are

$$\begin{aligned} &U_{0,j} = U_{I,j} = 0\,, && j = 0, 1, \ldots, J, \\ &U_{i,0} = u_+(x_i)\,, \quad U_{i,J} = u_-(x_i)\,, && i = 0, 1, \ldots, I. \end{aligned} \tag{3.14}$$

The BFGS method requires the gradient of (3.12); this is given by

$$\begin{aligned} \frac{\partial S_T}{\partial U_{i,j}} = &\left(2\Delta x + \frac{\Delta x \Delta t}{\delta} V''\left(\frac{U_{i,j-1} + U_{i,j}}{2}\right) + \frac{2\delta \Delta t}{\Delta x}\right) P_{i,j-1/2} \\ &- \left(2\Delta x - \frac{\Delta x \Delta t}{\delta} V''\left(\frac{U_{i,j} + U_{i,j+1}}{2}\right) - \frac{2\delta \Delta t}{\Delta x}\right) P_{i,j+1/2} \\ &- \frac{\delta \Delta t}{\Delta x}\left(P_{i-1,j-1/2} + P_{i-1,j+1/2} + P_{i+1,j-1/2} + P_{i+1,j+1/2}\right) \end{aligned}$$

for $i = 1, 2, \ldots, I-1$, $j = 1, 2, \ldots, J-1$.

In the BFGS method, the initial approximation to the inverse of the Hessian plays the role of preconditioner. A good choice of the preconditioner is vital for the efficiency of the algorithm. A general choice is to use the diagonal matrix

$$H_k^0 = \gamma_k I \tag{3.15}$$

where γ_k is defined in the appendix. For our problem, we found that the linear part of the Hessian of the action (3.8), which contains the highest-order derivatives, is a better preconditioner:

$$H_k^0 = B^{-1} = \left(-\partial_t^2 + \delta^2 \partial_x^4\right)^{-1}. \tag{3.16}$$

The operator B can be inverted efficiently by the fast Fourier transform (FFT). In the following, we will compare the behavior for different choices of H_k^0.

3.2 Numerical Results

In Table 3.1 we illustrate the behavior of the L-BFGS method for different choices of H_k^0 and the memory parameter m. In this problem, the parameter δ is 0.05, and T is 1, and we used 100×100 points in the discretization. For a complex system, the evaluation of the objective function and its gradient dominates the cost of the computation. Therefore, in the table we present the number of function and gradient evaluations n_{fg} required for the L_2-norm of the gradient to reach a certain tolerance. We also present the number of iterations, denoted by n_{it}, in parentheses. Note that n_{fg} is always bigger than n_{it}, since at each iteration step, several function

	$H_k^0 = I$ $n_{\mathrm{fg}}(n_{\mathrm{it}})$	$H_k^0 = \gamma_k I$ $n_{\mathrm{fg}}(n_{\mathrm{it}})$	$H_k^0 = B^{-1}$ $n_{\mathrm{fg}}(n_{\mathrm{it}})$
$m = 4$	16634 (8298)	18697 (17838)	744 (709)
$m = 5$	17492 (8742)	13130 (12714)	658 (638)
$m = 6$	18219 (9092)	10070 (9771)	573 (547)
$m = 7$	16902 (8421)	9745 (9500)	570 (553)
$m = 8$	17928 (8935)	10200 (9938)	561 (544)
$m = 9$	17400 (8656)	9446 (9208)	564 (550)
$m = 10$	18624 (9265)	8765 (8547)	540 (520)

TABLE 3.1. Performance of L-BFGS for various choices of the storage parameter m and initial approximations to the Hessian. n_{fg} is the total number of function and gradient evaluations, and n_{it} is the total number of iterations required to decrease the L_2-norm of the gradient to 10^{-10}.

and gradient evaluations might be needed in the line search if the search direction p_k is not well scaled (see the appendix for the definition of p_k).

The first conclusion that can be drawn from this table is that the number of function and gradient evaluations is insensitive to the memory parameter m. Since the cost of each iteration increases with the amount of storage, the method is most efficient for moderate m, for example, $m = 6$ or 7. Second, the table shows that for fixed m, the Hessian approximation by (3.16) improves the convergence significantly, which indicates that (3.16) is a good preconditioner for this problem. Furthermore, for $H_k^0 = \gamma_k I$ and $H_k^0 = B^{-1}$, the two values n_{fg} and n_{it} are of the same order, which indicates that the search direction p_k is well scaled, and, as a result, the initial step length $\alpha_k = 1$ is accepted in most iterations. However, for $H_k^0 = I$, the number n_{fg} is about twice the size of n_{it}, which means that p_k is not well scaled, and the line search is required to obtain a suitable step length.

Figure 3.2 shows the sequence of profiles of the optimal path u at different times in $[0, T]$ for various values of T at a fixed $\delta = 0.03$. The switching proceeds by nucleation followed by propagation of domain walls. For large T, the switching proceeds by propagating one domain wall from one boundary to another, as shown in Figure 3.2(a). As T becomes smaller, however, the number of nucleation events increases. In Figure 3.3, we display the values of the action against T for the various local minimizers shown in Figure 3.2. Figure 3.4 shows the space-time plot of the square of the force $p(x, t)$ corresponding to the switching event in Figure 3.2(f). This can be interpreted as the minimal noise necessary to induce the switching. The peaks correspond to the nucleations. These results are explained next.

3.3 Nucleation Versus Propagation in One Dimension

The results in the last section can be understood as follows. (Our discussion here is rather qualitative; more quantitative results will be presented in [22]. For a

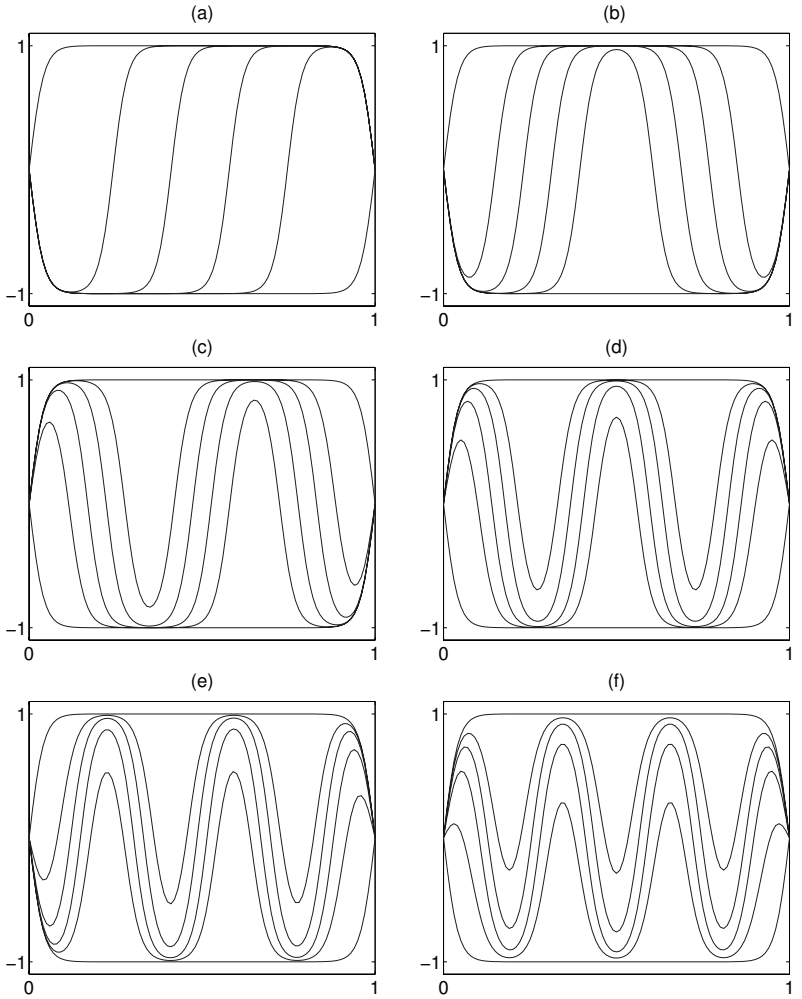

FIGURE 3.2. Snapshots of profiles of the minimizer u during a switching from u_+ (top curve) to u_- (bottom curve) at equally spaced times in $[0, T]$ for different T at fixed $\delta = 0.03$. (a) $T = 7$; (b) $T = 2$; (c) $T = 1$; (d) $T = 0.8$; (e) $T = 0.6$; (f) $T = 0.4$.

complementary discussion of nucleation versus propagation in one dimension, see [8].) Consider the critical points of the energy (3.4), i.e., the solutions of

$$\delta u_{xx} - \delta^{-1} V'(u) = 0 \tag{3.17}$$

with $u|_{x=0} = u|_{x=1} = 0$. Besides u_+ and u_-, corresponding to the minimizers of the energy, there are also saddle point configurations with an increasing number of domain walls. Our result shows that, for large T, the switching path crosses the saddle point configuration with minimum energy, i.e., the configuration with a single domain wall shown in Figure 3.2(a) (this result is standard; see [6] and also

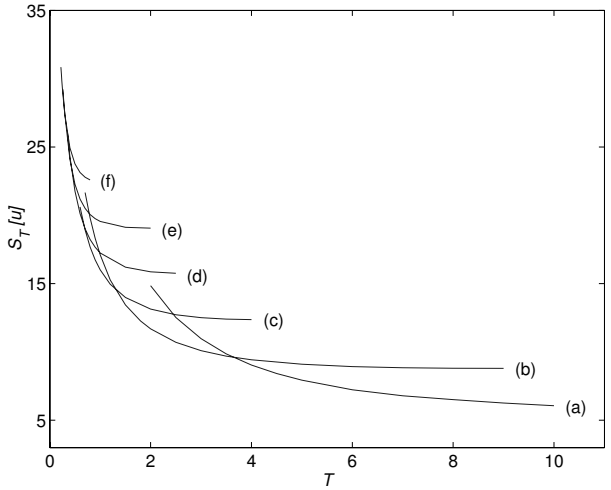

FIGURE 3.3. The action as a function of the switching time T for the six minimizers shown in Figure 3.2.

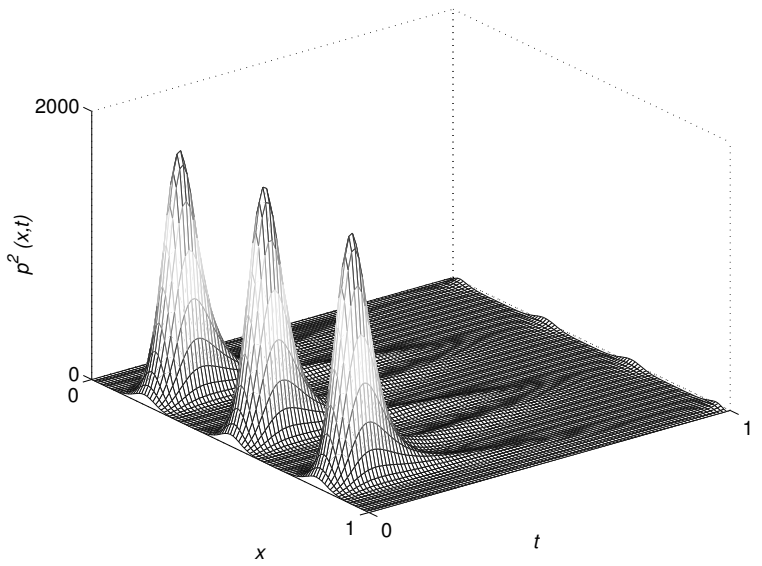

FIGURE 3.4. The space-time value of $p^2(x, t)$ for the minimizer (f) in Figure 3.2.

[14]). As T is decreased, the optimal switching path crosses (the vicinity of) saddle point configurations with increasing energy; i.e., it involves more nucleations and therefore more domain walls, giving rise to a cascade of nucleation events. The reason is that both the nucleation and domain wall motion are noise induced. As T

decreases, at a fixed number of nucleations the speed of propagation of the domain wall must increase in order to complete the switching during the allowed time T. This is energy consuming and, at certain critical values of T, it becomes more favorable to make an additional nucleation. The same type of cascade of nucleations is also observed if T is fixed but δ is decreased.

To quantify these observations, we use test functions to find an upper bound for the action. The numerical results indicate that switching occurs via domain wall propagation at constant speed c. Therefore, to describe the motion of a domain wall localized at $x = x^\star + ct$, we take

$$u(x,t) = \tanh\left(\frac{\eta}{\sqrt{2}}\right) \quad \text{with } \eta = \frac{x - x^\star - ct}{\delta}\,. \tag{3.18}$$

This hyperbolic tangent profile fits very well the minimizer obtained numerically, which suggests that it will lead to a rather sharp bound on the action. (3.18) is a solution of (3.17) with the boundary conditions $\lim_{x\to\pm\infty} u = \pm 1$. (3.18) is valid locally in the strip $|x - x^\star - ct| \ll \lambda$ with $\delta \ll \lambda \ll 1$ and must be modified to describe the other domain walls outside of this strip. Inserting (3.18) into (3.8), one obtains

$$S_T\left[\tanh\left(\frac{\eta}{\sqrt{2}}\right)\theta(x)\right] \approx \frac{c^2}{4\delta}\int_0^T \int_{\mathbb{R}} \operatorname{sech}^4\left(\frac{\eta}{\sqrt{2}}\right) d\eta\, dt = \frac{\sqrt{2}}{3} T c^2 \delta^{-1}.$$

Since this is the cost per domain wall, and their velocity is $1/(2nT)$ if there are $2n$ of them (for simplicity, we consider only the situation where the number of domain walls is even; i.e., there is no domain wall starting from the boundary of the domain—it gives the right scaling for large n anyway), the total cost of propagation is

$$A^{\text{prop}} = \frac{C_2}{Tn\delta} \quad \text{with } C_2 = \frac{\sqrt{2}}{6}\,. \tag{3.19}$$

This estimate indicates that propagation alone does not account for all the action cost of switching since otherwise one could make A^{prop} as small as one wishes by increasing n (and hence decreasing c). The additional cost comes from the nucleation events. To describe these we look for the minimizer of (3.8) localized in a region of size $O(\delta)$ both in space and time,

$$u(x,t) = v\left(\frac{x - x^\star}{\delta}, \frac{t}{\delta}\right),$$

for some $v(\cdot)$ to be determined later and obtain

$$S_T[v] \approx \frac{1}{2}\int_0^\infty \int_{\mathbb{R}} (v_\tau - v_{\xi\xi} + V'(v))^2\, d\xi\, d\tau\,. \tag{3.20}$$

The minimum cost of a nucleation event is given by the minimum of the above action constrained by

$$v(\xi, 0) = -1\,, \quad v(\xi, \infty) = +1 \quad \text{if } \xi \in \left[-\frac{1}{2}\lambda\delta^{-1}, \frac{1}{2}\lambda\delta^{-1}\right],$$

where, as before, λ satisfies $\delta \ll \lambda \ll 1$. The minimum is achieved by following backward in time the orbit associated with

$$v_\tau = v_{\xi\xi} - V'(v)\,,$$

which connects the state where $v = +1$ if $\xi \in [-\frac{1}{2}\lambda\delta^{-1}, \frac{1}{2}\lambda\delta^{-1}]$ to $v = -1$. Indeed, the action (3.20) can be written as

$$S_T[v] = \frac{1}{2}\int_0^\infty \int_{\mathbb{R}} (v_\tau + v_{\xi\xi} - V'(v))^2\, d\xi\, d\tau + 2\int_0^\infty \int_{\mathbb{R}} v_\tau (v_{\xi\xi} - V'(v)) d\xi\, d\tau$$

along the orbit; the first term vanishes, as can be seen by reversing time in the integration, whereas the second term reduces to twice the difference of energy between the state where $v = +1$ if $\xi \in [-\frac{1}{2}\lambda\delta^{-1}, \frac{1}{2}\lambda\delta^{-1}]$ and the minimum energy state $v = -1$ (which has zero energy). In the limit as $\lambda\delta^{-1} \to \infty$, the energy of the state where $v = +1$ if $\xi \in [-\frac{1}{2}\lambda\delta^{-1}, \frac{1}{2}\lambda\delta^{-1}]$ and $v = -1$ if $\xi \to \pm\infty$ can be estimated as twice the energy of the tangent hyperbolic profile already encountered before. We denote twice this energy by C_1 and therefore obtain the following estimate for the cost of n nucleations corresponding to $2n$ domain walls (assuming again that there is no nucleation at the boundary of the domain):

$$A^{\text{nucl}} = C_1 n\,. \tag{3.21}$$

Using (3.19) and (3.21), we obtain the total cost of the nucleation and propagation of $2n$ domain walls:

$$A(n, \delta, T) = C_1 n + \frac{C_2}{n\delta T}\,. \tag{3.22}$$

For fixed δ and T, the optimal number of nucleations is given by (ignoring the integer constraint)

$$n^\star = \arg\min_n A(n, \delta, T) = \left(\frac{C_2}{C_1\delta T}\right)^{\frac{1}{2}}, \tag{3.23}$$

and from (3.7) one has

$$\lim_{\varepsilon\to 0} \varepsilon \ln P_T = -A(n_\star, \delta, T) = -2\left(\frac{C_1 C_2}{\delta T}\right)^{\frac{1}{2}}, \tag{3.24}$$

which gives the envelope of the curves in Figure 3.3.

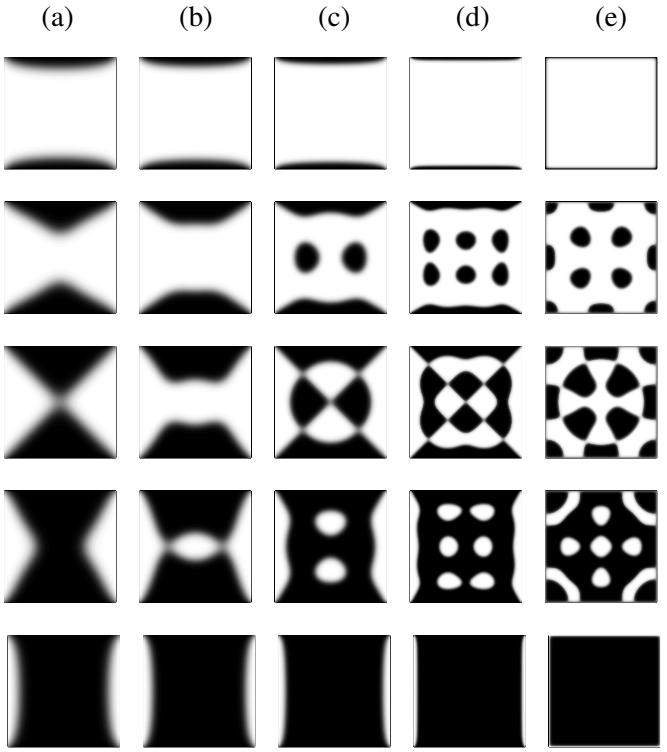

FIGURE 4.1. Snapshots of profiles of the minimizer u during a switching from u_+ (top figure in each column) to u_- (bottom figure in each column) at five equally spaced times in $[0, T]$ for $T = 1$ and different δ and different boundary conditions: (a)–(d) correspond to (4.2), (e) corresponds to (4.3). In (a) $\delta = 0.04$; (b) $\delta = 0.03$; (c) $\delta = 0.02$; (d) $\delta = 0.01$; (e) $\delta = 0.01$. The gray scale is from white for $u = 1$ to black for $u = -1$.

4 Application to the Two-Dimensional Ginzburg-Landau Model

4.1 Numerical Results

As an example in two dimensions, we consider

$$u_t = \delta \Delta u - \delta^{-1} V'(u) \tag{4.1}$$

in the unit square $\Omega = [0, 1] \times [0, 1]$. We will consider two different Dirichlet boundary conditions:

$$u|_{x=0} = u|_{x=1} = 1\,, \quad u|_{y=0} = u|_{y=1} = -1\,, \tag{4.2}$$

and

$$u|_{x=0} = u|_{x=1} = u|_{y=0} = u|_{y=1} = 0\,. \tag{4.3}$$

In both cases, (4.1) has two stable equilibrium states. One of them, u_+, which is the top figure in each column in Figure 4.1, is close to $u = 1$ except at the boundary layers at $y = 0$ and $y = 1$ in the case (4.2), or at the edge of the square in the case (4.3). Another stable state, u_-, which is the bottom figure in each column in Figure 4.1, is close to $u = -1$ except at the boundary layers at $x = 0$ and $x = 1$ in the case (4.2), or at the edge of the square in the case (4.3). Similar to the one-dimensional problem, assuming the system switches from u_+ to u_- before a given time T, we minimize the action

$$S_T[u] = \frac{1}{2}\int_0^T \int_\Omega \left(u_t - \delta\Delta u + \delta^{-1}V'(u)\right)^2 dx\, dy\, dt \tag{4.4}$$

to find the optimal switching path between the two states. The numerical algorithm we use is a direct extension of the method for the one-dimensional model.

4.2 Nucleation Versus Propagation in Higher Dimensions

In Figure 4.1, we show the time sequences of the switching process for different values of δ at fixed $T = 1$. The paths (a), (b), (c), and (d) correspond to the first case with boundary condition (4.2), and (e) corresponds to (4.3). The overall trend is consistent with what was found in the one-dimensional example; namely, there are more and more nucleation events as $\delta \to 0$.

As in the one-dimensional case, we can estimate the optimal number of nucleations for fixed T and δ in two dimensions by using test functions. As suggested by our numerical results, we will assume that the switching occurs via the dynamics of domain walls or, more appropriately, sharp interface motion. Consistent with this scenario, suppose the location of the interface at time t is specified by the curve $\gamma(t)$ (or the hypersurface in higher dimension), and assume the minimizer of the action can be approximated by the test function

$$u(x, y, t) = \tanh\left(\frac{\rho((x, y), \gamma(t))}{\sqrt{2}\,\delta}\right). \tag{4.5}$$

Here $\rho((x, y), \gamma(t))$ denotes the distance of the point (x, y) from the interface $\gamma(t)$. (4.5) is valid if $\rho((x, y), \gamma(t)) \le \lambda$ with $\delta \ll \lambda \ll 1$. Inserting (4.5) into (4.4) and using the properties that ρ_t is the normal velocity of the interface u_n, $-\Delta\rho$ its mean curvature κ, and $|\nabla\rho| = 1$ by definition, one obtains to leading order in δ

$$\begin{aligned} S_T\left[\tanh\left(\frac{\rho}{\sqrt{2}\,\delta}\right)\right] &\approx \frac{1}{4\delta}\int_0^T \int_{\gamma(t)} (u_n + \delta\kappa)^2\, ds\, dt \int_{\mathbb{R}} \operatorname{sech}^4\left(\frac{\rho}{\sqrt{2}}\right) d\rho \\ &= \frac{\sqrt{2}}{3\delta}\int_0^T \int_{\gamma(t)} (u_n + \delta\kappa)^2\, d\sigma\, dt\,. \end{aligned} \tag{4.6}$$

This limiting action is valid in dimension 2 or higher. Interestingly, to leading order in δ, (4.6) is complete, and there is no need to include an additional term to account

for nucleation events provided that their codimension is $d_c = d$ (i.e., we exclude scenarios with line nucleation in $d = 2$, line or surface nucleation in $d = 3$, etc.). This is to be contrasted with the situation in one dimension. To understand why this is the case, notice first that the curvature term in (4.6) guarantees that the number of nucleations in the minimizers of this action must be finite since a large number of nucleations leads to an interface with very high curvature and hence, from (4.6), very high cost.

To verify that the additional cost of nucleations is higher order in δ, let us estimate their cost in d dimensions with a test function of the form

$$u(\mathbf{x}, t) = v\left(\frac{\mathbf{x} - \mathbf{x}^\star}{\delta}, \frac{t}{\delta}\right),$$

where $\mathbf{x}, \mathbf{x}^\star \in \mathbb{R}^d$, and $\mathbf{x}^\star$ is the nucleation point. Inserting this function in (4.4) gives

$$S_T[v] = \frac{1}{2}\delta^{d-1} \int_0^\infty \int_{\mathbb{R}^d} (v_\tau - \Delta_\xi v + V'(v))^2 \, d^d\xi \, d\tau = O(\delta^{d-1}). \tag{4.7}$$

Next we estimate the minimum of the action in (4.6) and show that it is lower order in δ (i.e., the cost of nucleations accounted by (4.7) is indeed negligible).

To estimate the order of magnitude of the action, we look for an interface composed of n spheres with radius $R(t)$. In this case, $u_n = \dot{R}$, $\kappa = R^{-1}$. We impose that $R(T/2) = \frac{1}{2}n^{-1/d}$; i.e., the n circular interfaces at $t = T/2$ just touch each other as they should. We also assume that the motion is consistent with the scaling of flow by curvature; i.e., we take

$$R(t) = \frac{1}{\sqrt{2}} n^{-\frac{1}{d}} \left(\frac{t}{T}\right)^{\frac{1}{2}}. \tag{4.8}$$

This last assumption can in fact be removed and the scaling of (4.8) can be deduced by minimization of (4.6) with circular interface satisfying $R(T/2) = \frac{1}{2}n^{-1/d}$. Here we use (4.8) directly for simplicity of presentation. Inserting (4.8) in (4.6) gives

$$\begin{aligned} S_T(n) &= \frac{\sqrt{2}}{3} \delta^{-1} n S_d \int_0^T \left(\dot{R} + \frac{\delta}{R}\right)^2 R^{d-1} \, dt \\ &= \frac{2^{d/2}}{3} S_d \delta^{-1} n^{\frac{1}{d}} \left(n^{-\frac{1}{d}} T^{-\frac{1}{2}} + \delta n^{\frac{1}{d}} T^{\frac{1}{2}}\right)^2, \end{aligned}$$

where S_d is the surface of the unit sphere in d dimensions. Here n is the only parameter that remains to be minimized over. Since

$$S_T\left(\nu(\delta T)^{-\frac{d}{2}}\right) = \frac{2^{d/2}}{3} S_d (T\delta)^{-\frac{1}{2}} \nu^{\frac{1}{d}} \left(\nu^{-\frac{1}{d}} + \nu^{\frac{1}{d}}\right)^2,$$

it follows that

$$\min_n S_T(n) = C_3 (T\delta)^{-\frac{1}{2}}, \tag{4.9}$$

where

$$C_3 = \frac{2^{d/2}}{3} S_d \nu_\star \quad \text{with } \nu_\star = \arg\min_{\nu} \nu^{\frac{1}{d}} \left(\nu^{-\frac{1}{d}} + \nu^{\frac{1}{d}}\right)^2 .$$

Thus the minimal cost of (4.6) is bounded above by $O(\delta^{-1/2})$. Furthermore, the number of nucleations scales as

$$n_\star = (\delta T)^{-\frac{d}{2}} \nu_\star = O\left((\delta T)^{-\frac{d}{2}}\right). \tag{4.10}$$

This makes the additional contribution from (4.7) negligible as asserted since $n_\star = O(\delta^{-d/2})$ nucleations have a cost of the order of $O(\delta^{-d/2}\delta^{d-1}) = O(\delta^{d/2-1})$ only.

5 Nongradient Systems

The methodology can be applied equally well to nongradient systems. In this section we illustrate this on the example of the Brusselator:

$$\begin{cases} u_t = \dfrac{1}{\alpha}(u_{xx} + 1 + u^2 v - (1+A)u) \\ v_t = v_{xx} + Au - u^2 v\,, \end{cases} \tag{5.1}$$

where A and α are two parameters. We consider these equations on $x \in [0, 1]$ and impose Neumann boundary conditions

$$u_x(0,t) = u_x(1,t) = 0\,, \quad v_x(0,t) = v_x(1,t) = 0\,. \tag{5.2}$$

The Brusselator was introduced as a simple model of a nonlinear chemical system in which the relative concentration of products can oscillate in time as in, for example, the Belousov-Zhabotinski reaction [17, 20].

The Brusselator has a stable fixed point at $(1, A)$. We will be interested in finding the optimal path from the stable fixed point to another point in the phase space under the influence of a small noise,

$$\begin{cases} u_t = \dfrac{1}{\alpha}(u_{xx} + 1 + u^2 v - (1+A)u) + \sqrt{2\varepsilon}\,\eta_1(x,t) \\ v_t = v_{xx} + Au - u^2 v + \sqrt{2\varepsilon}\,\eta_2(x,t)\,, \end{cases} \tag{5.3}$$

where η_1 and η_2 are two independent spatiotemporal white noises and we impose the reflecting boundary condition at $\{u = 0\}$ and $\{v = 0\}$.

The action functional associated with (5.3) is

$$\begin{aligned} S_T[u,v] = {} & \frac{1}{2}\int_0^T \int_0^1 \left(u_t - \frac{1}{\alpha}\left(u_{xx} + 1 + u^2 v - (1+A)u\right)\right)^2 dx\,dt \\ & + \frac{1}{2}\int_0^T \int_0^1 \left(v_t - v_{xx} - Au + u^2 v\right)^2 dx\,dt\,. \end{aligned} \tag{5.4}$$

We minimize this action functional using a straightforward extension of the method described earlier. In Figure 5.1, we plot the spatial means $\bar{u} = \int_0^1 u\,dx$ and $\bar{v} = \int_0^1 v\,dx$ of the optimal pathways for various final states and the contour lines of the corresponding actions.

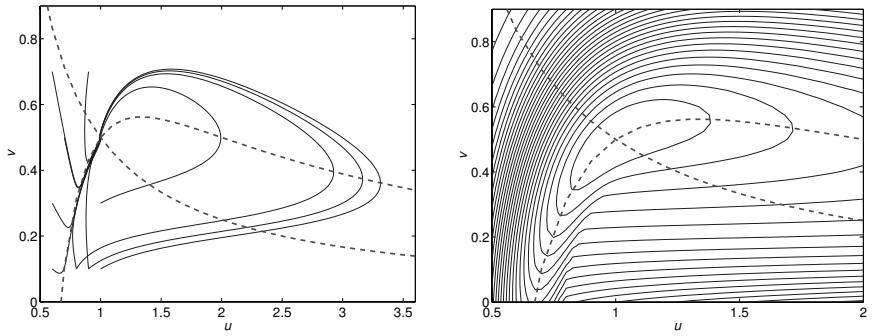

FIGURE 5.1. Left panel: The spatial means $\bar{u} = \int_0^1 u\,dx$ and $\bar{v} = \int_0^1 v\,dx$ of the optimal switching pathways from the stable position (1,A) to various other states. Note that there is more than one local minimizer of the action in the vicinity of $(0.8, 0.1)$. Also shown are the two nullclines (dashed lines). Right panel: The contour lines of the corresponding action as a function of the final states. In the calculation, the parameters are $\alpha = 0.1$, $A = 0.5$, and $T = 2$.

One can see from Figure 5.1 that the optimal paths contain large excursions. This is one of the special features of nongradient systems. More thorough discussion of this example as well as other examples of infinite-dimensional, nongradient systems will be presented elsewhere [16].

6 Discussion

In summary, the least-action principle provided by the Wentzell-Freidlin theory of large deviations is exploited as a numerical tool for finding the optimal dynamical path for spatially extended systems driven by a small noise. We presented the numerical results for the Ginzburg-Landau system in one and two dimensions, as well as applications to nongradient systems. A quasi-Newton method, the limited memory BFGS method, is used to minimize the Wentzell-Freidlin action. Other numerical issues such as the preconditioners are also discussed. We also present analytical results on the nucleation and propagation of domain walls for the Ginzburg-Landau models. The theoretical estimates agree very well with the numerical results.

Although it requires the calculation of the Hessian of the energy functional and it is in general more expensive than methods based on computing the minimal energy path (such as the string method or NEB), the minimum action method has the advantage of being applicable to nongradient systems and finite-time events. Although the latter have exponentially small probability in small samples, they are relevant in large samples because their probability increases with system size and eventually tends to 1.

Appendix: Limited Memory BFGS Method

The BFGS method is one of the most popular and efficient quasi-Newton methods for large-scale problems [18]. It only requires that the gradient of the objective function be supplied at each iterate, and yet achieves superlinear convergence. Suppose $f(x)$ is the objective function that we are minimizing and x_k is the current iterate. The BFGS method defines the next one as

$$x_{k+1} = x_k + \alpha_k p_k \,. \tag{A.1}$$

The step length α_k is chosen to satisfy the Wolfe conditions:

$$\begin{cases} f_{k+1} \le f_k + c_1 \alpha_k \nabla f_k^T p_k \\ \nabla f_{k+1}^T p_k \ge c_2 \nabla f_k^T p_k \end{cases} \tag{A.2}$$

with $0 < c_1 < c_2 < 1$. The search direction p_k is given by

$$p_k = -H_k \nabla f_k \tag{A.3}$$

where H_k is an approximation of the inverse of the Hessian of $f(x)$ at x_k. In the BFGS method, H_k is defined recursively by

$$H_{k+1} = \left(I - \rho_k s_k y_k^T\right) H_k \left(I - \rho_k y_k s_k^T\right) + \rho_k s_k s_k^T \tag{A.4}$$

where

$$s_k = x_{k+1} - x_k \,, \quad y_k = \nabla f_{k+1} - \nabla f_k \,, \quad \text{and} \quad \rho_k = \frac{1}{y_k^T s_k} \,. \tag{A.5}$$

The initial Hessian approximation H_0 plays the role of preconditioner for the problem. In fact, it can easily be shown that the BFGS method is identical to the preconditioned conjugate gradient method with the preconditioner $B_0 = H_0^{-1}$ when applied to strongly convex quadratic functions. One general strategy for the choice of H_0 that has proven to be effective in practice is to use $H_0 = \gamma_1 I$, where the scaling factor γ_1 is defined by

$$\gamma_k = \frac{y_{k-1}^T s_{k-1}}{y_{k-1}^T y_{k-1}} \,. \tag{A.6}$$

When applied to a smooth convex function with an arbitrary starting point x_0 and a symmetric positive definite matrix H_0, the BFGS method can be proven to be globally convergent. Furthermore, the rate of convergence is superlinear, which is adequate for practical problems.

In the BFGS method, H_k defined by (A.4) is usually a dense matrix, so the cost of storing and manipulating it is prohibitively large for large-scale systems. The limited memory BFGS method, abbreviated as L-BFGS, can be used to circumvent this problem. The idea is to store and use the m most recent vector pairs $\{s_i, y_i\}$ to construct the approximation of the Hessian. Once the new iterate x_{k+1} is computed, the oldest vector pair in the set $\{s_i, y_i\}$, which is less likely to be relevant to the

behavior of the objective function at the current iteration, is discarded in order to save the storage and the computational cost.

Acknowledgments. We are grateful to Bob Kohn, Cyrill Muratov, Felix Otto, and Maria Reznikoff for stimulating discussions. The work of E is partially supported by NSF Grant DMS01-30107. Ren is partially supported by NSF Grant DMS97-29992. Vanden-Eijnden is partially supported by NSF Grant DMS02-09959 and by AMIAS (Association of Members of the Institute for Advanced Study).

Bibliography

[1] Berkov, D. V. Numerical calculation of the energy barrier distribution in disordered many-particle systems: the path integral method. *J. Magn. Magn. Mat.* **186** (1998), 199–213.

[2] Da Prato, G.; Zabczyk, J. *Stochastic equations in infinite dimensions.* Encyclopedia of Mathematics and Its Applications, 44. Cambridge University, Cambridge, 1992.

[3] E, W.; Ren, W.; Vanden-Eijnden, E. String method for the study of rare events. *Phys. Rev. B* **66** (2002), 052301.

[4] E, W.; Ren, W.; Vanden-Eijnden, E. Energy landscape and thermally activated switching of submicron-sized ferromagnetic elements. *J. Appl. Phys.* **93** (2003), 2275–2282.

[5] E, W.; Ren, W.; Vanden-Eijnden, E. Probing multi-scale energy landscapes using the string method. Unpublished paper.

[6] Faris, W. G.; Jona-Lasinio, G. Large fluctuations for a nonlinear heat equation with noise. *J. Phys. A* **15** (1982), no. 10, 3025–3055.

[7] Feynman, R. P.; Hibbs, A. R. *Quantum mechanics and path integrals.* New York, McGraw-Hill, 1965.

[8] Fogedby, H. C.; Hertz, J.; Svane, A. Soliton-dynamical approach to a noisy Ginzburg-Landau model. Preprint.

[9] Freidlin, M. I. Limit theorems for large deviations and reaction-diffusion equations. *Ann. Probab.* **13** (1985), no. 3, 639–675.

[10] Freidlin, M. I. *Semi-linear PDEs and limit theorems for large deviations.* Ecole d'Eté de Probabilités de Saint-Flour XX—1990, 1–109. Lecture Notes in Mathematics, 1527. Springer, Berlin, 1992.

[11] Freidlin, M. I.; Wentzell, A. D. *Random perturbations of dynamical systems*. 2nd ed. Springer, New York, 1998.

[12] Holden, H.; Øksendal, B.; Ubøe, J.; Zhang, T. *Stochastic partial differential equations.* Birkhaüser, Boston, 1996.

[13] Jónsson, H.; Mills, G.; Jacobsen, K. W. Nudged elastic band method for finding minimum energy paths of transitions. *Classical and quantum dynamics in condensed phase simulations.* B. J. Berne, G. Ciccotti, and D. F. Coker, eds. World Scientific, Singapore–Hong Kong–London, 1998.

[14] Maier, R. S.; Stein, D. L. Droplet nucleation and domain wall motion in a bounded interval. *Phys. Rev. Lett.* **87** (2001), no. 27(I), 601–800.

[15] Marcus, R. Parabolic Itô equations with monotone nonlinearities. *J. Funct. Anal.* **29** (1978), no. 3, 275–286.

[16] Muratov, C; Ren, W.; Vanden-Eijnden, E.; E, W. Minimum action method for non gradient systems. In preparation.

[17] Nicolis, G.; Prigogine, I. *Self-organization in nonequilibrium systems.* Wiley-Interscience, New York, 1977.

[18] Nocedal, J. *Numerical optimization*. Springer Series in Operations Research. Springer, New York, 1999.

[19] Olender, R.; Elber, R. Calculation of classical trajectories with a very large time step: formalism and numerical examples. *J. Chem. Phys.* **105** (1996), 9299–9315.

[20] Reichl, L. E. *A modern course in statistical physics.* University of Texas, Austin, 1980.

[21] Ren, W. Numerical methods for the study of energy landscapes and rare events. Doctoral dissertation, New York University, 2002.

[22] Reznikoff, M.; Ren, W.; E, W.; Kohn, R. V.; Otto, F.; Vanden-Eijnden, E. Non equilibrium interface dynamics. In preparation.

WEINAN E
Princeton University
Department of Mathematics and PACM
Fine Hall, Washington Road
Princeton, NJ 08544
E-mail: weinan@princeton.edu

WEIQING REN
Institute for Advanced Study
School of Mathematics
1 Einstein Drive
Princeton, NJ 08540
E-mail: weiqing@ias.edu

ERIC VANDEN-EIJNDEN
Courant Institute
251 Mercer Street
New York, NY 10012-1185
E-mail: eve2@cims.nyu.edu

Received May 2003.
Revised August 2003.