MULTISCALE MODEL. SIMUL.
Vol. 11, No. 4, pp. 1000–1032

ANOMALOUS SHOCK DISPLACEMENT PROBABILITIES FOR A PERTURBED SCALAR CONSERVATION LAW*

JOSSELIN GARNIER†, GEORGE PAPANICOLAOU‡, AND TZU-WEI YANG‡

Abstract. We consider a one-dimensional conservation law with random space-time forcing and calculate the exponentially small probabilities of anomalous shock profile displacements using large deviations theory. Under suitable hypotheses on the spatial support and structure of random forces, we analyze the scaling behavior of the rate function, which is the exponential decay rate of the displacement probabilities. For small displacements we show that the rate function is bounded above and below by the square of the displacement divided by time. For large displacements the corresponding bounds for the rate function are proportional to the displacement. We calculate numerically the rate function under different conditions and show that the theoretical analysis of scaling behavior is confirmed. We also apply a large-deviation-based importance sampling Monte Carlo strategy to estimate the displacement probabilities. We use a biased distribution centered on the forcing that gives the most probable transition path for the anomalous shock profile, which is the minimizer of the rate function. The numerical simulations indicate that this strategy is much more effective and robust than basic Monte Carlo.

Key words. conservation laws, shock profiles, large deviations, Monte Carlo methods, importance sampling

AMS subject classifications. 60F10, 35L65, 35L67, 65C05

DOI. 10.1137/120883864

1. Introduction.

It is well known that nonlinear waves are not very sensitive to perturbations in initial conditions or ambient medium inhomogeneities [10, Chap. 8, 19]. This is in contrast to linear waves in random media, where even weak inhomogeneities can affect significantly wave propagation over long times and distances. It is natural, therefore, to consider perturbations of shock profiles of randomly forced conservation laws as rare events and use large deviations theory. The purpose of this paper is to calculate probabilities of anomalous shock profile displacements for randomly perturbed one-dimensional conservation laws. We analyze the rate function that characterizes the exponential decay of displacement probabilities and show that under suitable hypotheses on the random forcing they have scaling behavior relative to the size of the displacement (with respect to the noise-free case) and the time interval on which it occurs.

The theory of large deviations for conservation laws with random forcing is an extension of the Freidlin–Wentzell theory of large deviations [11, 7] to partial differential equations. This has been carried out extensively [1, 2, 19] and we use this theory here. We are interested in a more detailed analysis of the exponential probabilities of anomalous shock profile displacements, which leads to an analysis of the rate function associated with large deviations for this particular class of rare events. We derive upper and lower bounds for the exponential decay rate of the small probabilities, using

*Received by the editors July 9, 2012; accepted for publication (in revised form) July 2, 2013; published electronically October 16, 2013. This work was supported partly by Department of Energy (National Nuclear Security Administration) award NA28614 and partly by AFOSR grant FA9550-11-1-0266.

http://www.siam.org/journals/mms/11-4/88386.html

†Laboratoire de Probabilités et Modèles Aléatoires & Laboratoire Jacques-Louis Lions, Université Paris VII, 75205 Paris Cedex 13, France (garnier@math.univ-paris-diderot.fr).

‡Department of Mathematics, Stanford University, Stanford, CA 94305 (papanicolaou@stanford.edu, twyang@stanford.edu).

suitable test functions for the variational problem involving the rate function. This is the main result of this paper. The applications we have in mind come from uncertainty quantification [14] in connection with simplified models of flow and combustion in a scramjet. Numerical calculations of the exponential decay rates from variational principles associated with the rate function have been carried out before [9, 27]. We carry out such numerical calculations here and confirm the scaling behavior of bounds obtained theoretically. We use a gradient descent method to do the optimization numerically, and we note that its convergence is quite robust even though the functional under consideration is not known to be convex. This robustness suggests the Monte Carlo simulations with importance sampling using a change of measure based on the minimizer of the discrete rate function is likely to be effective. Our simulations show that indeed such importance sampling Monte Carlo performs much better than the basic Monte Carlo method.

The paper is organized as follows. In section 2 we formulate the one-dimensional conservation law problem. In section 3 we state the large deviation principle and identify the rate function which we will use. In section 4 we give a simple, explicitly computable case of shock profile displacement probabilities that can be used to compare with the results of the large deviation theory. Section 5 contains the main results of the paper, which are the upper and lower bounds for the exponential decay rate of the displacement probabilities, under different conditions on the random forcing. We identify scaling behavior of these probabilities, relative to size of the displacement and the time interval of interest. In section 6 we introduce a discrete form for the conservation law and the associated large deviations, and calculate numerically the displacement probabilities from the discrete variational principle. In section 7 we implement importance sampling Monte Carlo based on the minimizer of the discrete rate function and we compare it with the basic Monte Carlo method. We end with a brief section summarizing the paper and our conclusions.

2. The perturbed conservation law. We consider the scalar viscous conservation law:

$$u_t + (F(u))_x = (Du_x)_x, \qquad x \in \mathbb{R}, \quad t \in [0, \infty), \tag{2.1}$$

$$u(t = 0, x) = u_0(x), \qquad x \in \mathbb{R}. \tag{2.2}$$

Here D is constant, $F \in \mathcal{C}^\infty(\mathbb{R})$, and the initial condition satisfies $u_0(x) \to u_\pm$ as $x \to \pm\infty$, where $u_- < u_+$. We are interested in traveling wave solutions of the form $U(x - \gamma t)$, where the smooth profile U satisfies $U(x) \to u_\pm$ as $x \to \pm\infty$ and the wave speed γ is given by the Rankine–Hugoniot condition

$$\gamma = \frac{F(u_+) - F(u_-)}{u_+ - u_-}. \tag{2.3}$$

A traveling wave U with wave speed γ exists, provided the following conditions are fulfilled:

$$F(u) - F(u_-) > \gamma(u - u_-) \quad \forall u \in (u_+, u_-), \tag{2.4}$$

$$F'(u_+) < \gamma < F'(u_-). \tag{2.5}$$

The first condition is the Oleinik entropy condition [15] and the second one is the Lax entropy condition [17]. Under these conditions the traveling wave profile exists, it is the solution of the ordinary differential equation

$$U_x = \frac{1}{D}\big(F(U) - F(u_-) - \gamma(U - u_-)\big), \qquad U(x) \overset{x \to -\infty}{\longrightarrow} u_-,$$

and it is orbitally stable, which means that perturbations of the profile decay in time, and thus initial conditions near the traveling wave profile converge to it. Note that a physically admissible viscous profile must have the stability property; otherwise, it would not be observable. As noted in [16], the motivating idea behind the orbital stability result is that in the stabilizing process, information is transferred from spatial decay of the profile U at infinity to temporal decay of the perturbation.

The purpose of this paper is to address another type of stability, that is, the stability with respect to external noise. We consider the perturbed scalar viscous conservation law with additive noise:

$$u_t^\varepsilon + (F(u^\varepsilon))_x = (Du_x^\varepsilon)_x + \varepsilon \dot{W}(t,x), \qquad x \in \mathbb{R}, \quad t \in [0,\infty), \tag{2.6}$$

$$u^\varepsilon(t=0,x) = U(x), \qquad x \in \mathbb{R}. \tag{2.7}$$

Here ε is a small parameter, $W(t,x)$ is a zero-mean random process (described below), and the dot stands for the time derivative. We would like to address the stability of the traveling wave $U(x-\gamma t)$ driven by the noise $\varepsilon\dot{W}$. Motivated by an application modeling combustion in a scramjet [14, 26], we have in mind a specific rare event, which is an exceptional or anomalous shift of the position of the traveling wave compared to the unperturbed motion with the constant velocity γ.

We consider mild solutions which satisfy (denoting $u^\varepsilon(t) = (u^\varepsilon(t,x))_{x\in\mathbb{R}}$)

$$u^\varepsilon(t) = S(t)U - \int_0^t S(t-s)N(u^\varepsilon(s))ds + \varepsilon \int_0^t S(t-s)dW(s), \tag{2.8}$$

where $S(t)$ is the heat semigroup with kernel

$$S(t,x,y) = \frac{1}{\sqrt{4\pi D t}} \exp\Big(- \frac{(x-y)^2}{4Dt}\Big),$$

and $N(u)(x) = (F(u(x)))_x$. The main result about the heat kernel [5, Chap. XVI, sect. 3] is as follows.

LEMMA 2.1. *If $f = (f(t))_{t\in[0,T]} \in L^2([0,T], H^{-1}(\mathbb{R}))$, then the function $\int_0^t S(t-s)f(s)ds$ is in $L^2([0,T], H^1(\mathbb{R})) \cap \mathcal{C}([0,T], L^2(\mathbb{R}))$.*

A white noise or cylindrical Wiener process $B(t,x)$ in the Hilbert space $L^2(\mathbb{R})$ is such that for any complete orthonormal system $(f_n(x))_{n\geq 1}$ of $L^2(\mathbb{R})$, there exists a sequence of independent Brownian motions $(\beta_n(t))_{n\geq 1}$ such that

$$B(t,x) = \sum_{n=1}^\infty \beta_n(t) f_n(x). \tag{2.9}$$

$B(t,x)$ can be seen as the (formal) spatial derivative of the Brownian sheet on $[0,\infty)\times\mathbb{R}$, which means that it is the Gaussian process with mean zero and covariance $\mathbb{E}[B(t,x)B(s,y)] = s\wedge t\, \delta(x-y)$. Note that the sum (2.9) does not converge in L^2 but in any Hilbert space H such that the embedding from $L^2(\mathbb{R})$ to H is Hilbert–Schmidt. Therefore, the image of the process B by a linear mapping on $L^2(\mathbb{R})$ is a well-defined process in the Sobolev space $H^k(\mathbb{R})$ (for $k=0$, we take the convention $H^0 = L^2$) when the mapping is Hilbert–Schmidt from $L^2(\mathbb{R})$ into $H^k(\mathbb{R})$. Then, if the kernel Φ is Hilbert–Schmidt in the sense that

$$\sum_{n=1}^\infty \|\Phi f_n\|_{H^k(\mathbb{R})}^2 = \sum_{j=0}^k \|\partial_x^j \Phi(\cdot,\cdot)\|_{L^2(\mathbb{R}\times\mathbb{R})}^2 < \infty,$$

then the random process $W = \Phi B$ in (2.6), where

$$W(t,x) = \int \Phi(x,x')B(t,x')dx'$$

is well-defined in $H^k(\mathbb{R})$. It is a zero-mean Gaussian process with covariance function given by

$$\mathbb{E}[W(t,x)W(t',x')] = t \wedge t'\, C(x,x'),$$
$$C(x,x') = \Phi\Phi^T(x,x') = \int \Phi(x,x'')\Phi(x',x'')dx'',$$

where Φ^T stands for the adjoint of Φ.

As an example, we may think of $\Phi(x,x') = \Phi_0(x)\Phi_1(x-x')$ with $\Phi_0 \in H^k(\mathbb{R})$ and $\Phi_1 \in H^k(\mathbb{R})$. In this case Φ_0 characterizes the spatial support of the additive noise and Φ_1 characterizes its local correlation function. In the limit $\Phi_0(x) = 1$ and $\Phi_1(x-x') = \delta(x-x')$, the process $\dot{W}(t,x)$ in (2.6) is a space-time white noise.

By adapting the technique used in [5, Chap. XVI, sect. 3] we obtain the following lemma.

LEMMA 2.2. *If Φ is Hilbert–Schmidt from L^2 into L^2, then $Z(t) = \int_0^t S(t-s)dW(s)$ belongs to $L^2([0,T],H^1(\mathbb{R})) \cap \mathcal{C}([0,T],L^2(\mathbb{R}))$ almost surely. If Φ is Hilbert-Schmidt from L^2 into H^1, then $Z(t)$ belongs to $L^2([0,T],H^2(\mathbb{R})) \cap \mathcal{C}([0,T],H^1(\mathbb{R}))$ almost surely.*

Proof. See Appendix A. □

3. Large deviation principle. In this section we state a large deviation principle (LDP) for the solution $(u^\varepsilon(t,x))_{t\in[0,T],x\in\mathbb{R}}$ of the randomly perturbed scalar conservation law (2.6). It generalizes the classical Freidlin–Wentzell principle for finite-dimensional diffusions. Throughout this paper, we assume that the flux F is a $\mathcal{C}^\infty$ function with bounded first and second derivatives.

Assumption 3.1. In this paper, the flux F is a $\mathcal{C}^\infty$ function and there exists $C_F < \infty$ such that $\|F'\|_{L^\infty(\mathbb{R})}$ and $\|F''\|_{L^\infty(\mathbb{R})}$ are bounded by C_F.

Remark. Assumption 3.1 is merely a technical assumption to simplify the proofs of Propositions 3.2 and 3.3, and obviously it violates the convexity of fluxes in general in conservation laws. From the physical point of view, for a general flux F, we can choose a very large constant M, let $F_M(u) = F(u)$ for $|u| \le M$, and saturate $F_M(u)$ for $|u| > M$. If $[-M,M]$ can cover the range of interest of u, then $F_M(u) = F(u)$. Without a proof, we point out that this physical argument can be proven mathematically: if u_t^ε, u_x^ε and u_{xx}^ε in (2.6) are continuous on $[0,T]\times\mathbb{R}$ and $\dot{W}$ is bounded on $[0,T]\times\mathbb{R}$, then the parabolic maximum principle implies that $|u^\varepsilon|$ is bounded on $[0,T]\times\mathbb{R}$ and thus we can find $M < \infty$. The technical difficulty is that W is not differentiable in time, so we can not apply the parabolic maximum principle directly. However, we can consider a modified $\tilde{u}^\varepsilon$ by replacing W by a smooth $\tilde{W}$ in (2.6) and show that $\|\tilde{u}^\varepsilon(t)\|_{L^\infty(\mathbb{R})} \to \|u^\varepsilon(t)\|_{L^\infty(\mathbb{R})}$ as $\tilde{W} \to W$.

We first describe the functional space to which the solution of the perturbed conservation law belongs.

PROPOSITION 3.2. *If Φ is Hilbert–Schmidt from L^2 to H^1, then for any $\varepsilon > 0$ there is a unique solution u^ε to (2.8) in the space $\mathcal{E}^1$ almost surely, where*

$$\mathcal{E}^1 = \big\{(u(t,x))_{t\in[0,T],x\in\mathbb{R}} :\ (u(t,x) - U(x))_{t\in[0,T],x\in\mathbb{R}} \in \mathcal{C}([0,T],H^1(\mathbb{R}))\big\}.$$

Proof. See Appendix B.1. □

The rate function of the LDP is defined in terms of the mild solution of the control problem

$$u_t + (F(u))_x = (Du_x)_x + \Phi h(t,x), \qquad x \in \mathbb{R}, \quad t \in [0,T], \tag{3.1}$$

$$u(t=0,x) = U(x), \qquad x \in \mathbb{R}, \tag{3.2}$$

where $h \in L^2([0,T], L^2(\mathbb{R}))$. The solution u to this problem is denoted by $\mathcal{H}[h]$ and $\mathcal{H}$ is a mapping from $L^2([0,T], L^2(\mathbb{R}))$ to $\mathcal{E}^1$ provided Φ is Hilbert–Schmidt from L^2 to H^1. The following proposition is an extension of the LDP proved in [1].

PROPOSITION 3.3. *If Φ is Hilbert–Schmidt from L^2 to H^1, then the solutions u^ε satisfy a large deviation principle in $\mathcal{E}^1$ with the good rate function*

$$I(u) = \inf_{h, u=\mathcal{H}[h]} \frac{1}{2} \int_0^T \|h(t,\cdot)\|_{L^2}^2 dt, \tag{3.3}$$

with the convention $\inf \emptyset = \infty$.

Proof. See Appendix B.2. □

In fact, $I(u) < \infty$ if and only if u is in the range of $\mathcal{H}$. The LDP means that, for any $A \subset \mathcal{E}^1$, we have

$$-\mathcal{J}(\mathring{A}) \le \liminf_{\varepsilon \to 0} \varepsilon^2 \log \mathbb{P}(u^\varepsilon \in A) \le \limsup_{\varepsilon \to 0} \varepsilon^2 \log \mathbb{P}(u^\varepsilon \in A) \le -\mathcal{J}(\overline{A}), \tag{3.4}$$

where

$$\mathcal{J}(A) = \inf_{u \in A} I(u), \tag{3.5}$$

with the convention $\mathcal{J}(\emptyset) = \infty$. Note that the interior $\mathring{A}$ and closure $\overline{A}$ are taken in the topology associated with $\mathcal{E}^1$. Assuming that A is closed and has nonempty interior, the most probable path u^* for the rare event A is defined as the minimizer of the rate function over A:

$$u^* := \arg\min_{u \in A} I(u) \tag{3.6}$$

Moreover, if u^* exists and is unique, $\mathbb{P}(u^\varepsilon \in A) \neq 0$, and $\mathcal{J}(\mathring{A}) = \mathcal{J}(A)$, then for any open neighborhood $\mathbf{N}(u^*) \subset \mathcal{E}^1$ containing u^*,

$$\begin{aligned} \mathbb{P}(u^\varepsilon \in \mathbf{N}(u^*) | u^\varepsilon \in A) &= 1 - \mathbb{P}(u^\varepsilon \in \mathbf{N}^C(u^*) | u^\varepsilon \in A) = 1 - \frac{\mathbb{P}(u^\varepsilon \in \mathbf{N}^C(u^*) \cap A)}{\mathbb{P}(u^\varepsilon \in A)} \\ &\gtrsim 1 - \frac{\exp(-\frac{1}{\varepsilon^2} \mathcal{J}(u^\varepsilon \in \mathbf{N}^C(u^*) \cap A))}{\exp(-\frac{1}{\varepsilon^2} \mathcal{J}(A))} \overset{\varepsilon \to 0}{\to} 1. \end{aligned} \tag{3.7}$$

In other words, the mass of the conditional probability concentrates exponentially fast around the most probable path u^*.

Although the convexity of the rate function I is unknown, it is possible to show that I satisfies the maximum principle: if the set of the rare event A does not contain the exact solution $U(x - \gamma t)$, then $\inf_{u \in A} I(u)$ attains its minimum at the boundary of A.

PROPOSITION 3.4. *If $U(x - \gamma t) \notin A$ and $u \in \mathring{A}$, then there exists a sequence $\{u^n\}$ in $\mathcal{E}^1$ such that $I(u^n) < I(u)$ and $u^n \to u$ in $\mathcal{E}^1$ as $n \to \infty$. As a consequence, any $u \in \mathring{A}$ can not be a local minimizer of I.*

Proof. See Appendix B.3. □

4. Wave displacement from an elementary point of view. In this paper we are interested in estimating the probability of large deviations from the deterministic path $U(x-\gamma t)$. In this section, we first study in a very elementary way how the center of the solution u^ε at time T can deviate from its unperturbed value. The center of a function $u \in \mathcal{E}^1$ is defined as

$$\mathcal{C}[u](t) = -\frac{\int_{-\infty}^{\infty}[u(t,x) - U(x)]dx}{u_- - u_+}, \tag{4.1}$$

provided that the integral is well-defined.

PROPOSITION 4.1. *If the covariance function C is in $L^1(\mathbb{R}\times\mathbb{R})$, then the center of u^ε is well-defined for any time $t \in [0,T]$. It is a Gaussian process with mean γt and covariance*

$$\mathbf{Cov}\big(\mathcal{C}[u^\varepsilon](t), \mathcal{C}[u^\varepsilon](t')\big) = \varepsilon^2 \frac{\iint C(x,x')dxdx'}{(u_- - u_+)^2}\, t \wedge t'. \tag{4.2}$$

Proof. See Appendix C. □

In the absence of noise, the center of the solution increases linearly as γt. In the presence of noise we can characterize the probability of the rare event

$$B = \big\{u \in \mathcal{E}^1,\, \mathcal{C}[u](T) \geq \gamma T + X_0\big\}, \tag{4.3}$$

where $X_0 \in [0,\infty)$.

PROPOSITION 4.2. *If the covariance function C is in $L^1(\mathbb{R}\times\mathbb{R})$, then*

$$\mathbb{P}(u^\varepsilon \in B) \sim \exp\Big(-\frac{X_0^2(u_- - u_+)^2}{2\varepsilon^2 T \iint C(x,x')dxdx'}\Big). \tag{4.4}$$

The approximate equality means that

$$\lim_{\varepsilon\to 0} \varepsilon^2 \log \mathbb{P}(u^\varepsilon \in B) = -\frac{X_0^2(u_- - u_+)^2}{2T\iint C(x,x')dxdx'}.$$

This proposition is a direct corollary of Proposition 4.1.

If $X_0 = 0$, then $U(x-\gamma t) \in B$ and $\mathbb{P}(u^\varepsilon \in B) = 1/2$. If $X_0 > 0$, then the set B is indeed exceptional in that it corresponds to the event in which the center of the profile is anomalously ahead of its expected position. Note that the scaling X_0^2/T in (4.4) corresponds to that of the exit problem of a Brownian particle. This is not surprising since the center $\mathcal{C}[u^\varepsilon](t)$ behaves like a Brownian motion by Proposition 4.1.

The LDP for u^ε stated in Proposition 3.3 is not used here for two reasons:

1. It does not give a good result because the interior of B in $\mathcal{E}^1$ is empty (since we can construct a sequence of functions bounded in H^1 that blows up in L^1).
2. The distribution of the center $\mathcal{C}[u^\varepsilon](T)$ is here explicitly known for any ε. This is fortunate and it is not always true. When the rare event is more complex, then the LDP for u^ε is useful as we will show in the next section.

5. Wave displacement from the large deviations point of view.

5.1. The framework and result. In this section, we use the large deviation principle to compute the probability of the rare event that the perturbed traveling wave at time T is close to the same profile but with the displacement X_0:

$$A_\delta = \{u \in \mathcal{E}^1 \text{ such that } u(0,\cdot) = U(x),\, \|u(T,\cdot) - U(\cdot - \gamma T - X_0)\|_{H^1} \leq \delta\}. \tag{5.1}$$

By Proposition 3.3, we have

$$-\mathcal{J}(\mathring{A}_\delta) \le \liminf_{\varepsilon\to 0} \varepsilon^2 \log \mathbb{P}(u^\varepsilon \in A_\delta) \le \limsup_{\varepsilon\to 0} \varepsilon^2 \log \mathbb{P}(u^\varepsilon \in A_\delta) \le -\mathcal{J}(A_\delta). \tag{5.2}$$

We define $A = A_\delta|_{\delta=0}$:

$$A = \{u \in \mathcal{E}^1 \text{ such that } u(0,x) = U(x),\ u(T,x) = U(x - \gamma T - X_0)\}. \tag{5.3}$$

The following lemma shows that $\mathcal{J}(A_\delta)$ converges to $\mathcal{J}(A)$ as $\delta \to 0$.

LEMMA 5.1. *By definition $\mathcal{J}(A_\delta)$ is a decreasing function with δ and bounded from above by $\mathcal{J}(A)$. In addition,*

$$\lim_{\delta\to 0} \mathcal{J}(A_\delta) = \mathcal{J}(A).$$

Proof. See Appendix D.1. In fact, the same proof works for any set A and A_δ with arbitrary initial and terminal conditions. ∎

By Lemma 5.1 and the fact that $\mathcal{J}(\mathring{A}_\delta) \le \mathcal{J}(A)$,

$$\begin{aligned} -\mathcal{J}(A) \le -\mathcal{J}(\mathring{A}_\delta) &\le \liminf_{\varepsilon\to 0} \varepsilon^2 \log \mathbb{P}(u^\varepsilon \in A_\delta) \\ &\le \limsup_{\varepsilon\to 0} \varepsilon^2 \log \mathbb{P}(u^\varepsilon \in A_\delta) \le -\mathcal{J}(A_\delta) \overset{\delta\to 0}{\to} -\mathcal{J}(A). \end{aligned}$$

Namely, for ε and δ small, we have

$$\mathbb{P}(u^\varepsilon \in A_\delta) \sim \exp\left(-\frac{1}{\varepsilon^2}\mathcal{J}(A)\right),$$

and it thus suffices to consider the bounds for $\mathcal{J}(A)$.

The value of the rate function heavily depends on the kernel Φ, and formally the simplest kernel we can have is the identity operator. Of course, the identity operator is not Hilbert–Schmidt, but with the given A, we can construct a Hilbert–Schmidt Φ such that Φ is approximately the identity in the region of interest.

Assumption 5.2.

1. We assume that the center of transition of $U(x)$ in (5.3) is 0 and $X_0 + \gamma T \ge 0$.
2. The kernel $\Phi^{l_c}_{L_0}$ has the following form:

$$\Phi^{l_c}_{L_0}(x,x') = \sigma\phi_0\Big(\frac{x}{L_0}\Big)\frac{1}{l_c}\phi_1\Big(\frac{x-x'}{l_c}\Big), \tag{5.4}$$

   where σ, L_0, and l_c are positive constants.
3. ϕ_0 and ϕ_1 are $L^1 \cap H^1$ functions so that their Fourier transforms are well-defined in $\mathbb{R}$ and Φ is Hilbert–Schmidt from L^2 to H^1 (see Lemma 5.4). ϕ_0 and ϕ_1 are normalized so that $\phi_0(0) = \hat{\phi}_1(0) = 1$.
4. $\phi_0 \in \mathcal{C}^\infty$ is positive valued and $\hat{\phi}_1$ is nonzero. In addition, $1/\phi_0(x)$ and $1/\hat{\phi}_1(\xi)$ have at most polynomial growth at $\pm\infty$.
5. $\phi_0(x)$ is increasing on $x \in (-\infty,-1)$, decreasing on $x \in (1,\infty)$ and identically equal to 1 on $x \in [-1,1]$. By this setting, the support of the noise kernel Φ is roughly $[-L_0, L_0]$.

The following theorem is the main result of this section, and the proof will be given in the next two subsections.

THEOREM 5.3. *Let A be defined by* (5.3). *Let $T_0 > 0$. There exist constants C_0, C_1, C_2 such that for all $T \le T_0$, for all X_0, for all $\Phi^{l_c}_{L_0}$ satisfying Assumption*

5.2 *with* $L_0 = \gamma T + X_0 + C_0$*, the quantity* $\mathcal{J}(A)$ *of the large deviation problem* (3.1) *generated by* $\Phi^{l_c}_{L_0}$ *satisfies*

$$C_1 \frac{X_0^2}{1+|X_0|}\frac{1}{T} \le \mathcal{J}(A) \le C_2 \frac{X_0^2}{1+|X_0|}\frac{1}{T} \tag{5.5}$$

if l_c *is small enough.*

Proof. See sections 5.2 and 5.3. ∎

Remark. $C_0 = \max\{C_v, C_w\}$, where C_v and C_w will be determined in Lemmas 5.5 and 5.9, respectively. $L_0 = \gamma T + X_0 + C_0$ means that the range of noise covers the region of interest of A so that the rare event A is possible and at the same time the range is not too large to cause waste of energy; the small l_c means that the noise in this region is weakly correlated and can be treated as the white noise.

We note that there are two occurrences of T in $\mathcal{J}(A)$: in (3.3), we integrate the $\|h(t,\cdot)\|^2_{L^2}$ from 0 to T, and the rare event A is T-dependent. These two occurrences of T make the large-T behavior of $\mathcal{J}(A)$ nontrivial, which is why Theorem 5.3 addresses only the behavior for $T \le T_0$.

In the rest of this subsection we discuss two relevant issues about this framework.

We first show that the given kernel $\Phi^{l_c}_{L_0}$ in Assumption 5.2 is indeed Hilbert–Schmidt from L^2 to H^1.

LEMMA 5.4. *If* ϕ_0 *and* ϕ_1 *are both in* $L^1 \cap H^k$ *for* $k \ge 1$*, then the kernel* $\Phi^{l_c}_{L_0}$ *of the form* (5.4) *is Hilbert–Schmidt from* L^2 *to* H^k*.*

Proof. See Appendix D.2. ∎

Second we compare the results stated in Theorem 5.3, valid under Assumption 5.2, and those stated in Proposition 4.4. It is easy to check that $A \subset B$. Note that

$$\iint C(x,x')dxdx' = \int \Big[\int \Phi(x,x')dx\Big]^2 dx'.$$

If the kernel $\Phi^{l_c}_{L_0}$ is of the form (5.4), then we find

$$\iint C(x,x')dxdx' = \frac{\sigma^2 L_0}{2\pi}\int |\hat\phi_1(\frac{l_c}{L_0}\xi)|^2 |\hat\phi_0(\xi)|^2 d\xi.$$

By assuming $l_c \ll L_0$ (and $\hat\phi_1(0) = 1$) as in Assumption 5.2, then this simplifies to

$$\iint C(x,x')dxdx' \simeq \sigma^2 L_0 \int \phi_0(x)^2 dx.$$

Therefore, from Assumption 5.2 (in which $L_0 = \gamma T + X_0 + C_0$), we get that there exist constants C_1', C_2' such that (for any $T \le T_0$ and for any X_0)

$$C_1'(1+|X_0|) \le \iint C(x,x')dxdx' \le C_2'(1+|X_0|)$$

for l_c small enough. This gives, with (4.4),

$$\mathbb{P}(u^\varepsilon \in B) \sim \exp\left(-\frac{1}{\varepsilon^2}\mathcal{J}_B\right), \quad C_1'\frac{X_0^2}{1+|X_0|}\frac{1}{T} \le \mathcal{J}_B \le C_2'\frac{X_0^2}{1+|X_0|}\frac{1}{T}. \tag{5.6}$$

We therefore recover the same asymptotic scales for $\mathcal{J}_B$ as the ones for $\mathcal{J}(A)$ in (5.5).

5.2. Proof of Theorem 5.3 for small $|X_0|$. Here we prove the bounds for $\mathcal{J}(A)$ in (5.5) when $|X_0|$ is small. We first consider the upper bound. By definition, $\mathcal{J}(A) \le I(u)$ for any $u \in A$; therefore the idea to obtain the upper bound is to find good test functions u.

For $|X_0|$ small, we use the linear shifted profile as the test function:

$$v(t,x) = U(x - (t/T)(\gamma T + X_0)).$$

If the kernel $\Phi^{l_c}_{L_0} \approx \sigma^{-2}$ in the region of interest, then

$$\mathcal{J}(A) = \inf_{u\in A} I(u) \le I(v) \lesssim \frac{1}{2\sigma^2}\int_0^T \|v_t + F(v)_x - Dv_{xx}\|^2_{L^2} dt. \tag{5.7}$$

We show (5.7) rigorously by the following lemmas.

LEMMA 5.5. *Given the traveling wave $U(x)$ in (5.3), there exists $C_v \ge 1$ such that, uniformly in X_0 and T and for all $t \in [0,T]$,*

$$\|[\phi_0^{-1}(\cdot/L_0) - 1]U_x(\cdot - (t/T)(\gamma T + X_0))\|^2_{L^2} \le 1, \tag{5.8}$$

where $L_0 = \gamma T + X_0 + C_v$.

Proof. See Appendix D.3. ∎

LEMMA 5.6. *Let $v(t,x) = U(x - (t/T)(\gamma T + X_0))$. Given $L_0 = X_0 + \gamma T + C_v$ in Lemma 5.5, $\Phi^{l_c}_{L_0}$ satisfying Assumption 5.2, and $h^{l_c}_{L_0}$ defined by*

$$v_t + F(v)_x - Dv_{xx} = \Phi^{l_c}_{L_0} h^{l_c}_{L_0}, \tag{5.9}$$

then

$$\int_0^T \|h^{l_c}_{L_0}(t,\cdot)\|^2_{L^2} dt \overset{l_c\to 0}{\to} \frac{1}{\sigma^2}\int_0^T \|(v_t + F(v)_x - Dv_{xx})/\phi_0(\cdot/L_0)\|^2_{L^2} dt. \tag{5.10}$$

Proof. See Appendix D.4. ∎

Now we are ready to prove (5.7). By Lemma 5.6, we have

$$I(v) \le \frac{1}{2}\int_0^T \|h^{l_c}_{L_0}(t,\cdot)\|^2_{L^2} dt \overset{l_c\to 0}{\to} \frac{1}{2\sigma^2}\int_0^T \|(v_t + F(v)_x - Dv_{xx})/\phi_0(\cdot/L_0)\|^2_{L^2} dt.$$

We next need to compute $\int_0^T \|(v_t + F(v)_x - Dv_{xx})/\phi_0(\cdot/L_0)\|^2_{L^2} dt$. Note that

$$\big(v_t + F(v)_x - Dv_{xx}\big)(t,x) = \frac{X_0}{T} U_x(x - (t/T)(\gamma T + X_0)).$$

With $L_0 = \gamma T + X_0 + C_v$ given in Lemma 5.5, we have

$$\|U_x(\cdot - (t/T)(\gamma T + X_0))/\phi_0(\cdot/L_0)\|^2_{L^2} \le 2\|U_x(\cdot - (t/T)(\gamma T + X_0))\|^2_{L^2} + 2,$$

and therefore

$$\begin{aligned}
&\int_0^T \|(v_t + F(v)_x - Dv_{xx})/\phi_0(\cdot/L_0)\|^2_{L^2} dt \\
&\quad = \frac{X_0^2}{T^2}\int_0^T \|U_x(\cdot - (t/T)(\gamma T + X_0))/\phi_0(\cdot/L_0)\|^2_{L^2} dt \\
&\quad \le \frac{X_0^2}{T^2}\int_0^T \big(2\|U_x(\cdot - (t/T)(\gamma T + X_0))\|^2_{L^2} + 2\big) dt = \frac{X_0^2}{T}\big(2\|U_x\|^2_{L^2} + 2\big).
\end{aligned}$$

Therefore, we have the asymptotic upper bound for $\mathcal{J}(A)$:

$$\mathcal{J}(A) \le I(v) \overset{l_c\to 0}{\le} \frac{1}{2\sigma^2}\frac{X_0^2}{T}\big(2\|U_x\|_{L^2}^2+2\big). \tag{5.11}$$

Now we find the lower bound for $\mathcal{J}(A)$. Let us first denote by 1 the function identically equal to 1, assume that $(\Phi_{L_0}^{l_c})^T 1 \in L^2$, and find the general form of the lower bound.

PROPOSITION 5.7. *Given $(\Phi_{L_0}^{l_c})^T 1 \in L^2$, for any $u \in A$,*

$$I(u) \ge \frac{1}{2}T^{-1}\|(\Phi_{L_0}^{l_c})^T 1\|_{L^2}^{-2}\left(\int [U(x-X_0)-U(x)]dx\right)^2. \tag{5.12}$$

Proof. See Appendix D.5. ∎

Then we show that $(\Phi_{L_0}^{l_c})^T 1$ is indeed in L^2.

LEMMA 5.8. *Let $\Phi_{L_0}^{l_c}$ satisfy Assumption 5.2. Then $(\Phi_{L_0}^{l_c})^T 1(x)$ is in $L^2(\mathbb{R})$, and $\|(\Phi_{L_0}^{l_c})^T 1\|_{L^2}^2 \to \sigma^2 L_0\|\phi_0\|_{L^2}^2$ as $l_c \to 0$.*

Proof. See Appendix D.6. ∎

From Lemma 5.5, $L_0 = \gamma T + X_0 + C_v$, where C_v is uniform in X_0 and T. Then for l_c small the general lower bound (5.12) becomes

$$\begin{aligned}\mathcal{J}(A) &\ge \frac{1}{2}T^{-1}\|(\Phi_{L_0}^{l_c})^T 1\|_{L^2}^{-2}\left(\int [U(x-X_0)-U(x)]dx\right)^2 \\ &\overset{l_c\to 0}{\to} \frac{1}{2\sigma^2}\|\phi_0\|_{L^2}^{-2}T^{-1}(X_0+\gamma T+C_v)^{-1}\left(\int [U(x-X_0)-U(x)]dx\right)^2.\end{aligned} \tag{5.13}$$

For $|X_0|$ small, $U(x-X_0) = U(x) - X_0U_x(x) + \mathcal{O}(X_0^2)$. Then

$$\begin{aligned}\mathcal{J}(A) &\overset{l_c\to 0}{\ge} \frac{1}{2\sigma^2}\|\phi_0\|_{L^2}^{-2}T^{-1}(X_0+\gamma T+C_v)^{-1}\left(\int [-X_0U_x(x)+\mathcal{O}(X_0^2)]dx\right)^2 \\ &= \frac{1}{2\sigma^2}\|\phi_0\|_{L^2}^{-2}T^{-1}(X_0+\gamma T+C_v)^{-1}[X_0^2(u_+-u_-)^2+\mathcal{O}(X_0^3)],\end{aligned} \tag{5.14}$$

which gives the asymptotic lower bound.

5.3. Proof of Theorem 5.3 for large $|X_0|$. Now we consider the bounds for $\mathcal{J}(A)$ in (5.5) when $|X_0|$ is large. We use the test function

$$w(t,x) = (1-t/T)U(x) + (t/T)U(x-\gamma T - X_0), \tag{5.15}$$

the linear interpolation of the two profiles, and show that

$$\mathcal{J}(A) = \inf_{u\in A} I(u) \le I(w) \lesssim \frac{1}{2\sigma^2}\int_0^T \|w_t + F(w)_x - Dw_{xx}\|_{L^2}^2 dt. \tag{5.16}$$

We also show (5.16) by two technical lemmas similar to Lemmas 5.5 and 5.6.

LEMMA 5.9. *Given the traveling wave $U(x)$ in (5.3) and $w(t,x)$ in (5.15), there exists $C_w \ge 1$ such that, uniformly in X_0 and T and for all $t \in [0,T]$,*

$$\begin{aligned}&\|[\phi_0^{-1}(\cdot/L_0)-1][U(\cdot-\gamma T-X_0)-U]\|_{L^2}^2 \le 1, \\ &\|[\phi_0^{-1}(\cdot/L_0)-1][F(w)_x - Dw_{xx}]\|_{L^2}^2 \le 1,\end{aligned} \tag{5.17}$$

where $L_0 = \gamma T + X_0 + C_w$.

Proof. See Appendix D.7. ∎

Because $w_t + F(w)_x - Dw_{xx} \in \mathcal{C}([0,T], \mathcal{S}(\mathbb{R}))$, we have the following lemma by replacing v by w in Lemma 5.6.

LEMMA 5.10. *Let* w *in* (5.15). *Given* $L_0 = X_0 + \gamma T + C_w$ *in Lemma* 5.9, $\Phi^{l_c}_{L_0}$ *satisfying Assumption* 5.2, *and* $h^{l_c}_{L_0}$ *defined by* $w_t + F(w)_x - Dw_{xx} = \Phi^{l_c}_{L_0} h^{l_c}_{L_0}$, *then*

$$I(w) \le \frac{1}{2}\int_0^T \|h^{l_c}_{L_0}(t,\cdot)\|^2_{L^2} dt \overset{l_c\to 0}{\to} \frac{1}{2\sigma^2}\int_0^T \|(w_t + F(w)_x - Dw_{xx})/\phi_0(\cdot/L_0)\|^2_{L^2} dt.$$

We next need to compute $\int_0^T \|(w_t + F(w)_x - Dw_{xx})/\phi_0(\cdot/L_0)\|^2_{L^2} dt$. Note that

$$w_t + F(w)_x - Dw_{xx} = \frac{1}{T}[U(x - \gamma T - X_0) - U(x)] + F(w)_x - Dw_{xx}.$$

With $L_0 = \gamma T + X_0 + C_w$ given in Lemma 5.9,

$$\begin{aligned}
&\int_0^T \|(w_t + F(w)_x - Dw_{xx})/\phi_0(\cdot/L_0)\|^2_{L^2} dt \\
&\le \frac{2}{T^2}\int_0^T \|[U(\cdot - \gamma T - X_0) - U]/\phi_0(\cdot/L_0)\|^2_{L^2} dt \\
&\quad + 2\int_0^T \|[F(w)_x - Dw_{xx}]/\phi_0(\cdot/L_0)\|^2_{L^2} dt \\
&\le \frac{2}{T^2}\int_0^T \big(2\|[U(\cdot - \gamma T - X_0) - U]\|^2_{L^2} + 2\big) dt \\
&\quad + 2\int_0^T \big(2\|F(w)_x - Dw_{xx}\|^2_{L^2} + 2\big) dt.
\end{aligned}$$

We note that for $|X_0|$ large, $\|U(\cdot - \gamma T - X_0) - U\|^2_{L^2}$ is of order $|X_0|$, and $\|F(w)_x - Dw_{xx}\|^2_{L^2}$ is of order 1. Then for $L_0 = \gamma T + X_0 + C_w$ given in Lemma 5.9, we have the desired upper bound for $\mathcal{J}(A)$ for $|X_0|$ large.

Now we consider the lower bound. We know that (5.13) still holds for $|X_0|$ large and with C_w given in Lemma 5.9. For $|X_0|$ large, $(\int [U(x - X_0) - U(x)]dx)^2$ is bounded from below by $C_1' X_0^2$, so

$$\mathcal{J}(A) \overset{l_c\to 0}{\ge} \frac{1}{2\sigma^2}\|\phi_0\|^{-2}_{L^2} T^{-1}(X_0 + \gamma T + C_w)^{-1} C_1' X_0^2, \tag{5.18}$$

which gives the desired lower bound for $\mathcal{J}(A)$ for $|X_0|$ large.

6. Large deviations for discrete conservation laws. Conservation laws can only be solved numerically, in general, so we need to consider space-time discretizations. For the calculation of small probabilities of large deviations, we may pass directly to the calculation of the infimum of the rate function, which we know analytically. First we discretize it in space and time and then use a suitable optimization method to find the approximate minimizer and the approximate value of the rate function. This way of calculating probabilities of large deviations has been carried out previously [9] for different stochastically driven partial differential equations. More involved methods that use adaptive meshes are discussed in [27].

In the next subsection we show briefly that the rate function for large deviations of discrete conservation laws using Euler schemes is, as expected, the corresponding discretization of the continuum rate function.

6.1. Large deviations with Euler schemes. To formulate the discrete problem, we discretize the space and time domains with uniform grids, with $L = x_0 < \cdots < x_M = R$, $M\Delta x = R - L$ and $0 = t_0 < \cdots < t_N = T$, $N\Delta t = T$. Let Q_m^n denote the average of u over the mth cell $[x_{m-1}, x_m]$ (whose center is $x_{m-1/2} = (x_{m-1} + x_m)/2$) at time $n\Delta t$, and Q^n denote the vector $(Q_1^n, \dots, Q_M^n)^T$. The Euler scheme for the conservation law is

$$Q_m^{n+1} = Q_m^n - \frac{\Delta t}{\Delta x}(F^n_{m+\frac{1}{2}} - F^n_{m-\frac{1}{2}}) + D\frac{\Delta t}{(\Delta x)^2}(Q^n_{m+1} - 2Q^n_m + Q^n_{m-1}) + \varepsilon \Delta W^n_m \tag{6.1}$$

for $m = 2, \dots, M-1$, $n = 0, \dots, N-1$, where $F^n_{m\pm 1/2}$ are numerical fluxes constructed by standard finite volume methods such as Godunov or local-Lax–Friedrichs (LLF) and possibly with higher order ENO (essentially non-oscillatory) reconstructions. The initial conditions $(Q_m^0)_{m=1,\dots,M}$ are given by (6.4), the boundary conditions $(Q_1^n)_{n=1,\dots,N}$ and $(Q_M^n)_{n=1,\dots,N}$ are given by (6.7), and the fluxes are given explicitly in (6.6), in the next subsection, for Burgers' equation. We let

$$b(Q^n) = (b_2(Q^n), \dots, b_{M-1}(Q^n))^T,$$

where

$$b_m(Q^n) = -\frac{1}{\Delta x}(F^n_{m+\frac{1}{2}} - F^n_{m-\frac{1}{2}}) + D\frac{1}{(\Delta x)^2}(Q^n_{m+1} - 2Q^n_m + Q^n_{m-1}).$$

Let $(\Delta W_m^n)_{m=2,\dots,M-1,n=1,\dots,N}$ be Gaussian random variables with mean 0 and covariance

$$\mathbb{E}(\Delta W^{n_1}_{m_1} \Delta W^{n_2}_{m_2}) = \begin{cases} \frac{\Delta t}{\Delta x} C_{m_1 m_2}, & n_1 = n_2, \\ 0 & \text{otherwise.} \end{cases}$$

The matrix $C = (C_{ij})_{i,j=2,\dots,M-1}$ is symmetric and nonnegative definite. For simplicity, we assume that C is positive definite, and then $C = \Phi\Phi^T$ for an invertible matrix Φ.

By the Markov property, the joint density function $f_{(Q^0,\dots,Q^N)}$ is

$$f_{(Q^0,\dots,Q^N)}(q^1, \dots, q^N) = \prod_{n=0}^{N-1} f_{Q^{n+1}|Q^n}(q^{n+1}; q^n),$$

where

$$\begin{aligned} &f_{Q^{n+1}|Q^n}(q^{n+1}; q^n) \\ &\quad = \frac{1}{Z} \exp\left[-\frac{\Delta t \Delta x}{2\varepsilon^2} \left(\frac{1}{\Delta t}(q^{n+1} - q^n) - b(q^n) \right)^T C^{-1} \left(\frac{1}{\Delta t}(q^{n+1} - q^n) - b(q^n) \right) \right] \\ &\quad = \frac{1}{Z} \exp\left[-\frac{\Delta t \Delta x}{2\varepsilon^2} \left\| \Phi^{-1} \left(\frac{1}{\Delta t}(q^{n+1} - q^n) - b(q^n) \right) \right\|_2^2 \right], \end{aligned}$$

where $Z^2 = (2\pi\varepsilon^2 \Delta t/\Delta x)^{M-2} \det C$ and the l^2-norm is here

$$\|q\|_2^2 = \sum_{j=2}^{M-1} q_j^2.$$

The exact probability that $Q = (Q^1, \dots, Q^N) \in A$ with the given Q^0 is therefore

$$\mathbb{P}(Q \in A) = \int_A \frac{1}{Z^N} \exp\left[-\frac{\Delta t \Delta x}{2\varepsilon^2} \sum_{n=0}^{N-1} \left\| \Phi^{-1}\left(\frac{1}{\Delta t}(q^{n+1} - q^n) - b(q^n) \right) \right\|_2^2 \right] dq^1 \cdots dq^N.$$

Because the problem is finite-dimensional, the LDP is a form of Laplace's method for the asymptotic evaluation of integrals [25] and the rate function for $q = (q^1, \dots, q^N)$ is

$$I(q) = \frac{\Delta t \Delta x}{2} \sum_{n=0}^{N-1} \left\| \Phi^{-1}\left(\frac{1}{\Delta t}(q^{n+1} - q^n) - b(q^n) \right) \right\|_2^2. \tag{6.2}$$

We observe that as Δx, $\Delta t \to 0$ and $N \to \infty$ with $N\Delta t = T$, if $q_j^n = u(n\Delta t, j\Delta x)$ with u smooth, then $I(q)$ converges to $\frac{1}{2}\iint \left(\Phi^{-1}(u_t + F(u)_x - (Du_x)_x)\right)^2 dxdt$, and the rate function of the discrete conservation law (6.2) converges to the rate function of the continuous conservation law (3.3). Note that, in general, the LDP of the discrete conservation law does not automatically imply the LDP for the continuous conservation law. To prove the LDP of the continuous conservation law from the discrete LDP, the limit in ε and the limit in Δt, Δx have to be interchangeable. In the language of large deviations, such interchangeability is called exponential equivalence [7]. In our paper we have established the LDP of the continuous conservation law and the LDP of the discrete conservation law separately, and we have observed that the rate function (6.2) of the discrete LDP is a discretization of the rate function of the continuous LDP.

6.2. Numerical calculation of rate functions for changes in traveling waves. We carry out numerical calculations with Burgers' equation as a simple but representative conservation law. For convenience, given a traveling wave U_0 with the speed γ, we consider Burgers' equation in moving coordinates:

$$u_t^\varepsilon + \left(\frac{1}{2}(u^\varepsilon - \gamma)^2 \right)_x = (Du_x^\varepsilon)_x + \varepsilon \dot{W}. \tag{6.3}$$

Thus U_0 is a traveling wave of (6.3) with speed zero. We are interested in the rare event that $u^\varepsilon(0, x) = U_0(x)$ and $u^\varepsilon(T, x) = U_T(x)$ because of $\varepsilon\dot{W}$, where U_0 and U_T are two traveling waves but $U_T(\cdot) \neq U_0(\cdot - \gamma T)$.

We therefore minimize the rate function (6.2) to compute the asymptotic probability. The initial conditions

$$q_m^0 = U_0(x_{m-1/2}) \quad \forall m = 1, \dots, M \tag{6.4}$$

and the terminal conditions

$$q_m^N = U_T(x_{m-1/2}) \quad \forall m = 1, \dots, M \tag{6.5}$$

are simply the values of the functions at the centers of the cells. For $m = 2, \dots, M-1$ we let $F_{m\pm1/2}^n$ be the numerical fluxes for $(u-\gamma)^2/2$ at $x_{m\pm1/2}$. We use Godunov's method [18] to construct $F_{m\pm1/2}^n$:

$$F_{m-1/2}^n = \begin{cases} \min_{q_{m-1}^n \le q \le q_m^n} \frac{1}{2}(q-\gamma)^2, & q_{m-1}^n \le q_m^n, \\ \max_{q_m^n \le q \le q_{m-1}^n} \frac{1}{2}(q-\gamma)^2, & q_m^n \le q_{m-1}^n. \end{cases} \tag{6.6}$$

When the final traveling wave solution U_T is the shifted initial profile $U_0(x - X_0)$ for some X_0, any $u \in \mathcal{E}^1$ have $\lim_{x\to\pm\infty} u(x) = u_\pm = \lim_{x\to\pm\infty} U_0(x)$ by Proposition 3.2. Then we impose the boundary conditions

$$q_1^n = u_- \text{ and } q_M^n = u_+ \quad \forall n = 1, \dots, N. \tag{6.7}$$

Equation (6.7) will be a good approximation for the boundary conditions at infinity if the spatial domain $[L, R]$ is sufficiently large. We will discuss the boundary conditions imposed in the other cases later on.

The covariance matrix of the random coefficients $(\Delta W_i^n)_{i=2,\dots,M-1}$ is set to be equal to $C_{ij} = \delta_{ij}$ and thus $\Phi = I_{M-2\times M-2}$. In other words, $(\Delta W_i^n)_{i=2,\dots,M-1}$ are i.i.d. Gaussian random variables. Since the discrete problem is finite dimensional, Φ is clearly a Hilbert–Schmidt matrix.

The objective is to minimize the rate function (6.2) with the initial, terminal, and boundary conditions (6.4), (6.5), and (6.7). The values of $\{q_m^0, q_m^N\}_{m=1}^M$ and $\{q_1^n, q_M^n\}_{n=1}^N$ are decided by (6.4), (6.5), and (6.7) and can be put into the rate function (6.2). Then we can minimize $I(q)$ over the rest of variables $\{q_m^n, 1 \le n \le N-1, 2 \le m \le M-1\}$ by unconstrained optimization methods. The BFGS quasi-Newton method [20] is our optimization algorithm.

An important issue in numerical optimization is that any gradient-based method, such as the BFGS method that we use, gives only a local optimum unless the objective function is convex. In our case, it is not clear that the discrete rate function (6.2) is convex or not. However, based on our numerical simulations we note the following.

1. Our numerical results of the most probable shifted profiles are consistent with the analytical predictions (5.5).
2. Instead of using a good initial guess for the minimizer in the numerical optimization, we have also numerically verified that completely random initial guesses give essentially the same result.
3. We have checked numerically to see if the rate function (6.2) is convex. We randomly pick two close by test paths to see if the midpoint convexity of (6.2) is satisfied. We find that in 99.98% out of 10^6 pairs the discrete rate function passes the convexity test. It is not known what causes the 0.02% failures. We conclude that based on numerical calculations (6.2) is essentially convex, if it is not fully convex. This explains the observed robustness of the numerical optimization.

6.3. The numerical setup. We use different U_0 and U_T to calculate probabilities of several rare events. Our main interest is the anomalous wave displacement, which is theoretically analyzed in the previous sections. The other cases are the wave speed change, the transition from a weak shock to a strong shock, and the transition from a strong shock to a weak shock. We have not carried out an analysis of the last three cases. However, the probabilities of these rare events can be calculated numerically and show how unlikely such events are compared to anomalous wave displacement.

In each configuration we consider the high viscosity case ($D = 1$) and the low viscosity case ($D = 0.01$). $T = 1$, $\Delta x = 0.2$, and $\Delta t = 0.02$ in all simulations and the linear interpolation of U_0 and U_T is the initial guess in the numerical optimization. As we noted before, a random initial guess gives essentially the same result, but we use the linear interpolation to speed up the optimization.

6.4. Anomalous wave displacements. In this case, we let U_0 be the traveling wave solution for Burgers' equation (6.3), and $U_T = U_0(x - X_0)$ represent a shifted

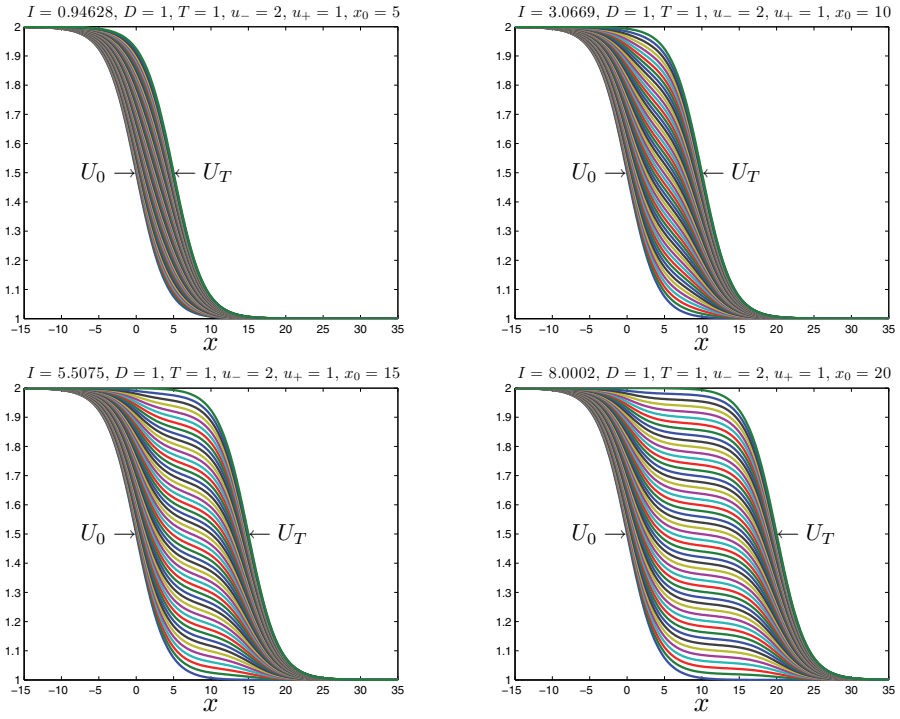

FIG. 1. *The most probable paths and their values of I in* (6.2) *in the case that $U_T(x) = U_0(x - X_0)$ for different X_0 with $D = 1$. In each figure, we plot the curves indicating the most probable path at time $0, \Delta t, 2\Delta t, \ldots, T = 1$. We can see that for X_0 small, the most probable path is close to the linear shift traveling wave and for X_0 large, the most probable path looks like the linear interpolation of U_0 and U_T. These observations are consistent with the test function v and w we use for the upper bounds in sections* 5.2 *and* 5.3.

traveling wave. This is the discrete version of (5.3), and we find that the numerical results are consistent with the analytical result (5.5).

We first consider Figures 1 and 2 with $D = 1$. The most probable path is close to the linear shift when X_0 is small and it looks like the linear interpolation when X_0 is large. This also motivates us to choose the test functions v and w for the upper bounds of $\mathcal{J}(A)$ in sections 5.2 and 5.3. Further, Figure 2 shows that the optimal I is quadratic near $X_0 = 0$, is linear for X_0 large, and is of order $1/T$ in T. These observations confirm the theoretical predictions (5.5).

We consider next Figures 3 and 4 with $D = 0.01$. As the transition regions are very narrow and separate very quickly in the low viscosity case, the most probable path is nearly the linear interpolation, and therefore the I versus X_0 plot is almost linear except around $X_0 = 0$. Moreover, the optimal rate function is still of order $1/T$ in T, which is also seen in the analysis (5.5).

6.5. Change of wave speeds. In this and the following subsections we consider the events in which $u(0, x) = U_0(x)$ and $u(T, x) = U_T(x) := U_1(x)$, where $U_1(x)$ is a traveling wave with $u_{1\pm} := \lim_{x\to\pm\infty} U_1(x)$ which are different from $u_\pm = \lim_{x\to\pm\infty} U_0(x)$. The fact that $u_{1\pm} \neq u_\pm$ violates the conclusion of Proposition 3.2 and thus the probabilities of such events are zero. Nevertheless, we can consider the minimization of the rate function over such events in a broader sense: we

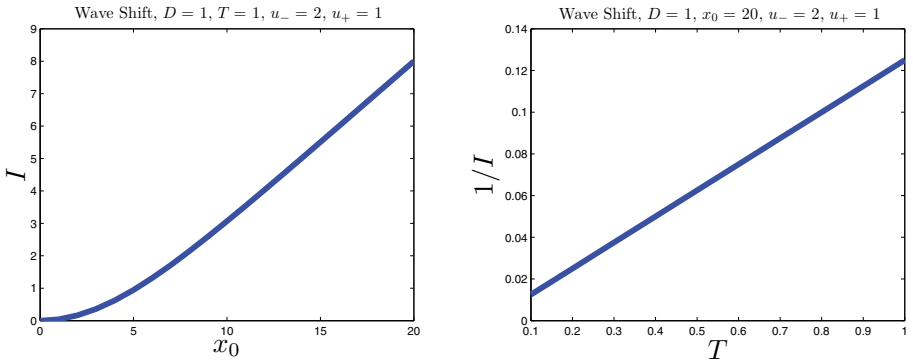

FIG. 2. *Left: The optimal values of I in (6.2) versus X_0 in the case that $U_T(x) = U_0(x - X_0)$ for $X_0 = 0, 1, 2, \ldots, 20$ with the same setting in Figure 1. We see that the curve is quadratic for X_0 small and linear for X_0 large (see (5.5)). Right: The reciprocal of the optimal values of I in (6.2) versus T in the case that $U_T(x) = U_0(x - X_0)$ for $X_0 = 20$ and $T = 0.1, 0.2, \ldots, 1$ with the same setting in Figure 1. We see that the curve is linear in T, which means the optimal I is of $\Theta(1/T)$ (see also (5.5)).*

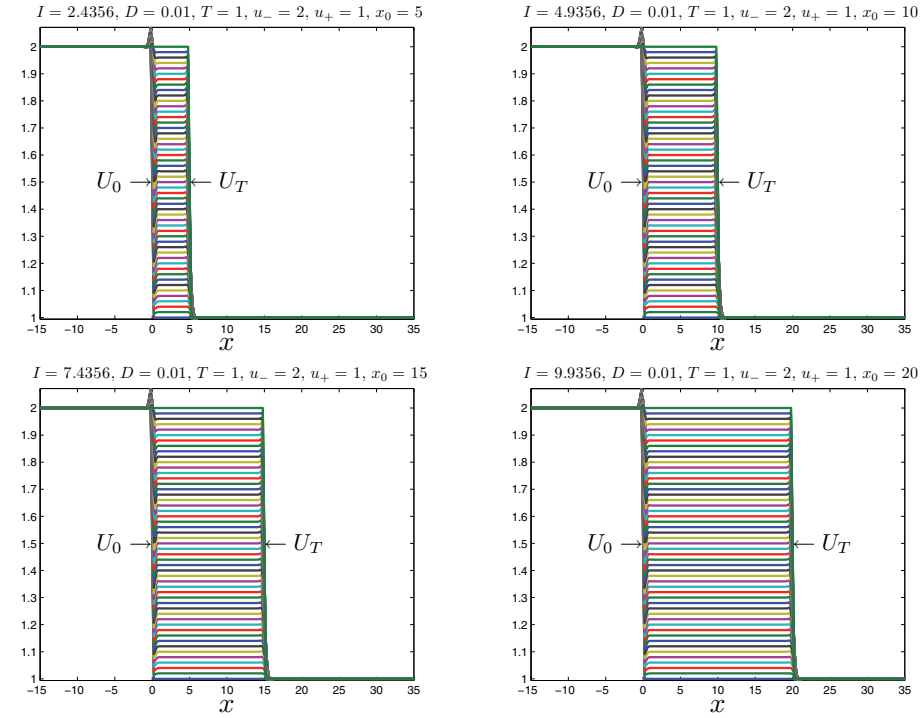

FIG. 3. *The most probable paths and their values of I in (6.2) in the case that $U_T(x) = U_0(x - X_0)$ for different X_0 with $D = 0.01$. In each figure, we plot the curves indicating the most probable path at time $0, \Delta t, 2\Delta t, \ldots, T = 1$. Different from Figure 1, as the transition region is very small, we see nearly no shifted part but only the linear interpolation.*

regard the rate function $I(q)$ in (6.2) as an energy functional of the path q, and the results in subsections 6.5 and 6.6 describe the minimizers of the energy functional I.

In this section we choose $u_{1\pm}$ such that $\gamma_1 = \frac{F(u_{1+}) - F(u_{1-})}{u_{1+} - u_{1-}}$ is different from

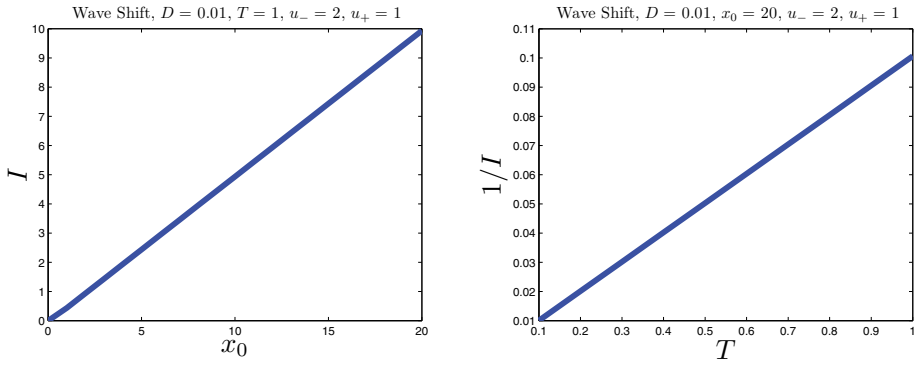

FIG. 4. *Left: The optimal values of I in (6.2) versus X_0 in the case that $U_T(x) = U_0(x - X_0)$ for $X_0 = 0, 1, 2, \ldots, 20$ with the same setting in Figure 3. As what we see in Figure 3, the most probable path is the linear interpolation and thus the curves is almost linear except a small perturbation around $X_0 = 0$. Right: The reciprocal of the optimal values of I in (6.2) versus T in the case that $U_T(x) = U_0(x - X_0)$ for $T = 0.1, 0.2, \ldots, 1$ with the same setting in Figure 3. We see that the curve is linear in T, which means the optimal I is $\mathcal{O}(1/T)$.*

$\gamma = \frac{F(u_+) - F(u_-)}{u_+ - u_-}$. The boundary conditions are more delicate to implement in this case. We know that U_0 and U_T are very close to constants when L and R are far from their transition regions. In this case, the LDP implies that the most probable path around the boundaries should be very close to the linear interpolations of U_0 and U_T in time. Therefore we let

$$q_1^n = \Big(1 - \frac{n}{N}\Big) q_1^0 + \frac{n}{N} q_1^N, \quad q_M^n = \Big(1 - \frac{n}{N}\Big) q_M^0 + \frac{n}{N} q_M^N.$$

It is possible not to set the boundary conditions and to optimize the boundary cells as well. Our numerical simulations show, however, that the solution in both cases is basically the same except for some oscillations near the boundaries. The oscillations come from the inappropriate discretization at the boundaries, and this is a limitation of the numerical discretization. We can have a few extra boundary conditions to reduce the unwanted oscillations at the boundaries. For example, we can additionally set

$$q_2^n = \Big(1 - \frac{n}{N}\Big) q_2^0 + \frac{n}{N} q_2^N, \quad q_{M-1}^n = \Big(1 - \frac{n}{N}\Big) q_{M-1}^0 + \frac{n}{N} q_{M-1}^N.$$

The results are shown in Figure 5. We let $\gamma = 0$ and $\gamma_1 = 3.5$ while we keep $u_- - u_+ = u_{1-} - u_{1+} = 1$ to indicate that U_0 and U_T have roughly the same transition magnitude. We see that the values of the rate function are much larger than that of the anomalous wave displacements. This means that it is very unlikely to have changes of wave speeds compared to wave displacements.

6.6. Weak shocks to strong shocks and strong shocks to weak shocks. We also consider the case that U_0 is a weak (strong) shock while U_T is a strong (weak) shock. By a strong shock we mean that the difference between u_- and u_+ is large. This case is also not in the range of our analytical framework, but we can still compute the rate function after we impose the suitable boundary conditions. We use the same

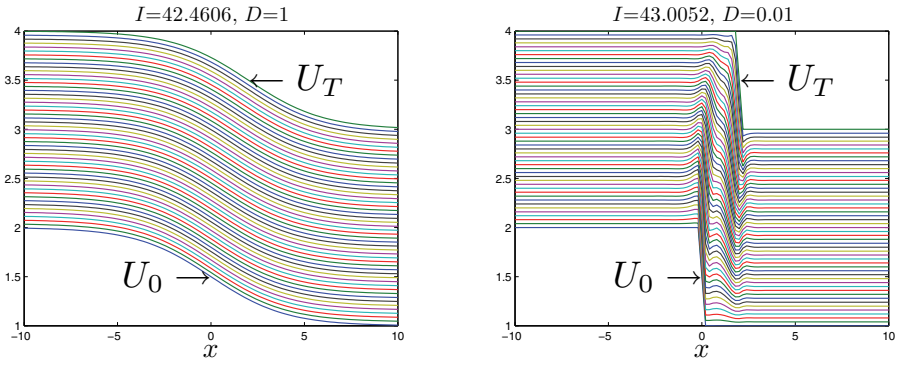

FIG. 5. *The most probable paths and their values of I in* (6.2) *in the case that U_0 and U_T have different speeds. Here we let $\gamma = 1.5$ in Burgers' equation* (6.3)*, and therefore the speed of U_0 is zero and the speed of U_T is* 3.5*. The differences between the left and right boundary values are kept the same so that U_0 and U_T have the same transition magnitude. In each figure, we plot the curves indicating the most probable path at time $0, \Delta t, 2\Delta t, \ldots, T = 1$.*

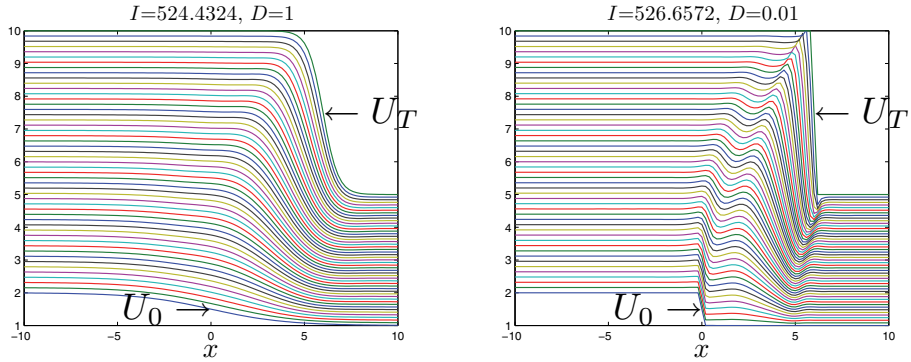

FIG. 6. *The most probable paths and their values of I in* (6.2) *in the case that U_0 is a weak shock while U_T is a strong shock. Here we let U_0 have a small difference in the boundary values while U_T has a large one. In each figure, we plot the curves indicating the most probable path at time $0, \Delta t, 2\Delta t, \ldots, T = 1$.*

boundary conditions as the ones in the previous subsection:

$$q_1^n = \Big(1 - \frac{n}{N}\Big)q_1^0 + \frac{n}{N}q_1^N, \quad q_M^n = \Big(1 - \frac{n}{N}\Big)q_M^0 + \frac{n}{N}q_M^N,$$
$$q_2^n = \Big(1 - \frac{n}{N}\Big)q_2^0 + \frac{n}{N}q_2^N, \quad q_{M-1}^n = \Big(1 - \frac{n}{N}\Big)q_{M-1}^0 + \frac{n}{N}q_{M-1}^N.$$

From Figures 6 and 7 we see that the most probable path of weak to strong and the one of strong to weak are significantly different, even if the reference strong and weak shocks are fixed. We note the very large value of the rate function compared to anomalous displacement, and how it depends on D. This confirms quantitatively the expectation that shock profiles are very stable and they are not easily perturbed except for displacements.

7. Direct numerical simulations with importance sampling. The large deviations probabilities calculated in sections 5 and 6 are only the exponential decay rates of the probabilities but not the actual probabilities. In this section we use Monte

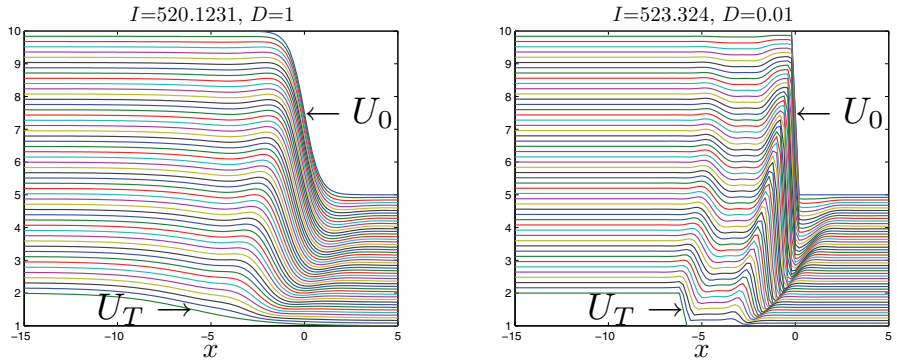

FIG. 7. *The most probable paths and their values of I in* (6.2) *in the case that U_0 is a strong shock while U_T is a weak shock. Here we let U_0 have a large difference in the boundary values while U_T has a small one. In each figure, we plot the curves indicating the most probable path at time $0, \Delta t, 2\Delta t, \dots, T = 1$.*

Carlo methods to compute the actual probabilities numerically.

7.1. Burgers' equation with spatially correlated random perturbations. We reformulate the discretized problem for Burgers' equation when we have spatially correlated random perturbations. Given a traveling wave solution $U_0(x-\gamma t)$ of Burgers' equation with $\lim_{x\to\pm\infty} U_0(x) = u_\pm$, we transform to moving coordinates

$$u_t^\varepsilon + \Big(\frac{1}{2}(u^\varepsilon - \gamma)^2\Big)_x = (Du_x^\varepsilon)_x + \varepsilon \dot{W}. \tag{7.1}$$

Then $U_0(x)$ is a stationary traveling wave of (7.1). The rare event we consider is

$$A_\delta = \{u \in \mathcal{E}^1 \text{ such that } u(0,\cdot) = U_0,\ \|u(T,\cdot) - U_0(\cdot - X_0)\|_{L^2} \le \delta\}.$$

Although for a discrete conservation law it is also possible to consider the other cases in section 6, their probabilities are too small to compute by the basic Monte Carlo method so we omit them.

To compute $\mathbb{P}(u \in A_\delta)$ numerically, we discretize the space and time domains uniformly as in section 6.1: $L = x_0 < \cdots < x_M = R$, $M\Delta x = R - L$, and $0 = t_0 < \cdots < t_N = T$, $N\Delta t = T$. Here Q_m^n denotes the average of u over the mth cell at time $n\Delta t$ and evolves by the Euler method:

$$Q_m^{n+1} = Q_m^n - \frac{\Delta t}{\Delta x}(F^n_{m+\frac12} - F^n_{m-\frac12}) + D\frac{\Delta t}{(\Delta x)^2}(Q^n_{m+1} - 2Q^n_m + Q^n_{m-1}) + \varepsilon \Delta W_m^n \tag{7.2}$$

for $m = 2, \dots, M-1$, $n = 0, \dots, N-1$, where $(F^n_{m-1/2})_{m=2,\dots,M}$ are numerical fluxes of $(u-\gamma)^2/2$ constructed from $(Q_m^n)_{m=1,\dots,M}$ by Godunov's method (6.6) and $(\Delta W_m^n)_{m=2,\dots,M-1,n=1,\dots,N}$ are Gaussian random variables with mean 0 and covariance

$$\mathbb{E}(\Delta W^{n_1}_{m_1} \Delta W^{n_2}_{m_2}) = \begin{cases} \frac{\Delta t}{\Delta x} C_{m_1 m_2}, & n_1 = n_2, \\ 0 & \text{otherwise.} \end{cases}$$

In order to make the problem more realistic, we assume that the variables ΔW_m^n are spatially correlated: $C_{m_1 m_2} = \sigma^2 \exp(-\frac{1}{l_c}|x_{m_1} - x_{m_2}|)$ and $C = (C_{ij})_{i,j=2,\dots,M-1} = \Phi\Phi^T$. Finally we impose the initial and boundary conditions: $Q_m^0 = U_0(x_{m-1/2})$, $Q_1^n = u_-$, and $Q_M^n = u_+$.

7.2. Introduction to importance sampling. To estimate $\mathbb{P}(u \in A_\delta)$, we may use the basic Monte Carlo method. The Monte Carlo strategy is as follows. We generate K independent samples $\Delta W^{(k)} = (\Delta W_m^{n,(k)})_{m=2,\dots,M-1,n=1,\dots,N}$, $k = 1,\dots,K$, of the Gaussian vector $(\Delta W_m^n)_{m=2,\dots,M-1,n=1,\dots,N}$, which give K independent samples $Q^{(k)} = (Q_m^{n,(k)})_{m=1,\dots,M,n=1,\dots,N}$, $k = 1,\dots,K$, of the random vector $Q = (Q_m^n)_{m=1,\dots,M,n=1,\dots,N}$. The basic Monte Carlo estimator is

$$\hat{P}^{MC} = \frac{1}{K}\sum_{k=1}^{K} 1_{A_\delta}(Q^{(k)}), \tag{7.3}$$

where

$$Q \in A_\delta \text{ if and only if } \Delta x \sum_{m=1}^{M} [Q_m^N - U_0(x_{m-\frac{1}{2}} - X_0)]^2 \le \delta^2. \tag{7.4}$$

In other words, $\hat{P}^{MC}$ is the empirical frequency that $Q^{(k)} \in A_\delta$. It is an unbiased estimator $\mathbb{E}[\hat{P}^{MC}] = \mathbb{P}(Q \in A_\delta)$. By the law of large numbers, it is strongly convergent $\hat{P}^{MC} \to \mathbb{P}(Q \in A_\delta)$ almost surely as $K \to \infty$. Its variance is given by

$$\mathbf{Var}(\hat{P}^{MC}) = \frac{1}{K}\mathbf{Var}(1_{A_\delta}(Q)) = \frac{1}{K}\big(\mathbb{P}(Q \in A_\delta) - \mathbb{P}(Q \in A_\delta)^2\big).$$

In order to have a meaningful estimation, the standard deviation of the estimator and $\mathbb{P}(Q \in A_\delta)$ should be of the same order. Namely, the relative error

$$\frac{\mathbf{Var}^{\frac{1}{2}}(\hat{P}^{MC})}{\mathbb{P}(Q \in A_\delta)} = \frac{1}{\sqrt{K}}\left(\frac{1}{\mathbb{P}(Q \in A_\delta)} - 1\right)^{\frac{1}{2}}$$

should be of order one (or smaller). This means that the number K of Monte Carlo samples should be at least of the order of the reciprocal of the probability $\mathbb{P}(Q \in A_\delta)$. We note that for ε small, $\mathbb{P}(Q \in A_\delta)$ decreases exponentially in ε^2 and so K should be increased exponentially; the exponential growth of K makes the basic Monte Carlo method computationally infeasible.

The well established way to overcome the difficulty of calculating rare event probabilities is to use importance sampling. The problem with basic Monte Carlo is that for small ε there are very few samples in A_δ under the original measure $\mathbb{P}$ so the estimator is inaccurate. In importance sampling we change the original measure so that there is a significant fraction of Q^k in A_δ under this new measure $\mathbb{Q}$, even for small ε. Since we use the biased measure $\mathbb{Q}$ to generate $Q^{(k)}$, it is necessary to weight the simulation outputs in order to get an unbiased estimator of $\mathbb{P}(Q \in A_\delta)$. The correct weight is the likelihood ratio since we have

$$\mathbb{P}(Q \in A_\delta) = \mathbb{E}_{\mathbb{P}}[1_{A_\delta}(Q)] = \mathbb{E}_{\mathbb{Q}}\left[1_{A_\delta}(Q)\frac{d\mathbb{P}}{d\mathbb{Q}}(Q)\right].$$

Then the importance sampling estimator is

$$\hat{P}^{IS} = \frac{1}{K}\sum_{k=1}^{K} 1_{A_\delta}(Q^{(k)})\frac{d\mathbb{P}}{d\mathbb{Q}}(Q^{(k)}), \tag{7.5}$$

where $Q^{(k)}$ are generated under $\mathbb{Q}$. The estimator $\hat{P}^{IS}$ is unbiased $\mathbb{E}_{\mathbb{Q}}[\hat{P}^{IS}] = \mathbb{P}(Q \in A_\delta)$, and its variance is

$$\mathbf{Var}_{\mathbb{Q}}(\hat{P}^{IS}) = \frac{1}{K}\mathbf{Var}_{\mathbb{Q}}\left(1_{A_\delta}(Q)\frac{d\mathbb{P}}{d\mathbb{Q}}(Q)\right).$$

The main issue in importance sampling is how to choose a good $\mathbb{Q}$ to have a low $\mathbf{Var}_{\mathbb{Q}}[1_{A_\delta}(Q)\frac{d\mathbb{P}}{d\mathbb{Q}}(Q)]$. In many cases (see for example, [23, 22, 3, 21]), it can be shown that the change of measure suggested by the most probable path of the LDP is asymptotically optimal as $\varepsilon \to 0$. However, it is also well known that in some cases (see [13]), the estimator by this strategy is worse than the basic Monte Carlo and may even have infinite variance. However, because the rare event A_δ is convex and the discrete rate function I is (numerically tested) essentially convex, the importance sampling estimator by this strategy is expected to be asymptotically optimal and we will see that it indeed works very well.

7.3. Importance sampling based on the most probable path. In this subsection we implement the importance sampling by using a biased distribution centered on the most probable path obtained in section 6. From section 6.1, $\bar{Q} := \arg\inf_{Q\in A_\delta} I(Q)$ is the most probable path as $\varepsilon \to 0$ by the large deviation principle. We choose $\bar{h} = (\bar{h}^n)_{n=1,\dots,N} = (\bar{h}^n_m)_{m=2,\dots,M-1,n=1,\dots,N}$ such that

$$\begin{aligned}\bar{Q}^{n+1}_m = \bar{Q}^n_m - \frac{\Delta t}{\Delta x}(\bar{F}^n_{m+\frac{1}{2}} - \bar{F}^n_{m-\frac{1}{2}}) \\ + D\frac{\Delta t}{(\Delta x)^2}(\bar{Q}^n_{m+1} - 2\bar{Q}^n_m + \bar{Q}^n_{m-1}) + (\Phi\bar{h}^n)_m\end{aligned} \tag{7.6}$$

for $m = 2,\dots,M-1$, $n = 1,\dots,N$, where $(\bar{F}^n_{m-1/2})_{m=2,\dots,M}$ are numerical fluxes of $(u-\gamma)^2/2$ constructed from $(\bar{Q}^n_m)_{m=1,\dots,M}$ by Godunov's method (6.6).

Assume that under the probability $\mathbb{Q}$, the vector $\Delta W^n := (\Delta W^n_m)_{m=2,\dots,M-1}$ in (7.2) is multivariate Gaussian $\mathcal{N}(\varepsilon^{-1}\Phi\bar{h}^n, \frac{\Delta t}{\Delta x}C)$ and $\mathbf{Cov}(\Delta W^{n_1}_{m_1}, \Delta W^{n_2}_{m_2}) = 0$ if $n_1 \neq n_2$. Denoting $\Delta W = (\Delta W^n)_{n=1,\dots,N}$, the likelihood ratio $d\mathbb{P}/d\mathbb{Q}$ can be computed explicitly:

$$\frac{d\mathbb{P}}{d\mathbb{Q}}(\Delta W) = \exp\left(-\frac{1}{2}\sum_{n=1}^{N}[\|\Phi^{-1}\Delta W^n\|_2^2 - \|\Phi^{-1}\Delta W^n - \varepsilon^{-1}\bar{h}^n\|_2^2\right). \tag{7.7}$$

Note that under $\mathbb{Q}$, (7.2) can be written as

$$Q^{n+1}_m = Q^n_m - \frac{\Delta t}{\Delta x}(F^n_{m+\frac{1}{2}} - F^n_{m-\frac{1}{2}}) + D\frac{\Delta t}{(\Delta x)^2}(Q^n_{m+1} - 2Q^n_m + Q^n_{m-1}) + (\Phi\bar{h}^n)_m + \varepsilon\Delta\tilde{W}^n_m, \tag{7.8}$$

where $\Delta\tilde{W} = (\Delta\tilde{W}^n_m)_{m=2,\dots,M-1,n=1,\dots,N}$ is zero-mean, Gaussian with the spatial covariance $\frac{\Delta t}{\Delta x}C$, and white in time. Then (7.7) can be written as

$$\frac{d\mathbb{P}}{d\mathbb{Q}}(\Delta\tilde{W}) = \exp\left(-\frac{1}{2}\sum_{n=1}^{N}\|\Phi^{-1}\Delta\tilde{W}^n + \varepsilon^{-1}\bar{h}^n\|_2^2 - \|\Phi^{-1}\Delta\tilde{W}^n\|_2^2\right). \tag{7.9}$$

In summary, the importance sampling Monte Carlo strategy is implemented as follows:

1. Compute the most probable path $\bar{Q} = \arg\inf_{Q\in A_\delta} I(Q)$ and its residual $\bar{h}$ from (7.6).

TABLE 1
The parameters for the Monte Carlo simulations.

$\Delta x = 0.5$	$\Delta t = 0.05$	$L = -15$	$R = 20$	$T = 1$	$u_- = 2$	$u_+ = 1$
$\gamma = 1.5$	$D = 1$	$x_0 = 5$	$\delta = \sqrt{0.5}$	$\sigma = 1$	$l_c = 5$	$K = 10^4$

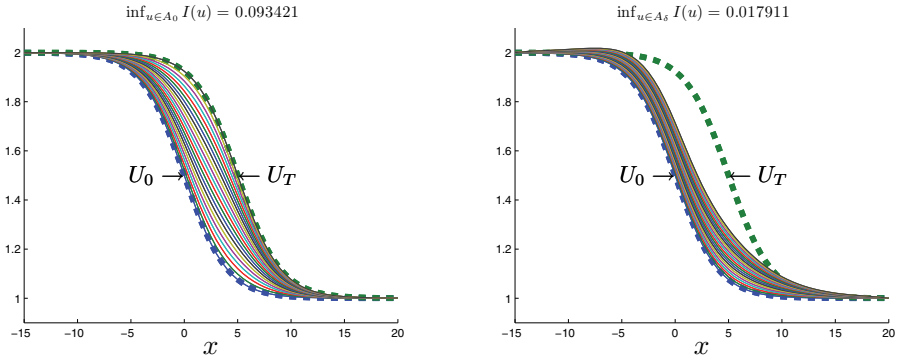

FIG. 8. *The most probable paths for* $\inf_{Q\in A_0} I(Q)$ *and* $\inf_{Q\in A_\delta} I(Q)$. *With the spatially correlated noise,* $\inf_{Q\in A_0} I(Q)$ *is much lower than the one with the spatially white noise (see Figure* 1*).*

2. Sample K independent $\Delta\tilde{W}^{(k)}$ with the zero-mean, Gaussian distribution with the covariance $\frac{\Delta t}{\Delta x}C$ in space and white in time. Compute the corresponding $Q^{(k)}$ by (7.8).
3. The importance sampling estimator is

$$\hat{P}^{IS} = \frac{1}{K}\sum_{k=1}^{K} 1_{A_\delta}(Q^{(k)})\frac{d\mathbb{P}}{d\mathbb{Q}}(\Delta\tilde{W}^{(k)}),$$

where $\frac{d\mathbb{P}}{d\mathbb{Q}}(\Delta\tilde{W})$ is defined in (7.9).

7.4. Simulations with importance sampling. We consider three estimators: the basic Monte estimator $\hat{P}^{MC}$ and two importance sampling estimators: $\hat{P}_0^{IS}$ and $\hat{P}_\delta^{IS}$, where $\hat{P}_0^{IS}$ uses $\bar{h}$ in (7.6) with $\bar{Q} = \arg\inf_{Q\in A_0} I(Q)$ while $\hat{P}_\delta^{IS}$ uses $\bar{h}$ in (7.6) with $\bar{Q} = \arg\inf_{Q\in A_\delta} I(Q)$. The parameters for the simulations are listed in Table 1.

As we mention before, we only test the wave displacement with $X_0 = 5$ and $D = 1$ because the probabilities of the other cases in section 6 are too small for $\hat{P}^{MC}$ to have meaningful samples.

First we find the most probable paths for $\inf_{Q\in A_0} I(Q)$ and $\inf_{Q\in A_\delta} I(Q)$. As before, $\inf_{Q\in A_0} I(Q)$ can be modeled as an unconstrained optimization problem and we solve it by the BFGS method while we solve $\inf_{Q\in A_\delta} I(Q)$ by sequential quadratic programming (SQP) [20].

From Figure 8 we note that $\inf_{Q\in A_0} I(Q)$ is much smaller than the corresponding one with spatially white noise. This is because with the correlated noise, it is easier to have uniform spatial increments. In addition, $\inf_{Q\in A_\delta} I(Q)$ is significantly different from $\inf_{Q\in A_0} I(Q)$ because $\delta = \sqrt{0.5}$ is not very small. we will see that this difference significantly affects the performances of $\hat{P}_0^{IS}$ and $\hat{P}_\delta^{IS}$.

Once $\arg\inf_{Q\in A_0} I(Q)$ and $\arg\inf_{Q\in A_\delta} I(Q)$ are obtained, we can construct $\hat{P}_0^{IS}$ and $\hat{P}_\delta^{IS}$. We estimate $\mathbb{P}(Q \in A_\delta)$ by these three estimators for 100 different ε taken uniformly in $[0.01, 0.2]$. For each ε and estimator, we use $K = 10^4$ samples and plot

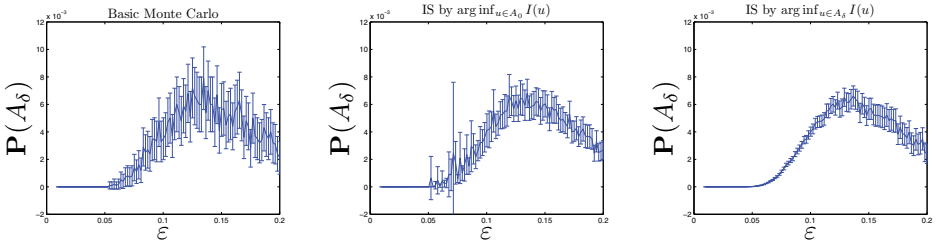

FIG. 9. *The estimated probabilities and their* 99% *confidence intervals by* $\hat{P}^{MC}$, $\hat{P}_0^{IS}$ *and* $\hat{P}_\delta^{IS}$. $\hat{P}_\delta^{IS}$ *has the best performance and* $\hat{P}_0^{IS}$ *also has good performance for* $0.1 \le \varepsilon \le 0.2$. *However* $\hat{P}_0^{IS}$ *works poorly for* $\epsilon < 0.1$.

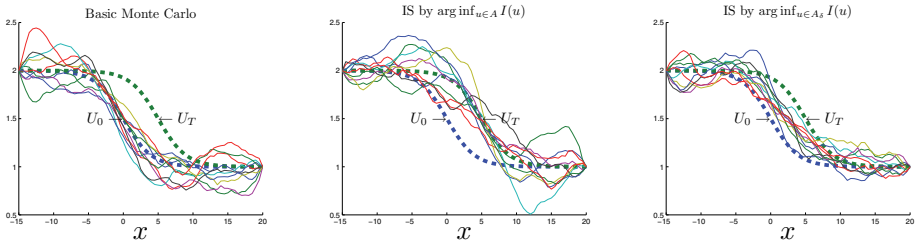

FIG. 10. *Sample paths under these three probability measures with* $\varepsilon = 0.1$.

the estimated probabilities and their 99% confidence intervals in Figure 9 We also plot typical sample paths under three different measures with $\varepsilon = 0.1$ in Figure 10. The (numerical) 99% confidence intervals are defined by $\left[\text{Mean}_K - 2.6\,\text{Std}_K, \text{Mean}_K + 2.6\,\text{Std}_K\right]$, where the (numerical) mean and standard deviation are

$$\text{Mean}_K = \frac{1}{K}\sum_{k=1}^{K} p^{(k)}, \qquad \text{Std}_K^2 = \frac{1}{K}\sum_{k=1}^{K} (p^{(k)})^2 - \text{Mean}_K^2,$$

with $p^{(k)} = \mathbf{1}_{A_\delta}(Q^{(k)})$ for $\hat{P}^{MC}$ and $p^{(k)} = d\mathbb{P}/d\mathbb{Q}(\Delta\tilde{W}^{(k)})\mathbf{1}_{A_\delta}(Q^{(k)})$ for $\hat{P}^{IS}$. We find that $\hat{P}_\delta^{IS}$ has the best performance and $\hat{P}_0^{IS}$ also has the good performance for $0.1 \le \varepsilon \le 0.2$. For $\epsilon < 0.1$, because $\arg\inf_{Q\in A_0} I(Q)$ is not the most probable path, $\hat{P}_0^{IS}$ is even worse than $\hat{P}^{MC}$ due to the inappropriate change of measure. For $\epsilon < 0.1$, because $\arg\inf_{Q\in A_\delta} I(Q)$ is the most probable path, $\hat{P}_\delta^{IS}$ dramatically outperforms $\hat{P}^{MC}$. This shows that large-deviations-driven importance sampling strategies can be efficient to estimate rare event probabilities in the context of perturbed scalar conservation laws.

We plot the estimated probabilities and the relative error, the ratio of the (numerical) standard deviation to the (numerical) mean, in log scale. We can see that the curves of the relative errors in Figure 11 (right) saturate at 100 for small ε. This can be explained as follows. When the estimated probability is dominated by the value $p^{(k_0)}$ of one realization over K, then $p^{(k_0)} = 1$ for $\hat{P}^{MC}$ and $p^{(k_0)} = d\mathbb{P}/d\mathbb{Q}(\Delta\tilde{W}^{(k_0)})$ for $\hat{P}^{IS}$, and the numerical variance is approximately $(p^{(k_0)})^2/K$. Therefore the relative error is about $\sqrt{K} = 100$. Of course when the relative error reaches $\sqrt{K}$, the estimator is determined by a single realization and is not reliable.

8. Conclusion. We have analyzed here the small probabilities of anomalous shock profile displacements due to random perturbations using the theory of large

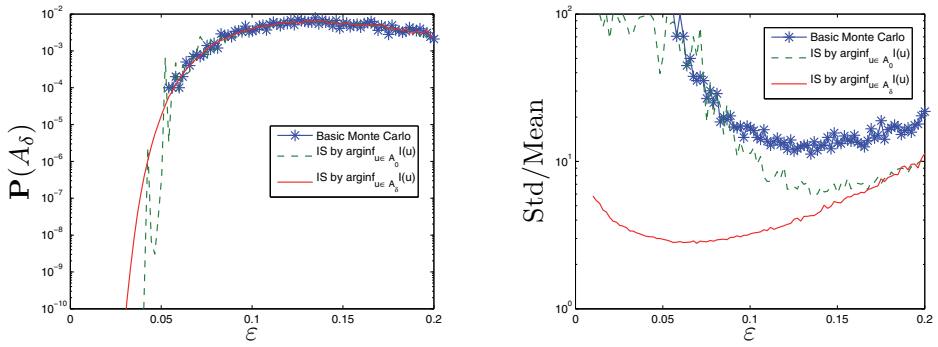

FIG. 11. *The semilog plots of the estimated probabilities (left) and the relative errors (right).*

deviations. We have obtained analytically upper and lower bounds for the exponential rate of decay of these probabilities with respect to ε and we have verified the accuracy of these bounds with numerical simulations. We have also used Monte Carlo simulations with importance sampling based on the analytically known rate function, which is efficient and gives very good results for rare event probabilities.

Appendix A. Proof of Lemma 2.2. Taking a Fourier transform in x we have

$$\hat{Z}(t,\xi) = \int_0^t e^{-D\xi^2(t-s)} d\hat{W}(s,\xi),$$

where $\hat{W}(t,\xi) = \int W(t,x)e^{-i\xi x}dx$ is a complex Gaussian process with mean zero and covariance

$$\mathbb{E}\big[\hat{W}(t,\xi)\overline{\hat{W}(t',\xi')}\big] = t \wedge t'\, \hat{C}(\xi,\xi'),$$

with $\hat{C}(\xi,\xi') = \int e^{-i\xi x + i\xi' x'} C(x,x')dxdx'$. We find that

$$\mathbb{E}[|\hat{Z}(t,\xi)|^2] = \frac{1-e^{-2D\xi^2 t}}{2D\xi^2}\hat{C}(\xi,\xi).$$

On the one hand, the fact that Φ is Hilbert–Schmidt from L^2 into L^2 implies that

$$\int \hat{C}(\xi,\xi)d\xi = 2\pi\|\Phi(\cdot,\cdot)\|^2_{L^2(\mathbb{R}\times\mathbb{R})}$$

is finite. On the other hand we have $(1-e^{-s})/s \le 1$ uniformly with respect to $s \in (0,\infty)$, so we get that for any $t \in [0,T]$

$$\int (1+\xi^2)\mathbb{E}[|\hat{Z}(t,\xi)|^2]d\xi \le 2\pi\Big(T + \frac{1}{2D}\Big)\|\Phi(\cdot,\cdot)\|^2_{L^2(\mathbb{R}\times\mathbb{R})},$$

which is equivalent by Parseval relation to

$$\mathbb{E}[\|Z(t)\|^2_{H^1}] \le \Big(T + \frac{1}{2D}\Big)\|\Phi(\cdot,\cdot)\|^2_{L^2(\mathbb{R}\times\mathbb{R})},$$

which gives that $Z(t) \in L^2([0,T], H^1(\mathbb{R}))$ almost surely. Similarly we find

$$\mathbb{E}\big[|\hat{Z}(t,\xi) - \hat{Z}(t',\xi)|^2\big] = \big[1 - e^{-D\xi^2|t-t'|}\big]\big[2 - e^{-2D\xi^2 t\wedge t'} + e^{-D\xi^2(t+t')}\big]\frac{\hat{C}(\xi,\xi)}{2D\xi^2}$$
$$\leq \big[1 - e^{-D\xi^2|t-t'|}\big]\frac{\hat{C}(\xi,\xi)}{D\xi^2} \leq |t-t'|\hat{C}(\xi,\xi),$$

which gives

$$\mathbb{E}\big[\|Z(t) - Z(t')\|_{L^2}^4\big] = \iint \mathbb{E}\big[|Z(t,x) - Z(t',x)|^2 |Z(t,x') - Z(t',x')|^2\big]dxdx'$$
$$\leq \iint \mathbb{E}\big[|Z(t,x) - Z(t',x)|^4\big]^{1/2}\mathbb{E}\big[|Z(t,x') - Z(t',x')|^4\big]^{1/2}dxdx'$$
$$= 3\iint \mathbb{E}\big[|Z(t,x) - Z(t',x)|^2\big]\mathbb{E}\big[|Z(t,x') - Z(t',x')|^2\big]dxdx'$$
$$= 3\Big(\int \mathbb{E}\big[|Z(t,x) - Z(t',x)|^2\big]dx\Big)^2 \leq 3(2\pi)^2|t-t'|^2\|\Phi(\cdot,\cdot)\|_{L^2(\mathbb{R}\times\mathbb{R})}^4,$$

where we have used the fact that $Z(t,x) - Z(t',x)$ is a Gaussian random variable and Parseval equality. By Kolmogorov's continuity criterion for Hilbert valued stochastic processes [4, Thm. 3.3] we find that that $Z(t) \in \mathcal{C}([0,T], L^2(\mathbb{R}))$ almost surely.

If Φ is Hilbert–Schmidt from L^2 into H^1, then

$$\int \xi^2 \hat{C}(\xi,\xi)d\xi = 2\pi\|\partial_x\Phi(\cdot,\cdot)\|_{L^2(\mathbb{R}\times\mathbb{R})}^2$$

is finite, and we can repeat the same arguments to show the final result.

Appendix B. Proofs in section 3.

B.1. Proof of Proposition 3.2. From Lemma 2.2, $Z(t) = \int_0^t S(t-s)dW(s)$ is in $\mathcal{C}([0,T], H^1(\mathbb{R}))$ almost surely. We want to prove the existence and uniqueness in $\mathcal{E}^1$ of the solution to the equation

$$u(t) = S(t)U - \int_0^t S(t-s)[F(u(s))]_x ds + \varepsilon Z(t), \tag{B.1}$$

where F is a $\mathcal{C}^\infty$ function with $\max\{\|F'\|_{L^\infty}, \|F''\|_{L^\infty}\} \leq C_F < \infty$.

To prove the existence we use the Picard iteration scheme: we define $u^0(t) \equiv U$ and

$$u^{n+1}(t) = S(t)U - \int_0^t S(t-s)[F(u^n(s))]_x ds + \varepsilon Z(t).$$

Therefore,

$$u^{n+1}(t) - U = S(t)U - U - \int_0^t S(t-s)[F'(u^n(s))(u_x^n(s) - U_x)]ds \tag{B.2}$$
$$- \int_0^t S(t-s)[F'(u^n(s))U_x]ds + \varepsilon Z(t).$$

It is easy to see that $u^n(t) - U \in \mathcal{C}([0,T], H^1(\mathbb{R}))$ for all n, and we first show that $\sup_{t\in[0,T]}\|u^n(t) - U\|_{H^1(\mathbb{R})}$ are uniformly bounded in n.

LEMMA B.1. *Let $a_n(t) = \|u^n(t) - U\|_{H^1}$. Then there exists a constant C_u such that $a_n(t) \le C_u$ for all n and $t \in [0,T]$. As a consequence, we can choose sufficiently large C_u such that $\|u_x^n(t)\|_{L^2} \le C_u$ for all n and $t \in [0,T]$.*

Proof. Note that $\sup_{t\in[0,T]} \|S(t)U - U\|_{H^1}$ and $\sup_{t\in[0,T]} \|\varepsilon Z(t)\|_{H^1}$ are bounded. In addition, by [24, Chap. 15, sect. 1], $\|S(t-s)\|_{\mathcal{L}(L^2,H^1)} = \mathcal{O}((t-s)^{-1/2})$; then

$$\begin{aligned}\left\|\int_0^t S(t-s)[F'(u^n(s))U_x]ds\right\|_{H^1} &\le \int_0^t \|S(t-s)\|_{\mathcal{L}(L^2,H^1)}\|F'(u^n(s))U_x\|_{L^2}ds \\ &\le C_F\|U_x\|_{L^2}\int_0^t \|S(t-s)\|_{\mathcal{L}(L^2,H^1)}ds\end{aligned}$$

are uniformly bounded in n, and

$$\begin{aligned}&\left\|\int_0^t S(t-s)[F'(u^n(s))(u_x^n(s) - U_x)]ds\right\|_{H^1} \\ &\quad\le \int_0^t \|S(t-s)\|_{\mathcal{L}(L^2,H^1)}\|F'(u^n(s))(u_x^n(s) - U_x)\|_{L^2}ds \\ &\quad\le C_F\int_0^t \|S(t-s)\|_{\mathcal{L}(L^2,H^1)}\|u_x^n(s) - U_x\|_{L^2}ds \\ &\quad\le C_F\int_0^t \|S(t-s)\|_{\mathcal{L}(L^2,H^1)}\|u^n(s) - U\|_{H^1}ds \\ &\quad\le C_F\left[\int_0^t \|S(t-s)\|^p_{\mathcal{L}(L^2,H^1)}ds\right]^{1/p}\left[\int_0^t \|u^n(s) - U\|^q_{H^1}ds\right]^{1/q}.\end{aligned}$$

By letting $1/p + 1/q = 1$ with $1 < p < 2$ and $q > 2$, we can have the following recursive inequality with a sufficiently large C:

$$a_{n+1}(t) \le \frac{C}{2} + \frac{C}{2}\left[\int_0^t a_n^q(s)ds\right]^{1/q}.$$

By the convexity of $x \mapsto x^q$,

$$a_{n+1}^q(t) \le \left(\frac{C}{2} + \frac{C}{2}\left[\int_0^t a_n^q(s)ds\right]^{1/q}\right)^q \le \frac{C^q}{2} + \frac{C^q}{2}\int_0^t a_n^q(s)ds. \tag{B.3}$$

By noting that $a_0(t) = 0$ and (B.3), it easy to see that $a_n^q(t)$ are uniformly bounded in n and $t \in [0,T]$ and so are $a_n(t)$. ∎

To prove the convergence of $u^n(t) - U$ in $\mathcal{C}([0,T], H^1(\mathbb{R}))$, it suffices to prove that $\sum_{n=0}^\infty \sup_{t\in[0,T]} \|u^{n+1}(t) - u^n(t)\|_{H^1} < \infty$.

LEMMA B.2. *Let $b_n(t) = \|u^n(t) - u^{n-1}(t)\|_{H^1}$. Then $\sum_{n=0}^\infty \sup_{t\in[0,T]} b_n(t) < \infty$. As a consequence, $u^n(t) - U$ converges to $u(t) - U$ in $\mathcal{C}([0,T], H^1(\mathbb{R}))$ as $n \to \infty$ and $u(t)$ solves* (B.1).

Proof. By noting (B.1), Lemma B.1, and $\|\cdot\|_{L^\infty} \le C_S\|\cdot\|_{H^1}$ (the Sobolev embedding), we have

$$\begin{aligned}&\|u^{n+1}(t) - u^n(t)\|_{H^1} \\ &\le \int_0^t \|S(t-s)\|_{\mathcal{L}(L^2,H^1)}\|F'(u^n(s))[u_x^n(s) - u_x^{n-1}(s)]\|_{L^2}ds \\ &\quad + \int_0^t \|S(t-s)\|_{\mathcal{L}(L^2,H^1)}\|u_x^{n-1}(s)[F'(u^n(s)) - F'(u^{n-1}(s))]\|_{L^2}ds\end{aligned}$$

$$
\begin{aligned}
&\le C_F \int_0^t \|S(t-s)\|_{\mathcal{L}(L^2,H^1)} \|u^n(s) - u^{n-1}(s)\|_{H^1} ds \\
&\quad + \int_0^t \|S(t-s)\|_{\mathcal{L}(L^2,H^1)} \|u_x^{n-1}(s)\|_{L^2} \|F'(u^n(s)) - F'(u^{n-1}(s))\|_{L^\infty} ds \\
&\le C_F \int_0^t \|S(t-s)\|_{\mathcal{L}(L^2,H^1)} \|u^n(s) - u^{n-1}(s)\|_{H^1} ds \\
&\quad + C_F C_S C_u \int_0^t \|S(t-s)\|_{\mathcal{L}(L^2,H^1)} \|u^n(s) - u^{n-1}(s)\|_{H^1} ds \\
&\le C_F(1 + C_S C_u) \left[\int_0^t \|S(t-s)\|^p_{\mathcal{L}(L^2,H^1)} ds \right]^{1/p} \left[\int_0^t \|u^n(s) - u^{n-1}(s)\|^q_{H^1} ds \right]^{1/q},
\end{aligned}
$$

where $1/p + 1/q = 1$ with $1 < p < 2$ and $q > 2$. Then there exists a constant C such that

$$
b^q_{n+1}(t) \le C \int_0^t b^q_n(s) ds.
$$

Then $b^q_n(t) \le b^q_0(t) C^n T^n / n!$ so $b_n(t) \le b_0(t)(C^n T^n / n!)^{1/q}$ and it is easy to see that $\sum_{n=0}^\infty \sup_{t\in[0,T]} b_n(t) < \infty$. ∎

Finally we show that (B.1) has a unique solution in $\mathcal{E}^1$.

LEMMA B.3. *If $u, v \in \mathcal{E}^1$ solve* (B.1)*, then* $\sup_{t\in[0,T]} \|u(t) - v(t)\|_{H^1} = 0$.

Proof. Let $u, v \in \mathcal{E}^1$ solve (B.1). Using the same calculations in the proof of Lemma B.2, we get

$$
\|u(t) - v(t)\|_{H^1} \le C_F(1 + C_S C_u) \int_0^t \|S(t-s)\|_{\mathcal{L}(L^2,H^1)} \|u(s) - v(s)\|_{H^1} ds.
$$

By noting that $\|u(0) - v(0)\|_{H^1} = 0$, $\sup_{t\in[0,T]} \|u(t) - v(t)\|_{H^1} = 0$ by Gronwall's inequality. ∎

B.2. Proof of Proposition 3.3. The LDP can be obtained following the strategy of [12, 6]. The first step of the proof uses the LDP for the laws of the stochastic convolution εZ on the space $\mathcal{C}([0,T], H^1(\mathbb{R}))$, where

$$
Z(t) = \int_0^t S(t-s) dW(s).
$$

The laws of εZ are Gaussian measures and the LDP with a good rate function is a consequence of the general result on LDP for centered Gaussian measures on real Banach spaces [8, 12].

The second step is to prove that the mapping $X(t) \mapsto u(t) - U$ is continuous from $\mathcal{C}([0,T], H^1(\mathbb{R}))$ into itself, where

$$
u(t) = S(t)U - \int_0^t S(t-s)(F(u(s)))_x ds + \int_0^t S(t-s) X(s) ds. \tag{B.4}
$$

Then the LDP for u^ε is obtained by the contraction principle [7].

To prove the continuity of the mapping, given two pairs (X, u) and (Y, v) satisfying (B.4), we show that $v - u \to 0$ as $Y \to X$ in $\mathcal{C}([0,T], H^1(\mathbb{R}))$. By using essentially

the same calculations in Lemma B.2,

$$\|v(t)-u(t)\|_{H^1} \le C_F(1+C_SC_u)\int_0^t \|S(t-s)\|_{\mathcal{L}(L^2,H^1)}\|v(s)-u(s)\|_{H^1}ds + \int_0^t \|S(t-s)\|_{\mathcal{L}(H^1,H^1)}\|Y(s)-X(s)\|_{H^1}ds,$$

where $C_u = \sup_{t\in[0,T]}\|u_x(t)\|_{L^2}$. Noting that $\|S(t-s)\|_{\mathcal{L}(H^1,H^1)} = \mathcal{O}(1)$ and $\|S(t-s)\|_{\mathcal{L}(L^2,H^1)} = \mathcal{O}((t-s)^{-1/2})$, and by Gronwall's inequality, there exists a constant C such that for $t\in[0,T]$

$$\|v(t)-u(t)\|_{H^1} \le Ce^{C\int_0^T (T-s)^{-1/2}ds}\int_0^T \|Y(s)-X(s)\|_{H^1}ds. \tag{B.5}$$

Because $\int_0^T (T-s)^{-1/2}ds<\infty$, $v-u\to 0$ as $Y\to X$ in $\mathcal{C}([0,T],H^1(\mathbb{R}))$.

B.3. Proof of Proposition 3.4. For each n, we let $h^n\in L^2([0,T],L^2(\mathbb{R}))$ such that $u=\mathcal{H}[h^n]$ and

$$\frac{1}{2}\int_0^T \|h^n(t,\cdot)\|_{L^2}^2dt - \frac{1}{n} < I(u) \le \frac{1}{2}\int_0^T \|h^n(t,\cdot)\|_{L^2}^2dt.$$

Because $u\neq U(x-\gamma t)$, $I(u)>0$. We let u^n be the mild solution of

$$u^n_t + (F(u^n))_x = (Du^n_x)_x + (1-(nI(u))^{-1})^{1/2}\Phi h^n,$$
$$u^n(0,x) = U(x).$$

Then $u^n - U\in\mathcal{C}([0,T],H^1(\mathbb{R}))$ and

$$I(u^n) \le \frac{1}{2}\int_0^T (1-(nI(u))^{-1})\|h^n(t,\cdot)\|_{L^2}^2dt \le \frac{1}{2}\int_0^T \|h^n(t,\cdot)\|_{L^2}^2dt - \frac{1}{n} < I(u).$$

We show that $u^n-u\to 0$ in $\mathcal{C}([0,T],H^1(\mathbb{R}))$. Let $\alpha^n = (1-(nI(u))^{-1})^{1/2}$ and replace (X,u) and (Y,v) by $(u,\Phi h^n)$ and $(u^n,\alpha^n\Phi h^n)$ respectively in (B.5). We have

$$\begin{aligned}\|u^n(t)-u(t)\|_{H^1} &\le Ce^{C\int_0^T (T-s)^{-1/2}ds}\int_0^T (1-\alpha^n)\|\Phi\|_{\mathcal{L}(L^2,H^1)}\|h^n(s)\|_{L^2}ds\\ &\le 2Ce^{C\int_0^T (T-s)^{-1/2}ds}(1-\alpha^n)\|\Phi\|_{\mathcal{L}(L^2,H^1)}\left(I(u)+\frac{1}{n}\right).\end{aligned}$$

Then $u^n-u\to 0$ in $\mathcal{C}([0,T],H^1(\mathbb{R}))$ as $\alpha^n\to 1$.

Appendix C. Proof of Proposition 4.1. We use the following technical result: If $f\in\mathcal{C}([0,T],H^1(\mathbb{R}))$, then $F(t)=\int_0^t S(t-s)\partial_x f(s)ds$ is such that $\int F(t,x)dx=0$ for any $t\in[0,T]$. Indeed, since $f(s)\in H^1$ and $H^1\subset L^\infty$, we have by integrating by parts

$$F(t,x) = -\int_0^t \frac{1}{\sqrt{2\pi(t-s)}}\int \partial_y\big(e^{-\frac{(x-y)^2}{2(t-s)}}\big)f(s,y)dyds,$$

so that, for any $a<b$

$$\int_a^b F(t,x)dx = \int_0^t \frac{1}{\sqrt{2\pi(t-s)}}\int \big(e^{-\frac{(b-y)^2}{2(t-s)}} - e^{-\frac{(a-y)^2}{2(t-s)}}\big)f(s,y)dyds.$$

Let $\delta > 0$. There exists M such that $\int_{[-M,M]^c} f(s,y)^2 dy < \delta^2$ for any $s \in [0,T]$. We can split the integral in y into two pieces. The integral over $[-M,M]^c$ can be bounded by Cauchy–Schwarz inequality so we get

$$\Big| \int_a^b F(t,x)dx \Big| = \frac{\sqrt{2}}{\sqrt[4]{\pi}}\delta + \int_0^t \frac{1}{\sqrt{2\pi(t-s)}} \int_{-M}^{M} \big(e^{-\frac{(b-y)^2}{2(t-s)}} + e^{-\frac{(a-y)^2}{2(t-s)}}\big) f(s,y)dyds,$$

and therefore, by Lebesgue's theorem,

$$\limsup_{a\to-\infty, b\to+\infty} \Big| \int_a^b F(t,x)dx \Big| \le \frac{\sqrt{2}}{\sqrt[4]{\pi}}\delta.$$

Since δ is arbitrary we get the desired technical result.

We then find by integrating (B.1) in x that the center of $u^\varepsilon(t) - S(t)U$ is equal to the center of $\varepsilon Z(t)$. With the condition that C is in $L^1(\mathbb{R}\times\mathbb{R})$, the center of $Z(t)$ is a Gaussian process with mean zero and covariance given by (4.2), which gives the desired result.

Appendix D. Proofs in section 5.

D.1. Proof of Lemma 5.1. It suffices to show the case that $\delta_n = 1/n$. For each n, let $u_n \in A_{1/n}$, such that $\mathcal{J}(A_{1/n}) \le I(u^n) < \mathcal{J}(A_{1/n}) + 1/n$; $\{I(u^n)\}$ are bounded from above by $\mathcal{J}(A_1)+1 < \infty$. Because I is a good rate function and compactness is equivalent to sequential compactness in $\mathcal{E}^1$, $\{u^n\}$ has a convergent subsequence $\{u^{n_l}\}$ whose limit u^* is in A. As I is lower semicontinuous, then

$$\mathcal{J}(A) \ge \lim_n \mathcal{J}(A_{1/n}) = \lim_l I(u^{n_l}) = \liminf_l I(u^{n_l}) \ge I(u^*) \ge \mathcal{J}(A).$$

D.2. Proof of Lemma 5.4. The values of σ, L_0 and l_c do not affect the result, so in the proof we set $\sigma = L_0 = l_c = 1$ and $\Phi^{l_c}_{L_0} = \Phi$ without loss of generality. $\Phi(x,x')$ is Hilbert–Schmidt from L^2 to H^k if and only if $\partial_x^j \Phi(x,x') \in L^2(\mathbb{R}\times\mathbb{R})$ for $0 \le j \le k$. Taking the two-dimensional Fourier transform on $\partial_x^j\Phi(x,x')$, we have

$$\begin{aligned}
\mathcal{F}_{\xi,\xi'}\{\partial_x^j\Phi(x,x')\} &= \mathcal{F}_\xi\{\mathcal{F}_{\xi'}\{\partial_x^j\Phi(x,x')\}\} = \mathcal{F}_\xi\{\partial_x^j[\phi_0(x)\mathcal{F}_{\xi'}\{\phi_1(x-x')\}]\} \\
&= \mathcal{F}_\xi\{\partial_x^j[\phi_0(x)e^{ix\xi'}\hat\phi_1(-\xi')\}]\} = (i\xi)^j\hat\phi_1(-\xi')\mathcal{F}_\xi\{\phi_0(x)e^{ix\xi'}\} \\
&= (i\xi)^j\hat\phi_1(-\xi')\hat\phi_0(\xi+\xi').
\end{aligned}$$

By a simple calculation it is easy to see that $(i\xi)^j\hat\phi_1(-\xi')\hat\phi_0(\xi+\xi') \in L^2(\mathbb{R}\times\mathbb{R})$ if ϕ_0 and ϕ_1 are both in $H^j(\mathbb{R})$.

D.3. Proof of Lemma 5.5. Because $\phi_0(x) \equiv 1$ on $x \in (-1,1)$ by Assumption 5.2,

$$\begin{aligned}
&\|[\phi_0^{-1}(\cdot/L_0) - 1]U_x(\cdot - (t/T)(\gamma T + X_0))\|_{L^2}^2 \\
&= \int_{-\infty}^{-L_0} + \int_{L_0}^{\infty} \{[\phi_0^{-1}(x/L_0) - 1]U_x(x - (t/T)(\gamma T + X_0))\}^2 dx \\
&= \int_{-\infty}^{-L_0-\frac{t}{T}(\gamma T+X_0)} + \int_{L_0-\frac{t}{T}(\gamma T+X_0)}^{\infty} \Big\{ \Big[\phi_0^{-1}\Big(x/L_0 + \frac{t}{T}(\gamma T + X_0)/L_0\Big) - 1\Big] U_x(x)\Big\}^2 dx.
\end{aligned}$$

Let $L_0 = \gamma T + X_0 + C_v$ with $C_v \ge 1$. Because $\phi_0^{-1}(x) \ge 1$ is monotonically increasing for $x \ge 0$ and $C_v \ge 1$,

$$\Big[\phi_0^{-1}\Big(x/L_0 + \frac{t}{T}(\gamma T + X_0)/L_0\Big) - 1\Big]^2 \le [\phi_0^{-1}(x+1) - 1]^2, \quad x \ge 0.$$

Similarly, because $\phi_0^{-1}(x) \geq 1$ is monotonically decreasing for $x \leq 1$,

$$\Big[\phi_0^{-1}\Big(x/L_0 + \frac{t}{T}(\gamma T + X_0)/L_0\Big) - 1\Big]^2 \leq [\phi_0^{-1}(x) - 1]^2, \quad x \leq 0.$$

In addition, $C_v \leq L_0 - \frac{t}{T}(\gamma T + X_0)$ and $-C_v \geq -L_0 - \frac{t}{T}(\gamma T + X_0)$. Therefore,

$$\begin{aligned}&\|[\phi_0^{-1}(\cdot/L_0) - 1]U_x(\cdot - (t/T)(\gamma T + X_0))\|_{L^2}^2 \\ &\quad \leq \int_{-\infty}^{-C_v} \{[\phi_0^{-1}(x) - 1]U_x(x)\}^2 dx + \int_{C_v}^{\infty} \{[\phi_0^{-1}(x+1) - 1]U_x(x)\}^2 dx.\end{aligned}$$

By noting that $\phi_0^{-1}(x)$ has at most polynomial growth, there exists a uniform $C_v \geq 1$ in X_0 and T for (5.8).

D.4. Proof of Lemma 5.6. It is easy to check from the properties of the traveling wave U that $v_t + F(v)_x - Dv_{xx} \in \mathcal{C}([0,T], \mathcal{S}(\mathbb{R}))$. By (5.4) and (5.9), we have

$$v_t + F(v)_x - Dv_{xx} = \sigma\phi_0\Big(\frac{x}{L_0}\Big)\left[\frac{1}{l_c}\phi_1\Big(\frac{x}{l_c}\Big) * h_{L_0}^{l_c}(t,x)\right].$$

Taking the Fourier transform on the both sides,

$$\hat{h}_{L_0}^{l_c}(t,\xi) = \sigma^{-1}(\hat{\phi}_1(l_c\xi))^{-1}\mathcal{F}_\xi\{(v_t + F(v)_x - Dv_{xx})/\phi_0(\cdot/L_0)\}.$$

Because $v_t + F(v)_x - Dv_{xx} \in \mathcal{S}(\mathbb{R})$ and $1/\phi_0(\cdot/L_0) \in \mathcal{C}^\infty$ has at most polynomial growth, $\mathcal{F}_\xi\{(v_t + F(v)_x - Dv_{xx})/\phi_0(\cdot/L_0)\}$ is well-defined and also in $\mathcal{C}([0,T], \mathcal{S}(\mathbb{R}))$. In addition, $1/|\hat{\phi}_1(l_c\xi)|$ also has at most polynomial growth and then indeed $h_{L_0}^{l_c}(t,\xi) \in L^2(\mathbb{R})$.

Because $\mathcal{F}_\xi\{(v_t + F(v)_x - Dv_{xx})/\phi_0(\cdot/L_0)\}$ is in $\mathcal{C}([0,T], \mathcal{S}(\mathbb{R}))$ and $1/|\hat{\phi}_1(\xi)|$ has at most polynomial growth,

$$\begin{aligned}\|h_{L_0}^{l_c}(t,\cdot)\|_{L^2}^2 &\to \frac{1}{2\pi\sigma^2}\int |\hat{\phi}_1(0)|^{-2}|\mathcal{F}_\xi\{(v_t + F(v)_x - Dv_{xx})/\phi_0(\cdot/L_0)\}|^2 d\xi \\ &= \frac{1}{\sigma^2}\|(v_t + F(v)_x - Dv_{xx})/\phi_0(\cdot/L_0)\|_{L^2}^2\end{aligned}$$

as $l_c \to 0$, uniformly in $t \in [0,T]$. Then we have (5.10).

D.5. Proof of Proposition 5.7. For any pair (u,h) satisfying (3.1),

$$\begin{aligned}\|h(t,\cdot)\|_{L^2}^2 &= \sup_{f, f\neq 0} \frac{\langle h(t,\cdot), f\rangle^2}{\langle f, f\rangle} \geq \frac{\langle h(t,\cdot), (\Phi_{L_0}^{l_c})^T 1\rangle^2}{\langle (\Phi_{L_0}^{l_c})^T 1, (\Phi_{L_0}^{l_c})^T 1\rangle} \\ &= \|(\Phi_{L_0}^{l_c})^T 1\|_{L^2}^{-2}\langle \Phi_{L_0}^{l_c} h(t,\cdot), 1\rangle^2 = \|(\Phi_{L_0}^{l_c})^T 1\|_{L^2}^{-2}\langle u_t + F(u)_x - Du_{xx}, 1\rangle^2.\end{aligned}$$

Because $\Phi_{L_0}^{l_c} h \in \mathcal{C}([0,T], H^1(\mathbb{R}))$, we have $u - U \in \mathcal{C}([0,T], H^1(\mathbb{R}))$. Therefore $\lim_{x\to\pm\infty} u(t,x) = u_\pm$ and $\lim_{x\to\pm\infty} u_x(t,x) = 0$ for all $t \in [0,T]$. Thus,

$$\begin{aligned}I(u) &\geq \frac{1}{2}\|(\Phi_{L_0}^{l_c})^T 1\|_{L^2}^{-2}\int_0^T \langle u_t + F(u)_x - Du_{xx}, 1\rangle^2 dt \\ &= \frac{1}{2}\|(\Phi_{L_0}^{l_c})^T 1\|_{L^2}^{-2}\int_0^T \left(\int u_t(t,x)dx + F(u_+) - F(u_-)\right)^2 dt.\end{aligned}$$

By the Rankine–Hugoniot condition (2.3)

$$\begin{aligned} I(u) &\geq \frac{1}{2}\|(\Phi_{L_0}^{l_c})^T 1\|_{L^2}^{-2} \int_0^T \left(\int u_t(t,x)dx + \gamma(u_+ - u_-) \right)^2 dt \\ &= \frac{1}{2}\|(\Phi_{L_0}^{l_c})^T 1\|_{L^2}^{-2} \int_0^T \left(\int [u_t(t,x) - \frac{d}{dt}U(x-\gamma t)]dx \right)^2 dt \\ &= \frac{1}{2}\|(\Phi_{L_0}^{l_c})^T 1\|_{L^2}^{-2} \int_0^T \left(\frac{d}{dt}\int [u(t,x) - U(x-\gamma t)]dx \right)^2 dt. \end{aligned}$$

By the Schwarz inequality and noting that $u \in A$,

$$\begin{aligned} I(u) &\geq \frac{1}{2}T^{-1}\|(\Phi_{L_0}^{l_c})^T 1\|_{L^2}^{-2} \left(\int_0^T \frac{d}{dt}\int [u(t,x) - U(x-\gamma t)]dxdt \right)^2 \\ &= \frac{1}{2}T^{-1}\|(\Phi_{L_0}^{l_c})^T 1\|_{L^2}^{-2} \left(\int [u(T,x) - U(x-\gamma T)]dx \right)^2 \\ &= \frac{1}{2}T^{-1}\|(\Phi_{L_0}^{l_c})^T 1\|_{L^2}^{-2} \left(\int [U(x-\gamma T - X_0) - U(x-\gamma T)]dx \right)^2 \\ &= \frac{1}{2}T^{-1}\|(\Phi_{L_0}^{l_c})^T 1\|_{L^2}^{-2} \left(\int [U(x-X_0) - U(x)]dx \right)^2. \end{aligned}$$

D.6. Proof of Lemma 5.8. We first compute $(\Phi_{L_0}^{l_c})^T$. For any test functions f and g,

$$\begin{aligned} \langle \Phi_{L_0}^{l_c} f, g\rangle &= \int \sigma\phi_0\Big(\frac{x}{L_0}\Big) \int \frac{1}{l_c}\phi_1\Big(\frac{x-x'}{l_c}\Big) f(x')dx' g(x)dx \\ &= \int f(x') \int \sigma\phi_0\Big(\frac{x}{L_0}\Big)\frac{1}{l_c}\phi_1\Big(\frac{x-x'}{l_c}\Big) g(x)dxdx' = \langle f, (\Phi_{L_0}^{l_c})^T g\rangle. \end{aligned}$$

Thus $(\Phi_{L_0}^{l_c})^T g(x) = \sigma[\phi_0(\frac{x}{L_0})g(x)] * \frac{1}{l_c}\phi_1(-\frac{x}{l_c})$ and $(\Phi_{L_0}^{l_c})^T 1(x) = \sigma\phi_0(\frac{x}{L_0}) * \frac{1}{l_c}\phi_1(-\frac{x}{l_c})$. Then

$$\begin{aligned} \|(\Phi_{L_0}^{l_c})^T 1\|_{L^2}^2 &= \frac{1}{2\pi}\sigma^2 L_0^2 \int \hat\phi_0^2(L_0\xi)\hat\phi_1^2(-l_c\xi)d\xi \\ &\overset{l_c\to 0}{\to} \frac{1}{2\pi}\sigma^2 L_0^2 \int \hat\phi_0^2(L_0\xi)\hat\phi_1^2(0)d\xi = \sigma^2 L_0 \|\phi_0\|_{L^2}^2. \end{aligned}$$

D.7. Proof of Lemma 5.9. The proof for $F(w)_x - Dw_{xx}$ is similar to the proof of Lemma 5.5 so we skip it. For $U(x-\gamma T - X_0) - U(x)$, because $\phi_0(x) \equiv 1$ on

$x \in (-1,1)$ by Assumption 5.2,

$$\begin{aligned}
&\|[\phi_0^{-1}(\cdot/L_0)-1][U(\cdot-\gamma T-X_0)-U]\|_{L^2}^2\\
&=\int_{-\infty}^{-L_0}+\int_{L_0}^{\infty}\{[\phi_0^{-1}(x/L_0)-1][U(x-\gamma T-X_0)-U(x)]\}^2dx\\
&\le\int_{-\infty}^{-L_0}\{[\phi_0^{-1}(x/L_0)-1][u_- -U(x)]\}^2dx\\
&\quad+\int_{L_0}^{\infty}\{[\phi_0^{-1}(x/L_0)-1][U(x-\gamma T-X_0)-u_+]\}^2dx\\
&=\int_{-\infty}^{-L_0}\{[\phi_0^{-1}(x/L_0)-1][u_- -U(x)]\}^2dx\\
&\quad+\int_{L_0-\gamma T-X_0}^{\infty}\{[\phi_0^{-1}(x/L_0+(\gamma T+X_0)/L_0)-1][U(x)-u_+]\}^2dx\\
&\le\int_{-\infty}^{-C_w}\{[\phi_0^{-1}(x)-1][u_- -U(x)]\}^2dx+\int_{C_w}^{\infty}\{[\phi_0^{-1}(x+1)-1][U(x)-u_+]\}^2dx.
\end{aligned}$$

The last inequality holds because $\phi_0^{-1}(x) \ge 1$ is increasing for $x \ge 0$ and decreasing for $x < 0$, and $L_0 \ge 1$. Then we can find a uniform $C_w \ge 1$ in X_0 and T for (5.17).

Acknowledgment. The authors thank the Institut des Hautes Etudes Scientifiques (IHES) for its hospitality while part of this work was carried out.

REFERENCES

[1] C. Cardon-Weber, *Large deviations for a Burgers'-type SPDE*, Stochastic Process. Appl., 84 (1999), pp. 53–70.

[2] C. Cardon-Weber and A. Millet, *A support theorem for a generalized Burgers SPDE*, Potential Anal., 15 (2001), pp. 361–408.

[3] J.-C. Chen, D. Lu, J. S. Sadowsky, and K. Yao, *On importance sampling in digital communications. I. Fundamentals*, IEEE Journal on Selected Areas in Communications, 11 (1993), pp. 289–299.

[4] G. Da Prato and J. Zabczyk, *Stochastic Equations in Infinite Dimensions*, Encyclopedia Math. Appl. 44, Cambridge University Press, Cambridge, 1992.

[5] R. Dautray and J.-L. Lions, *Mathematical Analysis and Numerical Methods for Science and Technology. Vol.* 5, Springer-Verlag, Berlin, 1992.

[6] A. de Bouard and E. Gautier, *Exit problems related to the persistence of solitons for the Korteweg-de Vries equation with small noise*, Discrete Contin. Dyn. Syst., 26 (2010), pp. 857–871.

[7] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*, Stoch. Model. Appl. Probab. 38, Springer-Verlag, Berlin, 2010.

[8] J.-D. Deuschel and D. W. Stroock, *Large Deviations*, Pure and Applied Mathematics 137, Academic Press, Boston, MA, 1989.

[9] W. E, W. Ren, and E. Vanden-Eijnden, *Minimum action method for the study of rare events*, Comm. Pure Appl. Math., 57 (2004), pp. 637–656.

[10] J.-P. Fouque, J. Garnier, G. Papanicolaou, and K. Sølna, *Wave Propagation and Time Reversal in Randomly Layered Media*, Stochastic Model. Appl. Probab. 56, Springer, New York, 2007.

[11] M. I. Freidlin and A. D. Wentzell, *Random Perturbations of Dynamical Systems*, Grundlehren Math. Wiss. 260, 2nd ed., Springer-Verlag, New York, 1998.

[12] É. Gautier, *Large deviations and support results for nonlinear Schrödinger equations with additive noise and applications*, ESAIM Probab. Stat., 9 (2005), pp. 74–97.

[13] P. Glasserman and Y. Wang, *Counterexamples in importance sampling for large deviations probabilities*, Ann. Appl. Probab., 7 (1997), pp. 731–746.

[14] G. Iaccarino, R. Pecnik, J. Glimm, and D. Sharp, *A QMU approach for characterizing the operability limits of air-breathing hypersonic vehicles*, Reliability Engineering & System Safety, 96 (2011), pp. 1150–1160.
[15] A. M. Il′in and O. A. Oleĭnik, *Asymptotic behavior of solutions of the Cauchy problem for some quasi-linear equations for large values of the time*, Mat. Sb. (N.S.), 51 (93) (1960), pp. 191–216.
[16] C. K. R. T. Jones, R. Gardner, and T. Kapitula, *Stability of travelling waves for non-convex scalar viscous conservation laws*, Comm. Pure Appl. Math., 46 (1993), pp. 505–526.
[17] P. D. Lax, *Hyperbolic systems of conservation laws* II, Comm. Pure Appl. Math., 10 (1957), pp. 537–566.
[18] R. J. LeVeque, *Finite Volume Methods for Hyperbolic Problems*, Cambridge Texts Appl. Math., Cambridge University Press, Cambridge, 2002.
[19] M. Mariani, *Large deviations principles for stochastic scalar conservation laws*, Probab. Theory Related Fields, 147 (2010), pp. 607–648.
[20] J. Nocedal and S. J. Wright, *Numerical Optimization*, Springer Ser. Oper. Res. Financ. Eng., 2nd ed., Springer, New York, 2006.
[21] J. S. Sadowsky, *On Monte Carlo estimation of large deviations probabilities*, Ann. Appl. Probab., 6 (1996), pp. 399–422.
[22] J. S. Sadowsky and J. A. Bucklew, *On large deviations theory and asymptotically efficient Monte Carlo estimation*, IEEE Trans. Inform. Theory, 36 (1990), pp. 579–588.
[23] D. Siegmund, *Importance sampling in the Monte Carlo study of sequential tests*, Ann. Statist., 4 (1976), pp. 673–684.
[24] M. E. Taylor, *Partial Differential Equations* III *Nonlinear Equations*, Applied Math. Sci. 117, 2nd ed., Springer, New York, 2011.
[25] S. R. S. Varadhan, *Asymptotic probabilities and differential equations*, Comm. Pure Appl. Math., 19 (1966), pp. 261–286.
[26] N. West, G. Papanicolaou, P. Glynn, and G. Iaccarino, *A numerical study of filtering and control for scramjet engine flow*, in 20th AIAA Computational Fluid Dynamics Conference, Vol. 4, 2011, pp. 3010–3028.
[27] X. Zhou, W. Ren, and W. E, *Adaptive minimum action method for the study of rare events*, The Journal of Chemical Physics, 128 (2008), 104111.