

Giovanni A. Bonaschi · Mark A. Peletier

# Quadratic and rate-independent limits for a large-deviations functional

Received: 8 January 2015 / Accepted: 23 July 2015 / Published online: 23 August 2015  
© The Author(s) 2015. This article is published with open access at Springerlink.com

**Abstract** We construct a stochastic model showing the relationship between noise, gradient flows and rate-independent systems. The model consists of a one-dimensional birth–death process on a lattice, with rates derived from Kramers’ law as an approximation of a Brownian motion on a wiggly energy landscape. Taking various limits, we show how to obtain a whole family of generalized gradient flows, ranging from quadratic to rate-independent ones, connected via ‘ $L \log L$ ’ gradient flows. This is achieved via Mosco-convergence of the renormalized large-deviations rate functional of the stochastic process.

**Keywords** Large deviations · Gamma convergence · Gradient flows · Markov chains · Rate-independent systems

## 1 Introduction

### 1.1 Variational evolution

Two of the most studied types of variational evolution, ‘gradient-flow evolution’ and ‘rate-independent evolution’, differ in quite a few aspects. Although both are driven by the variation in space and time of an energy, gradient flows are in fact driven by energy gradients, while in practice rate-independent systems are driven by changes in the external loading (represented by the time variation of the energy). As a result, gradient-flow systems have an intrinsic timescale, while rate-independent systems (as the name signals) do not, and the mathematical definitions of solutions of the two are rather different [5,27].

Despite this, they share a common structure. Both can be written, at least formally, as

$$0 \in \partial\psi(\dot{x}(t)) + D_x E(x(t), t). \quad (1)$$

---

Communicated by Andreas Öchsner.

G. A. Bonaschi (✉) · M. A. Peletier  
Department of Mathematics and Computer Science, Institute for Complex Molecular Systems, Technische Universiteit Eindhoven,  
P.O. Box 513, 5600 MB Eindhoven, The Netherlands  
E-mail: g.a.bonaschi@tue.nl

M. A. Peletier  
E-mail: m.a.peletier@tue.nl

G. A. Bonaschi  
Dipartimento di Matematica, Università di Pavia, 27100 Pavia, Italy

Here  $E$  is the energy that drives the system, and the convex function  $\psi$  is a *dissipation potential*, with subdifferential  $\partial\psi$ . For gradient flows,  $\psi$  typically is quadratic, and  $\partial\psi$  single-valued and linear; for rate-independent systems,  $\psi$  is 1-homogeneous, and  $\partial\psi$  is a degenerate monotone graph.

Rate-independent systems have some unusual properties. Solutions are expected to be discontinuous, and therefore, the concept of smooth solutions is meaningless. Currently, two rigorous definitions of weak solutions are used, which we refer to as ‘energetic solutions’ [26] and ‘BV solutions’ [32]. Heuristically, the first corresponds to the principle ‘jump whenever it lowers the energy’, while the second can be characterized as ‘don’t jump until you become unstable’. For time-dependent *convex* energies, the two definitions coincide, but in the non-convex case they need not be.

Various rigorous justifications of rate-independent evolutions have been constructed, which underpin the rate-independent nature by obtaining it via upscaling from a ‘microscopic’ underlying system (e.g. [1, 13, 14, 21, 28, 40, 43]), with a perturbative approach (e.g. [11]), or from a chain model of highly nonlinear viscous springs [38]. The general approach in these results is to choose a microscopic model with a component of gradient-flow type (quadratic dissipation, deemed more ‘natural’) and then take a limit which induces the vanishing of the quadratic behaviour and the appearance of the rate-independent behaviour.

While these results give a convincing explanation of the rate-independent nature, most are based on *deterministic* microscopic models (exceptions are [38, 43, 44]). Other arguments suggest that the rate independence may arise through the interplay between thermal noise and a rough energy landscape. A well-studied example of this is the non-trivial temperature dependence of the yield stress in metals, which shows that the process is thermally driven (e.g. [7]), together with many classical non-rigorous derivations of rate-independent behaviour [8, 25, 39].

Recently, stochasticity has also been shown to play a role in understanding the origin of various gradient-flow systems, such as those with Wasserstein-type metrics [2, 3, 15, 30, 41]. In this paper, we ask the question whether these different roles of noise can be related:

*What is the relationship between noise, gradient flows, and rate-independent systems?*

We will provide a partial answer to this question by studying a simple stochastic model below. By taking various limits in this model, we obtain a full continuum of behaviours, among which rate-independent and quadratic gradient flow can be considered extreme cases. In this sense, both rate-independent and quadratic dissipation arise naturally from the same stochastic model in different limits.

### 1.2 The model

The model of this paper is a continuous-time Markov jump process  $t \mapsto X_t^n$  on a one-dimensional lattice, as shown in Fig. 1. Denoting by  $1/n$  the lattice spacing, we will be interested in the continuum limit as  $n \rightarrow \infty$ .

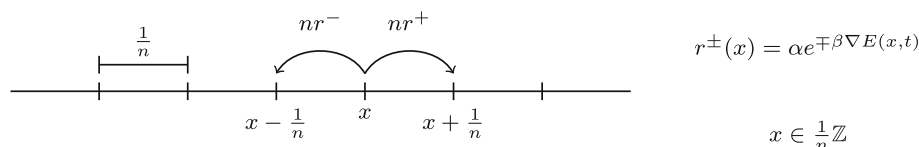
The evolution of the process can be described as follows. Assume that a smooth function  $(x, t) \mapsto E(x, t)$  is given and fix the origin as initial point. If the process is at the position  $x$  at time  $t$ , then it jumps in continuous time to its neighbours  $(x - 1/n)$  and  $(x + 1/n)$  with rate  $nr^-$  and  $nr^+$ , where  $r^\pm(x) = \alpha \exp(\mp\beta \nabla E(x, t))$  (throughout we use  $\nabla E(x, t)$  for the derivative with respect to  $x$ ).

The jump process of Fig. 1 has a bias in the direction  $-\nabla E$  of magnitude

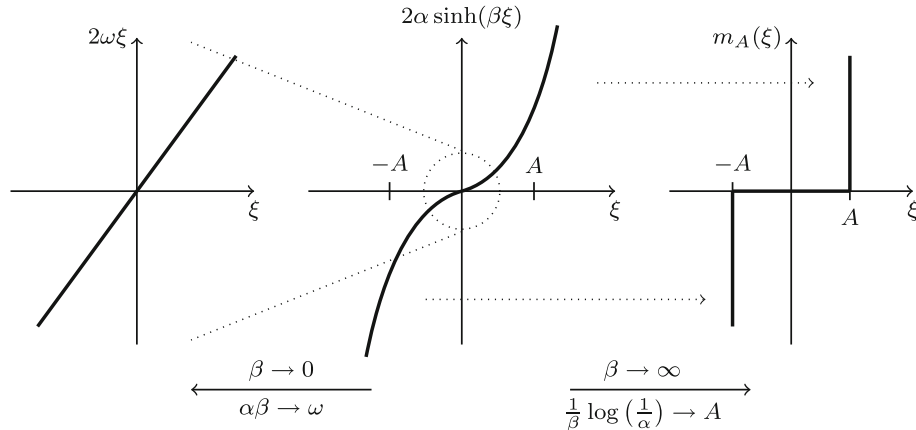
$$r^+ - r^- = -2\alpha \sinh(\beta \nabla E), \tag{2}$$

and we will see this expression return as a drift term in the limit problem. The parameter  $\alpha$  characterizes the rate of jumps and thus fixes the global timescale of the process; the parameter  $\beta$  should be thought as the inverse of temperature and characterizes the size of the noise.

In Sect. 1.6, we show the physics behind the choice of this particular model.



**Fig. 1** One-dimensional lattice with spacing  $1/n$ . The jump rates  $r^+$  and  $r^-$  depend on two parameters  $\alpha$  and  $\beta$  and on the derivative of the function  $E$



**Fig. 2** Middle graph shows the function  $\xi \mapsto 2\alpha \sinh(\beta\xi)$  for moderate values of  $\alpha$  and  $\beta$ . The left graph shows the limit for  $\beta \rightarrow 0$ , similar to zooming in to the region close to the origin; this limit is linear. The figure on the right shows the limiting behaviour when  $\beta \rightarrow \infty$ , for a specific scaling of  $\alpha$ . This second limit does not exist as a function, but only as a graph (a subset of the plane) defined in (5)

### 1.3 Heuristics

We now give a heuristic view of the dependence of this stochastic process on the parameters  $n$ ,  $\alpha$ , and  $\beta$ , and in doing so we look ahead at the rigorous results that we prove below.

First, as  $n \rightarrow \infty$ , the process  $X^n$  becomes deterministic, as might be expected, and its limit  $x$  satisfies the differential equation suggested by (2):

$$\dot{x}(t) = -2\alpha \sinh(\beta \nabla E(x(t), t)). \tag{3}$$

Equation (3) is of the form (1) with  $(\partial\psi)^{-1} = 2\alpha \sinh(\beta\cdot)$ . From the viewpoint of the gradient-flow-versus-rate-independence discussion above, the salient feature of the function  $\xi \mapsto 2\alpha \sinh(\beta\xi)$  is that it embodies both quadratic and rate-independent behaviour in one and the same function, in the form of limiting behaviours according to the values of the parameters  $\alpha$ ,  $\beta$ . This is illustrated in Fig. 2 as follows. On the one hand, if we construct a limit by zooming in to the origin, corresponding to  $\beta \rightarrow 0$ ,  $\alpha \sim \omega/\beta$  (the left-hand figure), then we find a limit that is linear; on the other hand, if we zoom out and rescale with  $\beta \rightarrow \infty$  and  $\alpha \sim e^{-\beta A}$ , then the exponential growth causes the limit to be the monotone graph in the right-hand side; see  $m_A$  in (5).

These two limiting cases give rise to a gradient-flow and a rate-independent behaviour, respectively. In formulas, as  $\alpha \rightarrow \infty$  and  $\beta \rightarrow 0$  with  $\alpha\beta \rightarrow \omega$  for fixed  $\omega > 0$ , then Eq. (3) converges to

$$\dot{x}(t) = -2\omega \nabla E(x(t), t), \tag{4}$$

which is a gradient flow of  $E$ . The limit  $\alpha \rightarrow \infty$  corresponds to large rate of jumps in the underlying stochastic process, while  $\beta \rightarrow 0$  corresponds to a weak influence of the energy gradient.

In the other case, as  $\alpha \rightarrow 0$  and  $\beta \rightarrow \infty$  with  $\alpha \sim \exp(-\beta A)$ , the rate of jumps is low, but the influence of the energy becomes large. Formally, we find the limiting equation

$$\dot{x}(t) \in m_A(-\nabla E(x(t), t)), \text{ with } m_A(\xi) = \begin{cases} \emptyset & \text{if } \xi < -A, \\ [-\infty, 0] & \text{if } \xi = -A, \\ \{0\} & \text{if } -A < \xi < A, \\ [0, \infty] & \text{if } \xi = A, \\ \emptyset & \text{if } \xi > A. \end{cases} \tag{5}$$

Again formally, in this limit the system can only move while  $\nabla E = \pm A$ ; whenever the force  $|\nabla E|$  is less than  $A$ , the system is frozen, while values of  $|\nabla E|$  larger than  $A$  should never appear. In Sect. 4, we obtain a rigorous version of this evolution as the limit system.

#### 1.4 Large deviations, gradient flows, and variational formulations

Before we describe the results of this paper, we comment on the methods that we use. We previously introduced the concept of gradient flows, and now we introduce the one of *large deviations* (both are defined precisely in Sect. 2).

In the context of stochastic systems, the theory of large deviations provides a characterization of the probability of rare events, as some parameter—in our case  $n$ —tends to infinity. In the case of stochastic processes, this leads to a *large-deviations rate function*  $\mathcal{I}$  that is defined on a suitable space of curves. It is now known that many gradient flows and large-deviations principles are strongly connected [2,3,16,30]. In abstract terms, the rate function  $\mathcal{I}$  of the large-deviations principle simultaneously figures as the defining quantity of the gradient flow, in the sense that

$$\begin{aligned} \mathcal{I} &\geq 0; \\ t \mapsto z(t) \text{ is a solution of the gradient flow} &\iff \mathcal{I}(z(\cdot)) = 0. \end{aligned}$$

The components of the gradient flow (the energy  $E$  and dissipation potential  $\psi$ ) can be recognized in  $\mathcal{I}$ . In [30], it was shown how jump processes may generate large-deviations rate functions with non-quadratic dissipation, leading to the concept of *generalized* gradient flows (see also [17,31,33]).

The central tool in this paper is this functional  $\mathcal{I}$  that characterizes both the large deviations of the stochastic process and the generalized gradient-flow structure of the limit. Our convergence proofs will be stated and proved using only this functional, giving a high level of coherence to the results.

#### 1.5 Results

In this section, we give a non-rigorous description of the results of this paper, with pointers towards the rigorous theorems later in the paper (Fig. 3).

Fix an energy  $E \in C^1(\mathbb{R} \times [0, T])$ . We start with the large-deviations result for the jump process  $X^n$  due to Wentzell [47,48].

**Statement 1** (Large deviations) *Let  $\alpha$  and  $\beta$  constant, then  $X^n$  satisfies a large-deviations principle for  $n \rightarrow \infty$ , with rate function  $\mathcal{I}_{\alpha,\beta}$  given in (15). Moreover, the minimizer of  $\mathcal{I}_{\alpha,\beta}$  satisfies the generalized gradient-flow equation (3).*

This result is stated in Theorem 1 with a sketch of the proof as it is presented in the introduction of [20]. In accordance with the discussion above,  $\mathcal{I}_{\alpha,\beta}(x) \geq 0$  for all curves  $x : [0, T] \rightarrow \mathbb{R}$ , and  $\mathcal{I}_{\alpha,\beta}(x) = 0$  if and only if  $x$  is a solution of (3).

Next we prove that the functional  $\mathcal{I}_{\alpha,\beta}$  converges to two functionals  $\mathcal{I}_Q$  and to  $\mathcal{I}_{RI}$  in the sense of Mosco-convergence, defined in (20) (in the spirit of pE convergence of [29]), when  $\alpha$  and  $\beta$  have the limiting behaviour of Fig. 2. The limiting functionals drive, respectively, a quadratic gradient-flow and a rate-independent evolution.

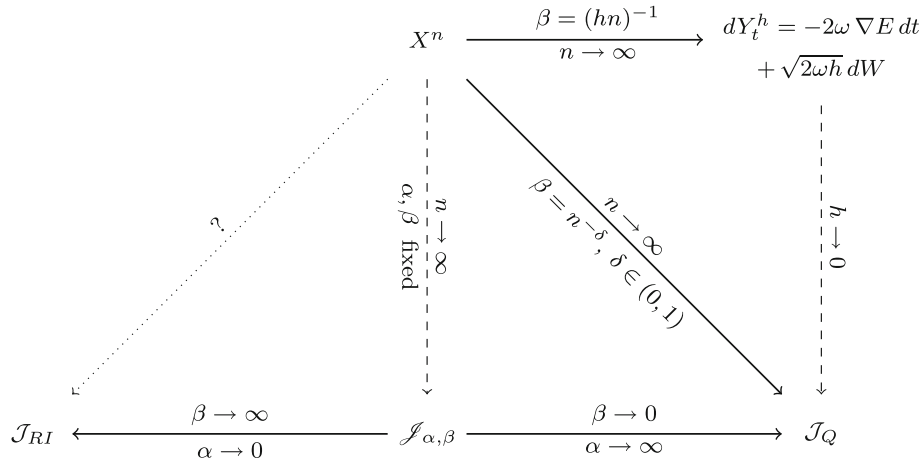
**Statement 2** (Connection)

- A1. *Let  $\alpha \rightarrow \infty$  and  $\beta \rightarrow 0$ , such that  $\alpha\beta \rightarrow \omega$ , for some  $\omega > 0$  fixed. Then, we have that, after rescaling,  $\mathcal{I}_{\alpha,\beta} \rightarrow \mathcal{I}_Q$ , where  $\mathcal{I}_Q$  is defined in (21); moreover  $\mathcal{I}_Q(x) = 0$  if and only if  $x$  solves (4).*
- A2. *Let  $\beta \rightarrow \infty$  and choose  $\alpha = e^{-\beta A}$ , for some  $A > 0$  fixed. Then, after an appropriate rescaling,  $\mathcal{I}_{\alpha,\beta} \rightarrow \mathcal{I}_{RI}$ , where  $\mathcal{I}_{RI}$  is given in (30); moreover  $\mathcal{I}_{RI}(x) = 0$  if and only if  $x$  is an appropriately defined solution of (5).*

The Mosco-convergence stated above also implies that minimizers  $x_{\alpha,\beta}$  of  $\mathcal{I}_{\alpha,\beta}$  converge to the minimizers of  $\mathcal{I}_Q$  and  $\mathcal{I}_{RI}$  in the corresponding cases, i.e. that the solutions  $x_{\alpha,\beta}$  of (3) converge to the solutions of (4) and (5). Point A1 is proven in Theorem 2, and point A2 in Theorem 5.

Together, Statements 1 and 2 describe a sequential limit process: first, we let  $n \rightarrow \infty$ , and then, we take limits in  $\alpha$  and  $\beta$ . For the quadratic case (A1), we can also combine the limits:

**Statement 3** (Combining the limits) *Let  $n \rightarrow \infty$ , and take  $\beta = \beta_n \sim n^{-\delta}$  for some  $0 \leq \delta \leq 1$ ; let  $\alpha = \alpha_n$  be such that  $\alpha_n \beta_n \rightarrow \omega$ , for some fixed  $\omega > 0$ .*



**Fig. 3** Schematic representation of this paper. In the top centre, there is the generator of the Markov process. The arrows starting from  $X^n$  represent the limiting behaviour for  $n \rightarrow \infty$  in different regimes. The *centre downwards arrow* represents the limit with  $\alpha$  and  $\beta$  fixed and ends at the rate functional  $\mathcal{J}_{\alpha, \beta}$  (Statement 1). Statement 3 is represented by the *right side of the figure*. The limits with  $\beta = \beta(n)$  show that the limiting behaviour may be either a Brownian motion with drift (B2a) or a gradient flow characterized by the rate functional  $\mathcal{J}_Q$  (B1). In the bottom part, there are the two Mosco-limits, representing Statement 2 where (A1) is the right arrow and (A2) the left one. *Dashed lines* are known results, the *thick lines are our contribution*, and the *dotted line is an open problem*

B1. First, let  $0 < \delta < 1$ . Then,  $X^n$  satisfies a large-deviations principle as  $n \rightarrow \infty$ , with rate function  $\mathcal{J}_Q$ ; the Markov process  $X^n$  has a deterministic limit (4), and this limit minimizes  $\mathcal{J}_Q$  (as we already mentioned).

B2a. In the case  $\delta = 1$ , let  $\alpha_n \beta_n \rightarrow \omega$ ,  $n \beta_n \rightarrow 1/h$ , for some  $\omega, h > 0$ ; then  $X^n$  converges to the process  $Y^h$  described by the SDE

$$dY_t^h = -2\omega \nabla E(Y_t^h, t) dt + \sqrt{2\omega h} dW_t. \tag{6}$$

B2b. The process  $Y^h$  in (6) satisfies a large-deviations principle for  $h \rightarrow 0$  with rate function  $\mathcal{J}_Q$ .

B3. The case  $\delta = 0$  corresponds to point A1 of Statement 2, where first  $n \rightarrow \infty$  and then  $\beta \rightarrow 0$  and  $\alpha \rightarrow \infty$ .

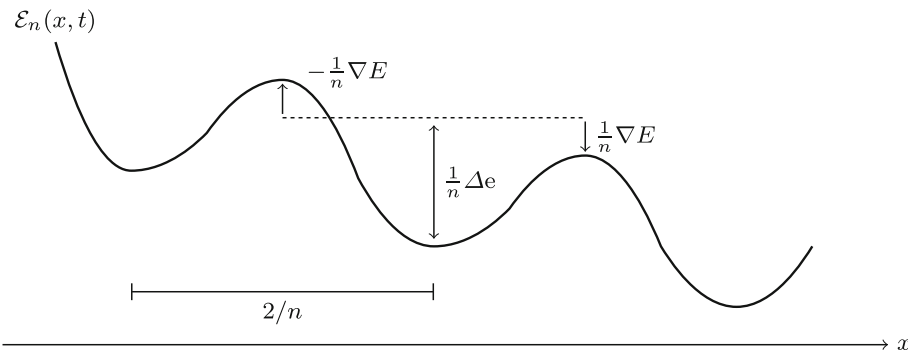
Point B1 of Statement 3 is given in Theorem 3. Point B2a is given in Theorem 4; point B2b is the well-known result of Freidlin and Wentzell [23, Ch. 4-Th. 1.1], and it is included in Theorem 3.

*Remark 1* In this paper, we consider only the one-dimensional case. We make this choice because the main goal of the paper is to show the connection and the interplay between large deviations and gradient flows, and the one-dimensionality allows us to avoid various technical complications. However, the generalization to higher dimension is in some cases just a change in the notation and does not require any relevant modification in some of the proofs. The rate-independent limit in higher dimensions is non-trivial, and it is the object of work in progress.

### 1.6 Modelling

The choice of this stochastic process is inspired by the noisy evolution of a particle in a wiggly energy landscape. An example could be that of a Brownian particle in an energy landscape of the form  $\mathcal{E}_n(x, t) = E(x, t) + n^{-1}e(nx)$ , where  $E$  is the smooth energy introduced above, and  $e$  is a fixed periodic function that can be thought of the type present in [1, 28]. If the noise is small with respect to the variation of  $e$  ( $\max e - \min e$ ), then this Brownian particle will spend most of its time near the wells of  $\mathcal{E}_n$ , which are close to the wells of  $e(n \cdot)$ . Kramers' formula [9, 24] provides an estimate of the rate at which the particle jumps from one well to the next; below we show how some approximations lead to the jump rates  $r^\pm$  of Fig. 1.

In the wiggly energy  $\mathcal{E}_n(x, t) = E(x, t) + n^{-1}e(nx)$ , assume that  $E$  varies slowly in both  $x$  and  $t$ , and that  $e$  has period 2 and let  $n$  be large. Then, a small patch of the energy landscape of  $\mathcal{E}_n$  looks like Fig. 4. The height of the energy barriers to the left and right of a well equals  $\Delta e := (\max e - \min e)$ , plus a perturbation from the smooth energy  $E$ , which to leading order has size  $n^{-1} \nabla E$ .



**Fig. 4** Global component  $E$  perturbs the height of the energy barriers, leading to the formula for  $r^\pm$  in Fig. 1

We now assume that the position  $Z_t$  of the system solves the SDE

$$dZ_t = -\nabla \mathcal{E}_n(Z_t, t) dt + \sqrt{\frac{2}{\beta n}} dW_t.$$

Here  $\beta$  characterizes the noise and can be interpreted as  $1/kT$  as usual, although in this case there is an additional scaling factor  $n$ . For sufficiently large  $\beta$ , the rates of escape from a well, to the left and to the right, are given by Kramers' law to be approximately

$$a \exp[\beta(-\Delta e \pm \nabla E)], \quad (7)$$

where the minus sign applies to the rate of leaving to the right. Here  $a$  is a constant depending on the form of  $e$  [9,24]. Writing

$$\alpha = a \exp[-\beta \Delta e], \quad (8)$$

we find the rates  $r^\pm$  of Fig. 1.

In this formula for  $\alpha$ , it appears that  $\alpha$  and  $\beta$  are coupled. From a modelling point of view, this is true: if one varies the temperature while keeping all other parameters fixed, then both  $\beta$  and  $\alpha$  will change. Note that the scaling regime  $\alpha \sim e^{-\beta A}$  is exactly this case, with  $A = \Delta e$ . On the other hand, the parameter  $a$  in (8) is still free, and this allows us to consider  $\alpha$  and  $\beta$  as independent parameters when necessary.

Note that by this derivation a natural alternative would be to define the jump rates  $r^\pm$  in terms of the energy differences rather than the gradient, i.e.

$$r^\pm(x) = \alpha \exp[-\beta n (E(x \pm 1/n, t) - E(x, t))] \quad (\text{compare to Fig. 1}).$$

This choice would also render the process time-reversible. However, for the purposes of this paper the difference is mathematically negligible, and we make the choice of Fig. 1 since it leads to simpler and easier-to-read formulas.

## 1.7 Outline

The paper evolves as described in the following. In Sect. 2, we introduce the concepts of (generalized) gradient flows and of large deviations and we show the connection between the two concepts. In Sect. 3, we prove point A1 of Statement 2 and the whole Statement 3. Then, in Sect. 4, we introduce the space of functions of bounded variations and rate-independent systems, and we prove point A2 of Statement 2. We end the paper in Sect. 5 with a final discussion.

## 2 Gradient flows and large deviations

In the introduction, we mentioned that the methods of this paper make use of a certain unity between gradient flows and large-deviations principles: the same functional  $\mathcal{J}$  that defines the gradient flow also appears as the rate function of a large-deviations principle. We now describe gradient flows, large deviations, and this functional  $\mathcal{J}$ .

2.1 Gradient flows

Given a  $C^1$  energy  $E : \mathbb{R} \rightarrow \mathbb{R}$ , we call a *gradient flow* of  $E$  the flow generated by the equation

$$\dot{x} = -\nabla E(x).$$

The energy  $E$  decreases along a solution, since

$$\frac{d}{dt} E(x(t)) = \nabla E \cdot \dot{x} = -|\nabla E|^2 = -|\dot{x}|^2.$$

Adopting the notation  $\psi(\xi) = \frac{\xi^2}{2}$  and  $\psi^*(\eta) = \frac{\eta^2}{2}$ , this identity can be integrated in time to find

$$\int_0^T (\psi(\dot{x}) + \psi^*(-\nabla E)) dt + E(x(T)) - E(x(0)) = 0$$

In this paper, we study a generalized concept of gradient flow, considering the energy equality for a broader class of couples  $\psi, \psi^*$ . We will allow the energy to be also dependent on time. We recall the definition of the Legendre transform: given  $\psi : \mathbb{R} \rightarrow \mathbb{R}$ , we define the transform  $\psi^*$  as

$$\psi^*(w) := \sup_v \{v \cdot w - \psi(v)\}.$$

In the following, apart from the rate-independent case, we will assume that  $\psi \in C^1(\mathbb{R})$  is symmetric, super-linear, and strictly convex, so that its Legendre transform  $\psi^*$  will share the same properties as well.

A curve  $x : [0, T] \rightarrow \mathbb{R}$  is an absolutely continuous curve, i.e.  $x \in AC(0, T)$ , if for every  $\varepsilon > 0$ , there exists a  $\delta > 0$  such that, for every finite sequence of pairwise disjoint intervals  $(t_j, \tau_j) \subset [0, T]$  satisfying  $\sum |t_j - \tau_j| < \delta$ , then  $\sum |x(t_j) - x(\tau_j)| < \varepsilon$ . The space  $AC(0, T)$  coincides with the Sobolev space  $W^{1,1}(0, T)$  [10, Ch. 8].

**Definition 1** (*Generalized gradient flow*) Given an energy  $E \in C^1(\mathbb{R} \times [0, T])$ , a convex dissipation potential  $\psi \in C^1(\mathbb{R})$  with  $\psi(v) = \psi(-v)$ , let  $\psi^*$  be its Legendre transform. Then, a curve  $x \in AC(0, T)$  is a (*generalized*) *gradient flow* of  $E$  with dissipation potential  $\psi$  in a given time interval  $[0, T]$ , if it satisfies the energy identity

$$\begin{aligned} & \int_0^T (\psi(\dot{x}(t)) + \psi^*(-\nabla E(x(t), t))) dt + E(x(T), T) - E(x(0), 0) \\ & - \int_0^T \partial_t E(x(t), t) dt = 0. \end{aligned} \tag{9}$$

Note that the left-hand side of (9) is non-negative for any function  $x$ , since

$$\begin{aligned} \frac{d}{dt} E(x(t), t) &= \nabla E(x(t), t) \dot{x}(t) + \partial_t E(x(t), t) \\ &\geq -\psi(\dot{x}(t)) - \psi^*(-\nabla E(x(t), t)) + \partial_t E(x(t), t). \end{aligned}$$

From this inequality, one deduces that *equality* in (9), as required by Definition 1, implies that for almost all  $t \in [0, T]$

$$\dot{x}(t) = \partial \psi^*(-\nabla E(x(t), t)), \tag{10}$$

where  $\partial \psi^*$  is the subdifferential of  $\psi^*$ . We will not use this form of Eq. (10); the arguments of this paper are based on Definition 1 instead.

Existence and uniqueness of classical gradient-flow solutions in  $\mathbb{R}$  (i.e. with quadratic  $\psi$ ) follows from classical ODE theory; in recent years, the theory has been extended to metric spaces and spaces of probability measures [5].

In our case, we will require the energy to satisfy the following conditions

$$\begin{cases} E \in C^1(\mathbb{R} \times [0, T]), E \geq 0, \\ |\nabla E| \leq R < \infty, \\ \nabla E \text{ is uniformly Lipschitz continuous in } t. \end{cases} \tag{11}$$

*Remark 2* In this paper, we analyse generalized gradient flow with  $\psi$  of Definition 1 chosen to be strictly convex and superlinear at infinity. The rate-independent evolution (5) is formally the case of a non-strictly convex, 1-homogeneous dissipation potential  $\psi$ , and for this case, there are several natural ways to define a rigorous solution concept. In Sect. 4, we show how generalized gradient flows for finite  $\alpha$  and  $\beta$ , with strictly convex  $\psi$ , converge to a *specific* rigorous rate-independent solution concept, the so-called *BV solutions* [31, 32]. We define this concept in Definition 3.

Returning to the unity between gradient flows and large deviations, for generalized gradient flows the functional  $\mathcal{J}$  mentioned before is the left-hand side of (9).

## 2.2 Large deviations

‘Large deviations’ of a random variable are rare events, and large-deviations theory characterizes the rarity of certain rare events for a sequence of random variables. Let  $\{X^n\}$  be such a sequence of random variables with values in some metric space.

**Definition 2** ([46])  $\{X^n\}$  satisfies a large-deviations principle (LDP) with speed  $a_n \rightarrow \infty$ , if there exists a lower semicontinuous function  $\mathcal{J} : S \rightarrow [0, \infty]$  with compact sublevel sets such that for each open set  $O$ ,

$$\liminf_{n \rightarrow \infty} \frac{1}{a_n} \log P(X^n \in O) \geq - \inf_{x \in O} \mathcal{J}(x),$$

and for each closed set  $C$

$$\limsup_{n \rightarrow \infty} \frac{1}{a_n} \log P(X^n \in C) \leq - \inf_{x \in C} \mathcal{J}(x).$$

The function  $\mathcal{J}$  is called the *rate function* for the large-deviations principle.

Intuitively, the two inequalities above state that

$$\text{Prob}(X^n \simeq x) \sim e^{-a_n \mathcal{J}(x)},$$

where we purposefully use the vague notations  $\simeq$  and  $\sim$ ; the rigorous versions of these symbols are exactly given by Definition 2.

*Remark 3* Typically, the rate function for Markov processes contains a term  $I_0$  characterizing the large deviations of the initial state  $X^n(0)$ . In the following, we will always assume that the starting point will be fixed, or at least that  $X^n(0) \rightarrow x_0$ , so that  $I_0(x)$  equals 0 if  $x = x_0$ , and  $+\infty$  otherwise, so we will disregard  $I_0$ .

## 2.3 The Feng–Kurtz method

Feng and Kurtz created a general method to prove large-deviations principles for Markov processes [20]. The method provides both a formal method to calculate the rate functional and a rigorous framework to prove the large-deviations principle. Here we present only the formal calculation.

Consider a sequence of Markov processes  $\{X^n\}$  in  $\mathbb{R}$ , which we take time-invariant for the moment, and consider the corresponding evolution semigroups  $\{S_n(t)\}$  defined by [see also (13)]

$$S_n(t)f(x) = \mathbb{E}[f(X^n(t)) \mid X^n(0) = x], \quad f \in C_b(\mathbb{R}),$$

satisfying

$$\frac{d}{dt} S_n(t)f = \Omega_n S_n(t)f, \quad S_n(0)f = f,$$

where  $\Omega_n$  is the generator of  $X_n$ . For any time interval  $[0, T]$ , where  $T$  may be infinite,  $X^n(\cdot)$  is an element of the Skorokhod space  $D([0, T])$ , the space of *cadlag* functions (right continuous and with limit from left). To obtain the rate functional, we define the *nonlinear generator*

$$(\mathbb{H}_n f)(x) := \frac{1}{a_n} e^{-a_n f(x)} \left( \Omega_n e^{a_n f} \right)(x).$$



If  $H_n \rightarrow H$  in some sense, and if  $Hf$  depends locally on  $\nabla f$ , we then define the *Hamiltonian*  $H(x, p)$  through

$$Hf(x) =: H(x, \nabla f(x)).$$

By computing the Legendre transform of  $H(x, p)$ , we obtain the *Lagrangian*

$$L(x, \dot{x}) = \sup_{p \in \mathbb{R}} \{ \dot{x} \cdot p - H(x, p) \}.$$

The Feng–Kurtz method then states, formally, that  $\{X^n\}$  satisfies a large-deviations principle in  $D([0, T])$  with speed  $a_n$ , with a rate function

$$\mathcal{I}(x) = \begin{cases} \int_0^T L(x, \dot{x}) dt & \text{if } x \in AC(0, T), \\ +\infty & \text{otherwise.} \end{cases}$$

In the book [20], a general method is described to make this algorithm rigorous.

### 2.4 Large deviations of $X^n$

We now apply this method to the process  $X^n$  described in the introduction. It is a continuous-time Markov chain, defined by its generator

$$\begin{aligned} \Omega_n f(x) := & n\alpha e^{-\beta \nabla E(x,t)} \left( f\left(x + \frac{1}{n}\right) - f(x) \right) \\ & + n\alpha e^{\beta \nabla E(x,t)} \left( f\left(x - \frac{1}{n}\right) - f(x) \right). \end{aligned} \tag{12}$$

For  $f \in C_b(\mathbb{R})$ , the expected value  $\mathbb{E}$  is defined as

$$\mathbb{E}(f(X_t^n) | X_0^n) = \int_{\mathbb{R}} f(z) d\mu_t^n(z), \tag{13}$$

with, denoting by  $\Omega_n^T$  the adjoint of  $\Omega_n$ ,

$$\begin{cases} \partial_t \mu_t = \Omega_n^T \mu_t, \\ \mu_0 = \delta_{X_0^n}. \end{cases} \tag{14}$$

where  $\mu_t^n$  is the law at time  $t$  of the process  $X^n$  started at the position  $X_0^n$  at time  $t = 0$ . Under the condition (11), the martingale problem (14) is well posed, since the operator  $\Omega_n$  is bounded in the uniform topology.

The rigorous proof of Statement 1 consists of the following theorem

**Theorem 1** (Large-deviations principle for  $\Omega_n$ ) *Let  $E : \mathbb{R} \times [0, T] \rightarrow \mathbb{R}$  satisfying condition (11). Consider the sequence of Markov processes  $\{X^n\}$  with generator  $\Omega_n$  defined in (12) and with  $X^n(0)$  converging to  $x_0$  for  $n \rightarrow \infty$ . Then, the sequence  $X^n$  satisfies a large-deviations principle in  $D([0, T])$  with speed  $n$  and rate function*

$$\mathcal{I}_{\alpha,\beta}(x) := \begin{cases} \beta \int_0^T \left( \psi_{\alpha,\beta}(\dot{x}) + \psi_{\alpha,\beta}^*(\nabla E) + \dot{x} \nabla E \right) dt & \text{for } x \in AC(0, T), \\ +\infty & \text{otherwise,} \end{cases} \tag{15}$$

where

$$\begin{aligned} \psi_{\alpha,\beta}(v) &:= \frac{v}{\beta} \log \left( \frac{v + \sqrt{v^2 + 4\alpha^2}}{2\alpha} \right) - \frac{1}{\beta} \sqrt{v^2 + 4\alpha^2} + \frac{2\alpha}{\beta}, \\ \psi_{\alpha,\beta}^*(w) &= \frac{2\alpha}{\beta} (\cosh(\beta w) - 1). \end{aligned} \tag{16}$$

The proof can be found in [23, Ch. 5-Th. 2.1] when the energy  $E$  is independent of time. In the general case of a time-dependent energy, the proof follows considering a space–time process, as shown in the proof of Theorem 3.

Note that solutions of  $\mathcal{I}_{\alpha,\beta}(x) = 0$  satisfy the gradient-flow Eq. (10), which in this case indeed is Eq. (3), i.e.

$$\dot{x} = -2\alpha \sinh(\beta \nabla E(x, t)).$$

In the remainder of this paper, we will consider sequences in  $\alpha$  and  $\beta$ ; to reduce notation, we will drop the double index, writing  $\psi_\beta$  and  $\psi_\beta^*$  for  $\psi_{\alpha,\beta}$  and  $\psi_{\alpha,\beta}^*$ ; similarly, we define the rescaled functional  $\mathcal{J}_\beta$ ,

$$\begin{aligned} \mathcal{J}_\beta(x) &:= \frac{1}{\beta} \mathcal{I}_{\alpha,\beta}(x) \\ &= \int_0^T \left( \psi_\beta(\dot{x}(t)) + \psi_\beta^*(-\nabla E(x(t), t)) + \dot{x}(t) \nabla E(x(t), t) \right) dt, \end{aligned} \tag{17}$$

for  $x \in AC(0, T)$  and  $\mathcal{J}_\beta(x) = +\infty$  otherwise.

*Remark 4* The theorem above shows how the rate function  $\mathcal{I}_{\alpha,\beta}$  can be interpreted as defining a generalized gradient flow. This illustrates the structure of the fairly widespread connection between gradient flows and large deviations: in many systems, the rate function not only defines the gradient-flow evolution, through its zero set, but the components of the gradient flow ( $E$  and  $\psi$ ) can be recognized in the rate function. This connection is explored more generally in [30] as we describe in the following. Define a so-called  $L$ -function  $\mathcal{L}(z, s)$ , positive, convex in  $s$  for all  $z$  and inducing an evolution equation. The authors of [30, Lemma 2.1 and Prop. 2.2] show that if  $D_s \mathcal{L}(z, 0)$  is an exact differential, say  $DS(z)$ , then it is possible to write  $\mathcal{L}$  as

$$\mathcal{L}(z, s) = \Psi(z, s) + \Psi^*(z, -DS(z)) + \langle DS(z), s \rangle, \tag{18}$$

where  $\Psi^*$  can be expressed in terms of the Legendre transform  $\mathcal{H}(z, \xi)$  of  $\mathcal{L}$  as

$$\Psi^*(z, \xi) := \mathcal{H}(z, DS(z) + \xi) - \mathcal{H}(z, DS(z)).$$

Applying the same procedure to our case, with  $\mathcal{L} = L$  defined in (19), we obtain after some calculations that

$$DS(z) := \frac{1}{2} \log \left( \frac{r^-(z)}{r^+(z)} \right).$$

Substituting in  $DS(z)$  our choice for  $r^+$  and  $r^-$ , namely

$$r^+ = \alpha e^{-\beta \nabla E(z)}, \quad r^- = \alpha e^{\beta \nabla E(z)},$$

it follows that  $S(z) = \beta E(z)$ .

### 2.5 Calculating the large-deviations rate functional for (12)

We conclude this section by calculating the rate function for the simpler situation when the jump rates  $r^\pm$  are constant in space and time, as it is shown in the introduction of [20]. This formally proves Theorem 1, substituting in the end the expression of  $r^\pm$  from (12).

With constant jump rates, the generator reduces to

$$\Omega_n f(x) = nr^+ \left[ f \left( x + \frac{1}{n} \right) - f(x) \right] + nr^- \left[ f \left( x - \frac{1}{n} \right) - f(x) \right],$$

and for  $n \rightarrow \infty$  it converges to  $\Omega f(x) = (r^+ - r^-) \nabla f(x)$ . As we said in the introduction, the process  $X^n$  has a deterministic limit, i.e.  $X^n \rightarrow x$  a.s., with  $\dot{x} = r^+ - r^-$ .

In order to calculate the rate functional, we compute the nonlinear generator and the limiting Hamiltonian and Lagrangian. We have

$$H_n f(x) = r^+ \left[ e^{n(f(x+1/n)-f(x))} - 1 \right] + r^- \left[ e^{n(f(x-1/n)-f(x))} - 1 \right],$$

so that

$$\lim_{n \rightarrow \infty} H_n = H(x, p) = r^+ (e^p - 1) + r^- (e^{-p} - 1).$$

We then obtain by an explicit calculation the Lagrangian

$$L(x, \dot{x}) = \dot{x} \log \left( \frac{\dot{x} + \sqrt{\dot{x}^2 + 4r^+r^-}}{2r^+} \right) - \sqrt{\dot{x}^2 + 4r^+r^-} + r^+ + r^-, \tag{19}$$

and substituting  $r^+$  and  $r^-$  with the corresponding ones from (12) we get

$$L(x, \dot{x}) = \beta (\psi(\dot{x}) + \psi^*(\nabla E) + \dot{x} \nabla E),$$

and we formally have proved Theorem 1.

### 3 The quadratic limit

In this section, we precisely state and prove point A1 of Statement 2 and the whole of Statement 3. We are in the regime where  $\beta \rightarrow 0, \alpha \rightarrow \infty$ , with  $\alpha\beta \rightarrow \omega$ .

First, we show heuristically why the functional  $\mathcal{J}_\beta$  defined in (17) is expected to converge to  $\mathcal{J}_Q$  defined in (21). Looking at the equation that minimizes the functional  $\mathcal{J}_\beta$ , and doing a Taylor expansion for  $\beta \ll 1$ ,

$$\dot{x} = -2\alpha \sinh(\beta \nabla E) \simeq -2\alpha\beta \nabla E \rightarrow -2\omega \nabla E.$$

Considering the functional  $\mathcal{J}_\beta$  for  $\beta \ll 1$ , it can be seen that

$$\begin{aligned} \psi_\beta^*(w) &= \frac{2\alpha}{\beta} (\cosh(\beta w) - 1) \simeq \alpha\beta w^2 \rightarrow \omega w^2, \\ \psi_\beta(v) &= \frac{v}{\beta} \log \left( \frac{v + \sqrt{v^2 + 4\alpha^2}}{2\alpha} \right) - \frac{1}{\beta} \sqrt{v^2 + 4\alpha^2} + \frac{2\alpha}{\beta} \simeq \frac{v^2}{4\alpha\beta}, \end{aligned}$$

implying that

$$\psi_\beta^*(w) \rightarrow \omega w^2, \quad \psi_\beta(v) \rightarrow \frac{v^2}{4\omega}.$$

We now turn to the rigorous proof of the convergence to the quadratic gradient flow and therefore point A1 of Statement 2. For this, we need the concept of Mosco-convergence. Given a sequence of functionals  $\phi_n$  and  $\phi$  defined on a space  $X$  with weak and strong topology,  $\phi_n$  is said to *Mosco-converge* to  $\phi$  ( $\phi_n \xrightarrow{M} \phi$ ) in the weak-strong topology of  $X$  if

$$\begin{cases} \forall x_n \rightharpoonup x \text{ weakly,} & \liminf \phi_n(x_n) \geq \phi(x), \\ \forall x \exists x_n \rightarrow x \text{ strongly such that} & \limsup \phi_n(x_n) \leq \phi(x). \end{cases} \tag{20}$$

The gradient-flow Definition 1 is based on the function space  $X := AC(0, T)$ , and we define weak and strong topologies on the space  $AC(0, T)$  by using the equivalence with  $W^{1,1}(0, T)$ . Let  $x, x_n \in AC(0, T)$ . We say that  $x_n$  converges weakly to  $x$  ( $x_n \rightharpoonup x$ ) if  $x_n \rightarrow x$  strongly in  $L^1(0, T)$  and  $\dot{x}_n \rightharpoonup \dot{x}$  weakly in  $L^1(0, T)$ , i.e. in  $\sigma(L^1, L^\infty)$ ; we say that  $x_n$  converges strongly to  $x$  ( $x_n \rightarrow x$ ) if in addition  $\dot{x}_n \rightarrow \dot{x}$  strongly in  $L^1(0, T)$ .

**Theorem 2** (Convergence to the quadratic limit) *Given  $E : \mathbb{R} \times [0, T] \rightarrow \mathbb{R}$  satisfying condition (11), for  $x \in AC(0, T)$  consider the functional  $\mathcal{J}_\beta$*

$$\mathcal{J}_\beta(x) = \int_0^T \left( \psi_\beta(\dot{x}(t)) + \psi_\beta^*(-\nabla E(x(t), t)) + \dot{x}(t)\nabla E(x(t), t) \right) dt,$$

then, for  $\alpha \rightarrow \infty$ ,  $\beta \rightarrow 0$ , and  $\alpha\beta \rightarrow \omega > 0$ ,  $\mathcal{J}_\beta \xrightarrow{M} \mathcal{J}_Q$  in the weak-strong topology of  $AC(0, T)$ , with

$$\mathcal{J}_Q(x) := \int_0^T \left( \frac{\dot{x}^2(t)}{4\omega} + \omega(\nabla E)^2(x(t), t) + \dot{x}(t)\nabla E(x(t), t) \right) dt. \tag{21}$$

Moreover, if a sequence  $\{x_\beta\}$  is such that  $\mathcal{J}_\beta(x_\beta)$  is bounded, and

$$E(x_\beta(0), 0) + \int_0^T \partial_t E(x_\beta(s), s) ds \leq C \quad \forall \beta,$$

then the sequence  $\{\dot{x}_\beta\}$  is relatively compact in the topology  $\sigma(L^1, L^\infty)$ .

*Proof* First, we prove the Mosco-convergence. The limsup condition follows because, for  $\beta \rightarrow 0$

$$\psi_\beta(\eta) \rightarrow \frac{\eta^2}{4\omega}, \quad \psi_\beta^*(\xi) \rightarrow \omega\xi^2, \quad \text{locally uniformly.}$$

By the local uniform convergence, we can choose the recovery sequence to be the trivial one.

Now we prove the liminf inequality. The uniform convergence of  $x_\beta$  to  $x$  implies that we can pass to the limit in the terms  $E(x_\beta(\cdot), \cdot)$  and  $\int \partial_t E(x_\beta, t) dt$ . Then, applying Fatou's lemma, we find

$$\begin{aligned} \liminf_{\beta \rightarrow 0} \int_0^T \frac{2\alpha}{\beta} (\cosh(\beta\nabla E(x_\beta, t)) - 1) dt &\geq \liminf_{\beta \rightarrow 0} \int_0^T \alpha\beta(\nabla E)^2(x_\beta, t) dt \\ &\geq \int_0^T \omega(\nabla E)^2(x, t) dt, \end{aligned}$$

where we used the inequality  $2 \cosh(\theta) \geq 2 + \theta^2$ .

Some algebraic manipulations are needed to estimate the lim inf for  $\psi_\beta$ . Let  $v \geq 0$

$$\begin{aligned} \psi_\beta(v) &= \frac{v}{\beta} \log \left( \frac{v + \sqrt{v^2 + 4\alpha^2}}{2\alpha} \right) - \frac{1}{\beta} \sqrt{v^2 + 4\alpha^2} + \frac{2\alpha}{\beta} \\ &= \frac{v}{\beta} \log \left( \frac{v}{2\alpha} + \sqrt{\frac{v^2}{4\alpha^2} + 1} \right) - \frac{2\alpha}{\beta} \sqrt{\frac{v^2}{4\alpha^2} + 1} + \frac{2\alpha}{\beta} \\ &\geq \frac{v}{\beta} \left( \frac{v}{2\alpha} - \frac{v^3}{48\alpha^3} \right) - \frac{2\alpha}{\beta} \left( 1 + \frac{v^2}{8\alpha} \right) + \frac{2\alpha}{\beta} = \frac{v^2}{4\alpha\beta} - \frac{v^4}{48\alpha^3\beta}, \end{aligned}$$

where we used that

$$\log \left( v + \sqrt{v^2 + 1} \right) > v - \frac{v^3}{6}, \quad \text{and} \quad \sqrt{1 + v} < 1 + \frac{v}{2}.$$

Then, considering that  $\alpha^3\beta \rightarrow \infty$ , we get

$$\liminf_{\beta \rightarrow 0} \int_0^T \psi_\beta(\dot{x}_\beta) dt \geq \liminf_{\beta \rightarrow 0} \int_0^T \left( \frac{\dot{x}_\beta^2}{4\alpha\beta} - \frac{\dot{x}_\beta^4}{48\alpha^3\beta} \right) dt \geq \int_0^T \frac{\dot{x}^2}{4\omega} dt,$$

and we conclude.

Now we prove the compactness.

Let us suppose that  $\mathcal{J}_\beta(x_\beta)$  and  $E(x_\beta(0), 0) + \int_0^T \partial_t E(x_\beta, t) dt$  are bounded. Then, by the positivity of  $\psi_\beta^*$ ,

$$\int_0^T \psi_\beta(\dot{x}_\beta) dt \leq C < \infty, \quad \forall \beta.$$

With the choice  $\bar{\beta} = 1$ , we have

$$\int_0^T \psi_1(\dot{x}_\beta) dt \leq C \quad \forall \beta \leq 1,$$

and the compactness of  $\{\dot{x}_\beta\}$  in  $\sigma(L^1, L^\infty)$  follows from the Dunford–Pettis theorem (e.g. [10, Th. 4.30]).  $\square$

Note that this result can also be obtained by the abstract method of Mielke [29, Th. 3.3]. Also note that the result can also be formulated in the weak-strict convergence of  $BV$  (see the definition after Eq. (25)); for the lower semicontinuity, this follows since the weak convergence in  $BV$  with bounded  $\mathcal{J}_\beta$  implies weak convergence in  $AC$ , and for the recovery sequence, it follows from our choice of the trivial sequence.

We end this section with two theorems completing the proof of Statement 3, pictured in the right-hand side of Fig. 3.

First, we define for each  $h > 0$  the SDE

$$dY_t^h = -2\omega \nabla E(Y_t^h, t) dt + \sqrt{2\omega h} dW_t, \tag{22}$$

where  $W_t$  is the Brownian motion on  $\mathbb{R}$ ,  $Y_0^h$  has law  $\delta_{x_0}$ , and its generator  $\Omega$  is defined as

$$\Omega f(x) = -2\omega \nabla E(x, t) \nabla f(x) + \omega h \Delta f(x). \tag{23}$$

For  $f \in C_b(\mathbb{R})$ , the expected value  $\mathbb{E}$  is defined as

$$\mathbb{E}(f(Y_t^h)|x_0) = \int_{\mathbb{R}} f(z) d\mu_t(z),$$

with, denoting by  $\Omega^T$  the adjoint of  $\Omega$ ,

$$\begin{cases} \partial_t \mu_t = \Omega^T \mu_t = -2\omega \nabla \cdot (\mu_t \nabla E) + \omega h \Delta \mu_t, \\ \mu_0 = \delta_{x_0}. \end{cases}$$

where  $\mu_t$  is the law at time  $t$  of the process  $Y^h$  started at the position  $x_0$  at time  $t = 0$ . Then, the following theorems hold.

**Theorem 3** (Large deviations for the processes  $X^n$  and  $Y^h$ ) *Given an energy  $E$  satisfying condition (11), fix  $0 < \delta < 1$  and consider the sequence of processes  $\{X^n\}$  with generator  $\Omega_n$  defined in (12) with  $\beta = \beta_n = n^{-\delta}$  and  $\alpha\beta = \alpha_n\beta_n \rightarrow \omega$  for  $n \rightarrow \infty$ . Then, if  $X^n(0) \rightarrow x_0$ , the process  $X^n$  satisfies a large-deviations principle in  $D([0, T])$  with speed  $n^{1-\delta}$  and with rate function the extension of  $\mathcal{J}_Q$  in (21) to  $BV$ :*

$$\mathcal{J}_Q(x) := \int_0^T \left( \frac{\dot{x}^2(t)}{4\omega} + \omega (\nabla E)^2(x(t), t) + \dot{x}(t) \nabla E(x(t), t) \right) dt,$$

when  $x \in AC(0, T)$  and  $\mathcal{J}_Q(x) = +\infty$  otherwise.

Moreover, as  $h \rightarrow 0$  the process  $Y^h$  defined in (22) satisfies a large-deviations principle in  $D([0, T])$  with speed  $h^{-1}$  and also with rate function  $\mathcal{J}_Q$ .

*Proof* The proof of the large-deviations principle for  $X^n$  relies on the fulfilment of three conditions, namely convergence of the operators  $H_n$ , exponential tightness for the sequence of processes  $X^n$ , and the comparison principle for the limiting operator  $H$ , following the steps of [20, Sec. 10.3].

We restrict ourselves, for the sake of simplicity, to the case  $\alpha = \omega/\beta$ , and let  $m = n^{1-\delta}$ . To treat the time dependence, we use the standard procedure of converting a time-dependent process into a time-independent process by adding the time to the state variable (see e.g. [19, Sec. 4.7]): consider the variable  $u = (x, t) \in \mathbb{R} \times [0, T]$ ,  $f \in C_c^2(\mathbb{R} \times [0, T])$ , and given  $\Omega_n$  defined in (12), we define  $Q_n$  as

$$Q_n f(u) = \Omega_n f(x, t) + \partial_t f(x, t). \tag{24}$$

With  $m = n^{1-\delta}$ , we have,

$$H_n f(u) = \frac{1}{m} e^{-mf(u)} Q_n e^{mf(u)} = \frac{1}{m} e^{-mf(u)} Q_n e^m f(u) + \partial_t f(u).$$

Now, with the convention  $u + 1/n = (x + 1/n, t)$ ,

$$\begin{aligned} H_n f(u) &= \omega n^{2\delta} \left\{ e^{-n^{-\delta} \nabla E(u)} \left( e^{n^{1-\delta}(f(u+1/n)-f(u))} - 1 \right) \right. \\ &\quad \left. + e^{n^{-\delta} \nabla E(u)} \left( e^{n^{1-\delta}(f(u-1/n)-f(u))} - 1 \right) \right\} + \partial_t f \\ &= \omega n^{2\delta} \left[ (1 - n^{-\delta} \nabla E + o(n^{-\delta})) \left( n^{-\delta} \nabla f(u) + n^{-2\delta} \frac{1}{2} (\nabla f)^2(u) + o(n^{-2\delta}) \right) \right. \\ &\quad \left. + (1 + n^{-\delta} \nabla E + o(n^{-\delta})) \left( -n^{-\delta} \nabla f(u) + n^{-2\delta} \frac{1}{2} (\nabla f)^2(u) + o(n^{-2\delta}) \right) \right] + \partial_t f \\ &= -2\omega \nabla E(u) \nabla f(u) + \omega (\nabla f)^2(u) + \partial_t f(u) + o(1), \end{aligned}$$

implying convergence in the uniform topology,

$$\lim_{n \rightarrow \infty} \|H_n f - Hf\|_\infty = 0,$$

to

$$Hf(u) = -2\omega \nabla E(u) \nabla f(u) + \omega (\nabla f)^2(u) + \partial_t f(u).$$

The *exponential tightness* holds by [20, Cor. 4.17]. The *comparison principle* can be proved as in [20, Example 6.11], modifying the definition of the auxiliary function  $\Phi_n$  used in the cited example, with an additional time-dependent term

$$\Phi_n(x, y, t, \tau) = \overline{f}(x, t) - \underline{f}(y, \tau) - n \frac{(x - y)^2}{1 + (x - y)^2} - n(t - \tau)^2,$$

and then the proof, mutatis mutandis, follows similarly.

Then, the large-deviations principle holds in  $D_{\mathbb{R} \times [0, T]}([0, T])$  with rate functional

$$J_Q(u) := \int_0^T L(x, \dot{x}, t, i) ds,$$

where  $L$  is the Legendre transform of  $H$  respect to the variables  $(\nabla f, \partial_t f)$ . It is just a calculation to check that

$$L(x, \dot{x}, t, i) = \frac{\dot{x}^2(s)}{2} + \nabla E(x(s), t(s)) \dot{x}(s) + \frac{\nabla E^2(x(s), t(s))}{2} + \mathbb{I}_1(i(s)),$$

where  $\mathbb{I}_1$  is the indicator function of the set  $\{1\}$ , i.e.

$$\mathbb{I}_1(i(s)) = \begin{cases} 0 & i(s) = 1, \\ +\infty & \text{otherwise.} \end{cases}$$

It is then clear that  $J_Q(u) = \mathcal{J}_Q(x)$ .

The large-deviations result for  $Y^h$  can be found in [23, Th. 1.1 of Ch. 4] in the case of a time-independent energy. The time-dependent case follows by the same modification as above.  $\square$

**Theorem 4** (Convergence to Brownian motion with gradient drift) *Let be given an energy  $E$  satisfying condition (11), with  $\nabla E$  uniformly continuous, let  $\alpha_n \beta_n \rightarrow \omega$ ,  $n \beta_n \rightarrow 1/h$ , and let  $\mu_t^n$  be the law of the process  $\{X^n(t)\}$  defined in (12) with  $\mu_0^n = \delta_{x_0^n}$ . If  $X_0^n \rightarrow x_0$ , then  $\mu_n$  weakly converge to  $\mu$  (in the duality with  $C_b(\mathbb{R})$ ), where  $\mu$  is the law of the Brownian motion with gradient drift (22) with  $\mu_0 = \delta_{x_0}$ .*

*Proof* This is a result of standard type, and we give a brief sketch of the proof for the case of time-independent  $E$ , using the semigroup convergence theorem of Trotter [45, Th. 5.2]. The assumptions of this theorem are satisfied by the existence of a single dense set on which  $\Omega$  and  $\Omega_n$  are defined, pointwise convergence of  $\Omega_n$  to  $\Omega$  on that dense set, and a dense range of  $\lambda - \Omega$  for sufficiently large  $\lambda$ . The assertion of Trotter’s theorem is pointwise convergence of the corresponding semigroups at each fixed  $t$ , which implies convergence of the dual semigroups in the dual topology, which is the statement of Theorem 4.

We set the system up as follows. Define the state space  $Y := C_b(\overline{\mathbb{R}})$  with the uniform norm, where  $\overline{\mathbb{R}}$  is the one-point compactification of  $\mathbb{R}$ ; define the core  $D := \{f \in C_b^2(\mathbb{R}) \cap C(\overline{\mathbb{R}}) : \Delta f \text{ uniformly continuous}\}$ , which is dense in  $Y$  for the uniform topology, and which will serve as the dense set of definition mentioned above for both  $\Omega_n$  and  $\Omega$ . For each  $f \in D$ ,  $\Omega_n f \rightarrow \Omega f$  in the uniform topology.

The density of the range of  $\lambda - \Omega$  is the solvability in  $D$  of the equation

$$-\omega h \Delta f + 2\omega \nabla E \nabla f + \lambda f = g, \quad \text{in } \mathbb{R},$$

for all  $g$  in a dense subset of  $Y$ ; we choose  $g \in C_c(\mathbb{R}) + \mathbb{R}$ . This is a standard result from PDE theory, which can be proved for instance as follows. First note that we can assume  $g \in C_c(\mathbb{R})$ , by adding a constant to both  $g$  and  $f$ . Secondly, for sufficiently large  $\lambda > 0$  the left-hand side generates a coercive bilinear form in  $H^1(\mathbb{R})$  in the sense of the Lax–Milgram lemma, and therefore, there exists a unique solution  $f \in H^1(\mathbb{R})$ . By bootstrap arguments, using the continuity and boundedness of  $\nabla E$ , we find  $f \in C_b^2(\mathbb{R})$ , and since  $f \in H^1(\mathbb{R}) \cap C_b^2(\mathbb{R})$ ,  $f(x)$  tends to zero at  $\pm\infty$ , implying that  $f \in C_b^2(\mathbb{R}) \cap C(\overline{\mathbb{R}})$ . Finally, since  $\nabla E$  is uniformly continuous, the same holds for  $\Delta f$ . This concludes the proof.  $\square$

### 4 Rate-independent limit

In this section, we prove point A2 of Statement 2 and the whole of Statement 3. We will prove point A2 with a theorem that holds in greater generality, without assuming the explicit form of the couple  $\psi_\beta, \psi_\beta^*$ , but only a few ‘reasonable’ assumptions and the limiting behaviour.

We are therefore in the regime where  $\beta \rightarrow \infty, \log \alpha = -\beta A$  for some  $A > 0$ .

#### 4.1 Functions of bounded variation and rate-independent systems

We now briefly recall the definition of the BV space of functions with bounded variation, following the notation of [33]. A full description of this space and its properties can be found in [4]. Given a function  $x : [0, T] \rightarrow \mathbb{R}$ , the total variation of  $x$  in the interval  $[0, T]$  is defined by

$$\text{Var}(x, [0, T]) := \sup \left\{ \sum_{j=1}^n |x(t_j) - x(t_{j-1})| : 0 = t_0 < \dots < t_n = T \right\}.$$

We say that  $x \in BV([0, T])$  if  $\text{Var}(x, [0, T]) < \infty$ . The function  $x$  then admits left and right limits  $x(t_-)$  and  $x(t_+)$  in every point  $t \in [0, T]$ , and we define the jump set of  $x$  as

$$J_x := \{t \in [0, T] : x(t_-) \neq x(t) \text{ or } x(t) \neq x(t_+)\},$$

and the pointwise variation in the jump set as

$$\text{Jmp}(x, [0, T]) := \sum_{t \in J_x} (|x(t_-) - x(t)| + |x(t) - x(t_+)|). \tag{25}$$

The total variation admits the representation

$$\text{Var}(x, [0, T]) = \int_0^T |\dot{x}(t)| dt + \int_0^T d|Cx| + \text{Jmp}(x, [0, T]),$$

where  $|\dot{x}|$  is the modulus of the absolutely continuous (a.c.) part of the distributional derivative of  $x$ ; the measure  $|Cx|$  is the Cantor part and  $\text{Jmp}$  represents the contribution of the (at most countable) jumps.

Given a sequence  $\{x_n\} \subset BV([0, T])$ , we again define two notions of convergence. We say that  $x_n$  *weakly* converges to  $x$  ( $x_n \rightharpoonup x$ ) if  $x_n(t)$  converges to  $x(t)$  for every  $t \in [0, T]$  and the variation is uniformly bounded, i.e.  $\sup_n \text{Var}(x_n, [0, T]) < \infty$ . We say that  $x_n$  *strictly* converges to  $x$  ( $x_n \rightarrow x$ ) if  $x_n \rightharpoonup x$  and in addition  $\text{Var}(x_n, [0, T])$  converges to  $\text{Var}(x, [0, T])$  as  $n \rightarrow \infty$ .

According to the general set-up of [32,33], we define the notion of rate-independent system based on an energy balance similar to equation (9), where now the dissipation  $\psi$  has a linear growth, i.e.  $\psi(\eta) = \psi_{RI}(\eta) = A|\eta|$  with  $A > 0$ .

We first define  $\text{Jmp}_E$ , which can be viewed as an energy-weighted jump term, as

$$\text{Jmp}_E(x, [0, T]) = \sum_{t \in J_x} [\Delta(x(t_-), x(t)) + \Delta(x(t_+), x(t))],$$

where using the notation  $a \vee b := \max\{a, b\}$

$$\Delta(x_0, x_1) := \inf_{\theta \in AC(0,1)} \left\{ \int_0^1 (|\nabla E(t, \theta(\tau))| \vee A) |\dot{\theta}(\tau)| d\tau : \theta(0) = x_0, \theta(1) = x_1 \right\}. \quad (26)$$

The relation between the definitions (26) and (25) becomes clear in the following inequality

$$\text{Jmp}_E(x, [0, T]) \geq A \text{Jmp}(x, [0, T]),$$

where equality can be achieved depending on the behaviour of  $E$ , e.g. trivially when  $|\nabla E(x, t)| \leq A$  for every  $x, t$ . Then, we can interpret  $\text{Jmp}_E$  as a modified jump term, with an  $E$ -dependent weight.

In analogy with the (generalized) gradient flow Definition 1, we define rate-independent systems. There is no unique way to define solutions for a rate-independent system. The so-called *energetic solutions* have been introduced and analysed in [35–37] and are based on the combination of a pointwise global minimality property and an energy balance. Here we concentrate on *BV solutions*, as defined in [32,34]. Our limiting system will be of this type.

Fix  $A > 0$ , the rate-independent dissipation  $\psi_{RI}$  and its Legendre transform  $\psi_{RI}^*$  are

$$\psi_{RI}(v) = A|v|, \quad \psi_{RI}^*(w) = \begin{cases} 0 & w \in [-A, A], \\ +\infty & \text{otherwise.} \end{cases} \quad (27)$$

**Definition 3** (*Rate-independent evolution, in the BV sense* [32,34]) Given an energy  $E : \mathbb{R} \times [0, T] \rightarrow \mathbb{R}$ , continuously differentiable, and  $A > 0$ , a curve  $x \in BV([0, T])$  is a *rate-independent evolution* of  $E$  in  $[0, T]$  if it satisfies the energy balance

$$\begin{aligned} & \int_0^T (\psi_{RI}(\dot{x}(t)) + \psi_{RI}^*(-\nabla E(x, t))) dt + A \int_0^T d|Cx| + \text{Jmp}_E(x, [0, T]) \\ & + E(x(T), T) - E(x(0), 0) - \int_0^T \partial_t E(x, t) dt = 0, \end{aligned} \quad (28)$$

with  $\psi_{RI}$  and  $\psi_{RI}^*$  are defined in (27).

## 4.2 Assumptions and the main result

In the rest of this section, we prove that the generalized gradient-flow evolution converges to the rate-independent one. This is point A2 of statement 2, formulated in Theorem 5 showing Mosco-convergence of  $\mathcal{J}_\beta$  to  $\mathcal{J}_{RI}$ , which is the left-hand side of (28). There are three main reasons why the convergence to a rate-independent system should be expected.

First, from a heuristic mathematical point of view, our choices of  $\alpha$  and  $\beta$  yield pointwise convergence of  $\psi_\beta$  and  $\psi_\beta^*$  to a one-homogeneous function and to its dual, the indicator function; this suggests a rate-independent limit. However, this argument does not explain which of the several rate-independent interpretations the limit should satisfy, nor does it explain the additional jump term.

Secondly, from a physical point of view, the underlying stochastic model mimics a rate-independent system. This can be recognized by keeping the lattice size finite but letting  $\beta \rightarrow \infty$ ; then, the rates either explode or converge to zero, depending on the value of  $\nabla E$ . We can interpret this in the sense that when a rate is infinite, with probability one a jump will occur to the nearest lattice point with zero jump rate.

Thirdly, considering the evolution, in the case  $1 \ll \beta < \infty$  the generalized gradient flow will present fast transitions when  $|\nabla E| > A$ . By slowing down time during these fast transitions, we can capture what is happening at the small timescale of these fast transitions—which become jumps in the limit. This is exactly how we construct the recovery sequence in Theorem 5.

The convergence will be proven in a greater generality; more precisely, we do not use the explicit formulas, but we require that  $\psi_\beta$  and  $\psi_\beta^*$  satisfy the following conditions:



- A  $\psi_\beta$  and  $\psi_\beta^*$  are both symmetric, convex and  $C^1$ ;
- B  $\psi_\beta^*$  converge pointwise to  $\psi_{RI}^*(w) = \begin{cases} +\infty & \text{for } |w| > A, \\ 0 & \text{for } |w| \leq A; \end{cases}$
- C  $\forall M > 0 \exists \delta_\beta \rightarrow 0$  such that as  $\beta \rightarrow \infty$ ,

$$K_\beta^{-1} := \partial\psi_\beta^*(A + \delta_\beta) \rightarrow \infty,$$

$$\sup_{|w| \leq R} \frac{\partial\psi_\beta^*(w + MK_\beta)}{\partial\psi_\beta^*(w \vee (A + \delta_\beta))} K_\beta \rightarrow 0, \quad \text{and}$$

$$\partial\psi_\beta^*(A + MK_\beta) K_\beta \rightarrow 0.$$

- D For each  $\alpha \geq 1$  and for each  $|w| \leq R$ , there exists  $\eta_\beta(w, \alpha) \geq 0$  such that

$$\partial\psi_\beta^*(w + \eta_\beta(w, \alpha)) = \alpha \partial\psi_\beta^*(w),$$

and  $\eta_\beta$  is bounded uniformly in  $\alpha, \beta$ , and  $|w| \leq R$ .

It is important to underline that the previous conditions C–D are needed in Theorem 5 only for the  $\Gamma$ -limsup, meanwhile they are not necessary for the  $\Gamma$ -liminf.

Conditions C–D are very technical and not intuitive, but they are satisfied by a large family of pairs  $(\psi_\beta, \psi_\beta^*)$ . In the following we show, with two examples, that our specific case and the vanishing-viscosity approach, respectively, are covered by the assumptions A–D. In the two examples, we implicitly underline the novelty of our approach and the difference between our model (with  $L \log L$ -type dissipation) versus the vanishing viscosity (see [34]).

*Dissipation (16):*  $\psi_\beta^*(w) = \beta^{-1} e^{-\beta A} \cosh(\beta w)$

Conditions A and B are trivially satisfied. Then, considering only  $w \geq A$  for simplicity, we get

$$\partial\psi_\beta^*(w) = e^{-\beta A} \sinh(\beta w) \simeq e^{\beta(w-A)}.$$

With the choice  $\delta_\beta = \beta^{-1} \log(\beta)$ , it holds that

$$K_\beta \simeq \beta^{-1}, \quad \partial\psi_\beta^*(A + MK_\beta) K_\beta \leq e^{\beta MK_\beta} K_\beta \rightarrow 0.$$

Then, condition C is satisfied with

$$\frac{\partial\psi_\beta^*(w + K_\beta)}{\partial\psi_\beta^*(w)} K_\beta \leq \frac{\exp(\beta(w + K_\beta - A)) + 1}{\exp(-\beta(A - w))} K_\beta \leq (\exp(\beta K_\beta) + 1) K_\beta \rightarrow 0.$$

Condition D is satisfied because for  $w \gg 1$  we have that  $\sinh(\beta w) \simeq \frac{1}{2} e^{\beta w}$ . Then, condition D approximately reads as

$$e^{\beta(w+\eta)} \simeq \alpha e^{\beta w},$$

which is satisfied for  $\eta \simeq \beta^{-1} \log \alpha$ .

*Viscoplasticity with vanishing viscosity:*  $\psi_\beta^*(w) = \beta(|w| - A)_+^2$

Also here, conditions A and B are immediately satisfied. Then, again considering  $w \geq A$ , we verify condition C by choosing  $\delta_\beta \simeq \beta^{-1/3}$  and  $\lambda = 1$ , so that

$$\partial\psi_\beta^*(w) = 2\beta(w - A) \implies K_\beta^{-1} = 2\beta\delta_\beta \rightarrow \infty.$$

Then, it is just a calculation to check that condition C is satisfied in this case. Now condition D requires that

$$2\beta(w + \eta - A) = 2\alpha\beta(w - A),$$

and so  $\eta = (\alpha - 1)(w - A)$  satisfies the condition.

**Theorem 5** (Convergence to the rate-independent evolution) *Given an energy  $E$  satisfying condition (11), a sequence of couple  $\psi_\beta - \psi_\beta^*$  satisfying conditions A–D, and for  $x \in BV([0, T])$  consider the functional  $\mathcal{J}_\beta$*

$$\mathcal{J}_\beta(x) = \int_0^T \left( \psi_\beta(\dot{x}(t)) + \psi_\beta^*(-\nabla E(x(t), t)) + \dot{x}(t)\nabla E(x(t), t) \right) dt \tag{29}$$

when  $x \in AC(0, T)$  and  $+\mathcal{J}_\beta(x) = +\infty$  otherwise. Then, as  $\beta \rightarrow \infty$ ,  $\mathcal{J}_\beta \xrightarrow{M} \mathcal{J}_{RI}$  with respect to the weak-strict topology of  $BV$ , where  $\mathcal{J}_{RI}$  is given by

$$\begin{aligned} \mathcal{J}_{RI}(x) := & \int_0^T \left( \psi_{RI}(\dot{x}(t)) + \psi_{RI}^*(-\nabla E(x(t), t)) \right) dt + A \int_0^T d|Cx| \\ & + \text{Imp}_E(x, [0, T]) + E(x(T), T) - E(x(0), 0) - \int_0^T \partial_t E(x(t), t) dt. \end{aligned} \tag{30}$$

Moreover, if the sequence  $\{x_\beta\}$  is such that  $\mathcal{J}_\beta(x_\beta)$  is bounded,  $\{x_\beta(0)\}$  is bounded, and

$$\int_0^t \partial_t E(x_\beta(s), s) ds \leq C \quad \forall \beta, t \in [0, T], \tag{31}$$

then  $\{x_\beta\}$  is weakly compact in  $BV([0, T])$ .

The proof is divided into three main steps. We first prove the compactness and the liminf inequality; this will follow as in [33, Th. 4.1,4.2]. We report them for completeness, and we translate their proof because we can avoid some technicalities. To finish the proof, we need to construct a recovery sequence. When minimizers with  $\mathcal{J}_{RI} \equiv 0$  are considered, then the recovery sequence is easy to construct; we just need to take a sequence  $x_\beta$  such that  $\mathcal{J}_\beta(x_\beta) = 0$  for every  $\beta$ . But for the full Mosco-convergence, we need to find a way to construct a recovery sequence also for non-minimizers of  $\mathcal{J}_{RI}$ . This is the last part of the proof, and it will be achieved using a parametrized-solution technique.

### 4.3 Proof of compactness and the lower bound

*Proof (Proof of compactness)* Recall that weak convergence in  $BV$  is equivalent to pointwise convergence supplemented with a global bound on the total variation (e.g. [4, Prop. 3.13]).

First, we show that  $|x_\beta(t) - x_\beta(0)|$  is bounded uniformly in  $t$  and  $\beta$ . We observe that

$$\psi_\beta(v) + \psi_\beta^*(A) \geq Av, \quad \text{for every } v, A \in \mathbb{R}, \tag{32}$$

and so we obtain

$$A|x_\beta(t) - x_\beta(0)| \leq A \int_0^t |\dot{x}_\beta| ds \leq \int_0^t \psi_\beta(\dot{x}_\beta) ds + t \sup_\beta \psi_\beta^*(A) \leq C < +\infty,$$

where the constant  $C$  may change from line to line. Then,

$$|x_\beta(t) - x_\beta(0)| \leq C \quad \text{for every } \beta \text{ and every } t \in [0, T].$$

The inequality above and the boundedness of  $x_\beta(0)$  imply that the whole sequence is bounded for every  $t \in [0, T]$ .

Next we show the existence of a converging subsequence.

For every  $0 \leq t_0 \leq t_1 \leq T$ , we recall the bound

$$A|x_\beta(t_1) - x_\beta(t_0)| \leq \int_{t_0}^{t_1} A|\dot{x}_\beta| dt \leq \int_{t_0}^{t_1} \left( \psi_\beta(\dot{x}_\beta) + \psi_\beta^*(A) \right) dt.$$

Defining the non-negative finite measures on  $[0, T]$

$$\nu_{\beta,A} := \left( \psi_\beta(\dot{x}_\beta) + \psi_\beta^*(A) \right) \mathcal{L}^1,$$

up to extracting a suitable subsequence, we can suppose that they weakly converge to a finite measure  $\nu_A$ , so that

$$A \limsup_{\beta \rightarrow \infty} |x_\beta(t_0) - x_\beta(t_1)| \leq \limsup_{\beta \rightarrow \infty} \nu_{\beta,A}([t_0, t_1]) \leq \nu_A([t_0, t_1]).$$

Defining the jump set  $J := \{t \in [0, T] : \nu_A(\{t\}) > 0\}$  and considering a countable set  $I \supset J$  that is dense in  $[0, T]$ , we can find a subsequence  $\beta_h$  such that  $x_{\beta_h} \xrightarrow{p.w.} x$  for every  $t \in I$  as  $\beta_h \rightarrow \infty$ . From now on, for simplicity, we will number the subsequence with the same index of the main sequence. Then,

$$A|x(t_0) - x(t_1)| \leq \nu_A([t_0, t_1]), \quad \text{for every } t_0, t_1 \in I. \tag{33}$$

The curve  $I \ni t \mapsto x(t)$  can be uniquely extended to a continuous curve in  $[0, T] \setminus J$ , that we will still denote by  $x$ . Arguing by contradiction, we show that the whole  $x_\beta(t)$  converges pointwise to  $x(t)$ . Fix a point  $t \in [0, T] \setminus I$ , if the pointwise convergence does not hold, we can find a subsequence that we still denote  $x_\beta(t)$  such that  $x_\beta(t) \rightarrow \tilde{x} \neq x(t)$ . From the pointwise convergence for  $t \in I$  and the continuity of  $x$  in  $[0, T] \setminus J$ , we can find a further subsequence  $I \ni t_{\beta_n} \rightarrow t$  such that  $x_{\beta_n}(t_{\beta_n}) \rightarrow x(t)$ , but this is in contradiction to the previous inequality, assuming for simplicity  $t_{\beta_n} \leq t$ ,

$$A|x(t) - \tilde{x}| \leq \liminf_{\beta_n \rightarrow \infty} A|x_{\beta_n}(t_{\beta_n}) - x_{\beta_n}(t)| \leq \limsup_{\beta_n \rightarrow \infty} \nu_{\beta_n,A}([t_{\beta_n}, t]) = \nu_A(\{t\}) = 0.$$

We have so proven the pointwise convergence of  $x_\beta$  to  $x$ ; the inequality (33) then gives a uniform bound on the BV norm of  $x_\beta$  and so we conclude. □

*Proof (Proof of the liminf inequality)* Let  $\{x_\beta\} \subset AC(0, T)$  be a sequence such that  $\mathcal{J}_\beta(x_\beta)$  is bounded, which converges weakly to  $x \in BV([0, T])$ . By the arguments above,  $x_\beta$  is bounded uniformly in  $t$  and  $\beta$  and every term in the functional is bounded itself by a constant independent of  $\beta$ . The following limits, follow from the pointwise convergence and Lebesgue’s dominated convergence theorem:

$$E(x_\beta(\cdot), \cdot) \rightarrow E(x(\cdot), \cdot), \quad \int_0^T \partial_t E(x_\beta(t), t) dt \rightarrow \int_0^T \partial_t E(x(t), t) dt.$$

As we said, the integral  $\int_0^T \psi_\beta^*(-\nabla E(x_\beta(t), t))dt$  is bounded for every  $\beta$  by a constant that we still denote by  $C$ . Because of the monotonicity of  $\psi_\beta^*$ , this bound implies that

$$\psi_\beta^*(a)\mathcal{L}^1\{t \in (0, T) : |\nabla E(x_\beta(t), t)| \geq a\} \leq K \quad \forall a \geq 0,$$

and since  $\psi_\beta^*(w) \rightarrow +\infty$  for  $|w| > A$ , we obtain

$$\lim_{\beta \rightarrow \infty} \mathcal{L}^1\{t \in (0, T) : |\nabla E(x_\beta(t), t)| \geq a\} = 0 \quad \forall a > A.$$

This proves that  $|\nabla E(x(t), t)| \leq A$  a.e. and therefore  $\int_0^T \psi_{RI}^*(\nabla E(x(t), t)) dt = 0$ ; it trivially follows that

$$\liminf_{\beta \rightarrow \infty} \int_0^T \psi_\beta^*(\nabla E(x_\beta(t), t)) dt \geq \int_0^T \psi_{RI}^*(\nabla E(x(t), t)) dt.$$

We now prove the second part of the inequality,

$$\liminf_{\beta \rightarrow \infty} \int_0^T (\psi_\beta(\dot{x}) + \psi_\beta^*(\nabla E)) dt \geq \int_0^T A|\dot{x}| dt + A \int_0^T d|Cx| + \text{Jmp}_E(x, [0, T]).$$

As in the proof of the compactness, we consider the non-negative finite measure on  $[0, T]$

$$\nu_\beta := (\psi_\beta(\dot{x}_\beta) + \psi_\beta^*(\nabla E(x_\beta, \cdot)))\mathcal{L}^1,$$

up to extracting a subsequence, we can suppose that they weakly converge to a finite measure

$$\nu_0 + \psi_{RI}^*(\nabla E(x, \cdot))\mathcal{L}^1.$$

Because  $|\nabla E(x, \cdot)| \leq A$   $\mathcal{L}$ -a.e. we obtain that, as in the proof of the compactness,

$$v_0 \geq A(|\dot{x}| + |Cx| + |Jx|) = \psi_{RI}(\dot{x}) + A(|Cx| + |Jx|).$$

This inequality is slightly too weak for us. The Cantor and the Lebesgue measurable parts are fine, but we need a stronger characterization of the jump part: for all  $t \in J_x$ ,

$$v_0(\{t\}) \geq \Delta(x(t_-), x(t)) + \Delta(x(t), x(t_+)).$$

To prove this, fix  $t \in J_x$  and take two sequences  $h_\beta^- < t < h_\beta^+$  converging monotonically to  $t$  such that

$$x_\beta(h_\beta^-) \rightarrow x(t_-), \quad x_\beta(h_\beta^+) \rightarrow x(t_+),$$

and define

$$s_\beta(h) := h + \int_t^h \left( \psi_\beta(\dot{x}_\beta(\tau)) + \psi_\beta^*(\nabla E(x_\beta(\tau), \tau)) \right) d\tau; \quad s_\beta^\pm := s_\beta(h_\beta^\pm).$$

Because of the convergence of  $v_\beta$ , we have

$$\limsup_{\beta \rightarrow \infty} (s_\beta^+ - s_\beta^-) \leq \limsup_{\beta \rightarrow \infty} v_\beta([s_\beta^-, s_\beta^+]) \leq v_0(\{t\}),$$

and up to extracting a subsequence we can assume that  $s_\beta^\pm \rightarrow s^\pm$ . Denote by  $h_\beta := s_\beta^{-1}$  the inverse map of  $s_\beta$ , we observe that  $h_\beta$  is 1-Lipschitz and monotone, and it maps  $[s_\beta^-, s_\beta^+]$  onto  $[h_\beta^-, h_\beta^+]$ . We can then define the following Lipschitz functions

$$\theta_\beta(s) := \begin{cases} x_\beta(h_\beta(s)) & \text{if } s \in [s_\beta^-, s_\beta^+], \\ x_\beta(h_\beta^+) & \text{if } s \geq s_\beta^+, \\ x_\beta(h_\beta^-) & \text{if } s \leq s_\beta^-. \end{cases}$$

The functions  $\theta_\beta$  are uniformly Lipschitz, since (writing  $\tau = h_\beta(s)$ )

$$\begin{aligned} |\dot{\theta}_\beta(s)| &= |\dot{x}_\beta(h_\beta(s))| |\dot{h}_\beta(s)| \leq \frac{|\dot{x}_\beta(\tau)|}{1 + \psi_\beta(\dot{x}_\beta(\tau)) + \psi_\beta^*(\nabla E(x_\beta(\tau), \tau))} \\ &\stackrel{(32)}{\leq} \frac{\max\{c, 1\}}{A}, \end{aligned}$$

and they take the special values

$$\theta_\beta(s_\beta^\pm) = x_\beta(h_\beta^\pm), \quad \theta_\beta(t) = x_\beta(t).$$

Therefore, denoting by  $I$  a compact interval containing the intervals  $[s_\beta^-, s_\beta^+]$  for all  $\beta$ , then up to a subsequence, we have that

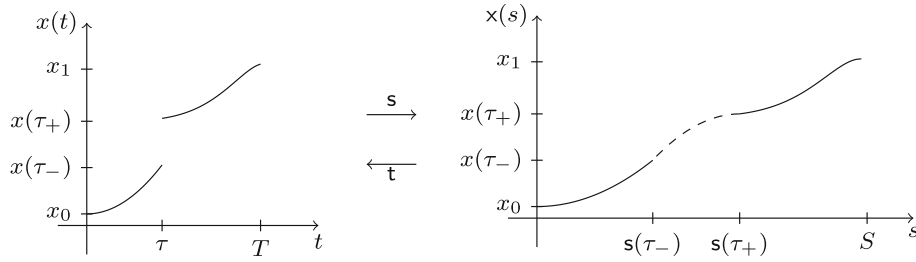
$$\theta_\beta(s) \rightarrow \theta(s), \quad |\dot{\theta}_\beta| \xrightarrow{*} m \text{ in } L^\infty(I) \quad \text{with } m \geq |\dot{\theta}|.$$

Moreover,  $\theta(s^\pm) = x(t_\pm)$  and  $\theta(t) = x(t)$ . Then, using the inequality

$$\psi_\beta(v) + \psi_\beta^*(w) \geq (|w| \vee A)v - \psi_\beta^*(A),$$

we obtain

$$\begin{aligned} v_0(\{t\}) &\geq \limsup_{\beta \rightarrow \infty} \int_{h_\beta^-}^{h_\beta^+} \left( \psi_\beta(\dot{x}_\beta(\tau)) + \psi_\beta^*(\nabla E(x_\beta(\tau), \tau)) \right) d\tau \\ &\geq \liminf_{\beta \rightarrow \infty} \int_{h_\beta^-}^{h_\beta^+} \left( (|\nabla E(x_\beta(\tau), \tau)| \vee A) |\dot{x}_\beta(\tau)| - \psi_\beta^*(A) \right) d\tau \\ &\geq \liminf_{\beta \rightarrow \infty} \int_{s_\beta^-}^{s_\beta^+} \left( (|\nabla E(\theta_\beta(s), h_\beta(s))| \vee A) |\dot{\theta}_\beta(s)| - (h_\beta^+ - h_\beta^-) \psi_\beta^*(A) \right) ds. \end{aligned}$$



**Fig. 5** Schematic representation of the time parametrization procedure. The curve  $\mathbf{x}$  is such that  $\mathbf{x}(s(t)) = x(t)$

The last term  $(h_\beta^+ - h_\beta^-)\psi_\beta^*(A)$  tends to zero as  $\beta \rightarrow \infty$ . Therefore,

$$\begin{aligned} v_0(\{t\}) &\geq \liminf_{\beta \rightarrow \infty} \int_I (|\nabla E(\theta_\beta(s), h_\beta(s))| \vee A) |\dot{\theta}_\beta|(s) \, ds \\ &\stackrel{(*)}{\geq} \int_I (|\nabla E(\theta(s), s)| \vee A) m(s) \, ds \geq \int_{s^-}^{s^+} (|\nabla E(\theta(s), s)| \vee A) |\dot{\theta}|(s) \, ds \\ &\geq \Delta(x(t_-), x(t)) + \Delta(x(t), x(t_+)). \end{aligned}$$

The inequality  $(*)$  follows from the technical Lemma [33, Lemma 4.3], and so we conclude.  $\square$

#### 4.4 Proof of the limsup inequality

*Proof* We assume that we are given  $x \in BV([0, T])$ ; we will construct a sequence  $x_\beta$  such that  $\mathcal{J}_\beta(x_\beta) \rightarrow \mathcal{J}_{RI}(x)$ .

*Reparametrization.* A central tool in this construction is a reparametrization of the curve  $x$  (as in Fig. 5), in terms of a new time-like parameter  $s$  on a domain  $[0, S]$ . The aim is to expand the jumps in  $x$  into smooth connections.

As in [32, Prop. 6.10], we define

$$\begin{aligned} \mathbf{s}(t) &:= t + \int_0^t (\psi_{RI}(\dot{x}) + \psi_{RI}^*(\nabla E(x, \tau))) \, d\tau + A \int_0^t d|Cx| \\ &\quad + \text{Jmp}_E(x, [0, t]), \end{aligned} \tag{34}$$

then there exists a Lipschitz parametrization  $(\mathbf{t}, \mathbf{x}) : [0, S] \rightarrow [0, T] \times \mathbb{R}$  such that  $\mathbf{t}$  is non-decreasing,

$$\mathbf{t}(\mathbf{s}(t)) = t, \quad \text{and} \quad \mathbf{x}(\mathbf{s}(t)) = x(t) \quad \text{for every } t \in [0, T], \tag{35}$$

and such that

$$\begin{aligned} &\int_0^S \mathbb{L}(\mathbf{x}, \mathbf{t}, \dot{\mathbf{x}}, \dot{\mathbf{t}}) \, ds \\ &= \int_0^T (\psi_{RI}(\dot{x}) + \psi_{RI}^*(\nabla E(x, \tau))) \, d\tau + A \int_0^T d|Cx| + \text{Jmp}_E(x, [0, T]), \end{aligned} \tag{36}$$

where recalling that  $\dot{\mathbf{t}} = 0$  in the jumps

$$\mathbb{L}(\mathbf{x}, \mathbf{t}, \dot{\mathbf{x}}, \dot{\mathbf{t}}) = \begin{cases} A|\dot{\mathbf{x}}| + \psi_{RI}^*(|\nabla E(\mathbf{x}, \mathbf{t})|) & \text{if } \dot{\mathbf{t}} > 0, \\ |\dot{\mathbf{x}}| (A \vee |\nabla E(\mathbf{x}, \mathbf{t})|) & \text{if } \dot{\mathbf{t}} = 0. \end{cases}$$

Moreover, it also holds that

$$\text{Var}(\mathbf{x}, [0, S]) = \text{Var}(x, [0, T]). \tag{37}$$

Note that  $\mathbb{L}(\mathbf{x}, \mathbf{t}, \dot{\mathbf{x}}, \dot{\mathbf{t}}) \geq |\dot{\mathbf{x}}| (A \vee |\nabla E(\mathbf{x}, \mathbf{t})|)$ , since  $\psi_{RI}^*(w)$  is only finite when  $|w| \leq A$ .

*Preliminary remarks.* The third term in  $\mathcal{J}_\beta(x_\beta)$  (see (29)) is equal to

$$E(x_\beta(T), T) - E(x_\beta(0), 0) - \int_0^T \partial_t E(x_\beta(t), t) dt,$$

and these three terms pass to the limit under the strict convergence  $x_\beta \rightarrow x$  that we prove below. We therefore focus on the other terms in  $\mathcal{J}_\beta$  and  $\mathcal{J}_{RI}$ . By (36), it is sufficient to prove that

$$\begin{aligned} & \limsup_{\beta \rightarrow \infty} \int_0^T \left[ \psi_\beta(\dot{x}_\beta(t)) + \psi_\beta^*(\nabla E(x_\beta(t), t)) \right] dt \\ & \leq \int_0^S \mathbb{L}(x(s), t(s), \dot{x}(s), \dot{t}(s)) ds. \end{aligned} \tag{38}$$

From condition (11), we have that  $\nabla E(x, t)$  is uniformly Lipschitz continuous in  $t$ ; let  $L$  be the Lipschitz constant. In order to define a time rescaling, we introduce an auxiliary function. We fix

$$2M := L \int_0^S |\dot{x}(s)| ds = L \text{Var}(x, [0, T]) = L \text{Var}(x, [0, S]),$$

and use Hypothesis C to obtain sequences  $\delta_\beta, K_\beta \rightarrow 0$  for this value of  $M$ . We now define

$$\begin{aligned} p_\beta(v, w) & := \inf_{\varepsilon > 0} \left\{ \varepsilon \psi_\beta\left(\frac{v}{\varepsilon}\right) + \varepsilon \psi_\beta^*(|w| \vee (A + \delta_\beta)) \right\} \\ & = |v| (|w| \vee (A + \delta_\beta)), \end{aligned} \tag{39}$$

and the infimum is achieved by

$$\varepsilon_\beta(v, w) := \frac{v}{\partial \psi_\beta^*(|w| \vee (A + \delta_\beta))}. \tag{40}$$

The function  $\varepsilon_\beta$  can be interpreted as an optimal time rescaling of a given speed  $v$  and a given force  $|w| \vee (A + \delta_\beta)$ .

*Definition of the new time  $t_\beta$  and the recovery sequence  $x_\beta$ .* For the sake of simplicity, in the following we construct a recovery sequence only for a curve  $x$  with jumps at 0 and  $T$ . Later in the proof we show that, in a similar way, a recovery sequence can be constructed for a curve  $x$  with countable jumps with transparent changes in the proof.

We construct the recovery sequence by first perturbing the time variable  $t$ . We define  $t_\beta : [0, S] \rightarrow [0, T_\beta]$  as the solution of the differential equation

$$\dot{t}_\beta(s) = \dot{t}(s) \vee \varepsilon_\beta(\dot{x}(s), \nabla E(x(s), t(s))), \quad t_\beta(0) = 0. \tag{41}$$

We can assume that  $|\dot{x}(s)| \neq 0$  for  $s \in [0, s(0)]$  and  $s \in [s(T^-), S]$  to guarantee the positivity of  $\varepsilon_\beta$ . Then, for  $s \in [s(0), s(T^-)]$ , we have

$$\dot{t}(s) = \frac{1}{\dot{s}(t)} \Big|_{t=t(s)} > 0,$$

so that  $\dot{t}_\beta(s) > 0$  for all  $s \in [0, S]$ . The range of  $t_\beta$  is  $[0, T_\beta]$ , with  $T_\beta \geq T$ ; since the recovery sequence  $x_\beta$  is to be defined on the interval  $[0, T]$ , we rescale  $t_\beta$  by

$$\lambda_\beta := \frac{T_\beta}{T} \geq 1,$$

and define our recovery sequence as follows:

$$x_\beta(t) := x\left(t_\beta^{-1}(t\lambda_\beta)\right), \quad \text{so that} \quad \dot{x}_\beta(t) = \frac{\dot{x}}{\dot{t}_\beta}\left(t_\beta^{-1}(t\lambda_\beta)\right)\lambda_\beta.$$

We now have that

$$\begin{aligned} & \int_0^T \left[ \psi_\beta (\dot{x}_\beta(t)) + \psi_\beta^* (\nabla E(x_\beta(t), t)) \right] dt \\ &= \int_0^T \left[ \psi_\beta \left( \frac{\dot{x}}{\dot{t}_\beta} \left( t_\beta^{-1}(t\lambda_\beta) \right) \lambda_\beta \right) + \psi_\beta^* \left( \nabla E \left( \mathbf{x}(t_\beta^{-1}(t\lambda_\beta)), t \right) \right) \right] dt \\ &= \int_0^S \left[ \psi_\beta \left( \frac{\dot{x}}{\dot{t}_\beta}(s)\lambda_\beta \right) + \psi_\beta^* \left( \nabla E(\mathbf{x}(s), t_\beta(s)\lambda_\beta^{-1}) \right) \right] \frac{\dot{t}_\beta(s)}{\lambda_\beta} ds. \end{aligned}$$

*Estimates.* The inequality (38) now follows from the following three estimates:

**Lemma 1** Write  $\varepsilon_\beta(s) := \varepsilon_\beta(\dot{\mathbf{x}}(s), \nabla E(\mathbf{x}(s), \mathbf{t}(s)))$ . Then, there exists  $C_\beta \rightarrow 0$  for  $\beta \rightarrow \infty$  such that

$$\begin{aligned} & \int_0^S \left[ \psi_\beta^* \left( |\nabla E(\mathbf{x}(s), \mathbf{t}_\beta(s)\lambda_\beta^{-1})| \right) \frac{\dot{t}_\beta(s)}{\lambda_\beta} - \psi_\beta^* \left( |\nabla E(\mathbf{x}(s), \mathbf{t}(s))| \right) \varepsilon_\beta(s) \right] ds \\ & \leq C_\beta S; \end{aligned} \tag{42}$$

$$\int_0^S \left[ \psi_\beta \left( \frac{\dot{x}}{\dot{t}_\beta}(s)\lambda_\beta \right) \frac{\dot{t}_\beta(s)}{\lambda_\beta} - \psi_\beta \left( \frac{\dot{x}}{\varepsilon_\beta}(s) \right) \varepsilon_\beta(s) \right] ds \leq C_\beta S; \tag{43}$$

$$\begin{aligned} & \int_0^S \left[ \psi_\beta \left( \frac{\dot{x}}{\varepsilon_\beta}(s) \right) + \psi_\beta^* \left( |\nabla E(\mathbf{x}(s), \mathbf{t}(s))| \right) \right] \varepsilon_\beta(s) ds \\ & \leq \int_0^S |\dot{\mathbf{x}}(s)| (A \vee |\nabla E(\mathbf{x}(s), \mathbf{t}(s))|) ds + C_\beta S. \end{aligned} \tag{44}$$

We prove this lemma below.

*Convergence and conclusion.* Strict convergence of  $x_\beta \rightarrow x$  follows if we prove the pointwise convergence  $x_\beta(t) \rightarrow x(t)$  for all  $t \in [0, T]$  and the convergence of the variation.

Recall the definition of  $\dot{t}_\beta(s) = \dot{t}(s) \vee \varepsilon_\beta(\dot{\mathbf{x}}(s), \nabla E(\mathbf{x}(s), \mathbf{t}(s)))$ , then

$$\begin{aligned} \limsup_{\beta \rightarrow \infty} \sup_s \varepsilon_\beta(s) &= \lim_{\beta \rightarrow \infty} \sup_s \frac{|\dot{\mathbf{x}}(s)|}{\partial \psi_\beta^*(|\nabla E(\mathbf{x}(s), \mathbf{t})| \vee (A + \delta_\beta))} \\ &\leq \lim_{\beta \rightarrow \infty} \frac{\sup_s |\dot{\mathbf{x}}(s)|}{\partial \psi_\beta^*(A + \delta_\beta)} = 0, \end{aligned}$$

implies  $\dot{t}_\beta(s) \rightarrow \dot{t}(s)$ , and so it also holds

$$t_\beta(s) \rightarrow t(s) \quad \implies \quad t_\beta^{-1}(t\lambda_\beta) \rightarrow \mathbf{s}(t) \quad \forall t \in (0, T).$$

Moreover,  $\dot{t}_\beta(s) > 0$  implies that  $t_\beta^{-1}(0) = 0$  and  $t_\beta^{-1}(T\lambda_\beta) = S$ , and so we have that

$$x_\beta(t) = \mathbf{x} \left( t_\beta^{-1}(t\lambda_\beta) \right) \rightarrow \mathbf{x}(\mathbf{s}(t)) \stackrel{(35)}{=} x(t) \quad \forall t \in [0, T].$$

The convergence of the variation is automatic, since by definition of  $x_\beta$

$$\int_0^T |\dot{x}_\beta(t)| dt = \int_0^S |\dot{\mathbf{x}}(s)| ds = \text{Var}(\mathbf{x}, [0, S]) \stackrel{(37)}{=} \text{Var}(x, [0, T]).$$

*Recovery sequence for a general curve  $x$ .* Now we show how to construct a recovery sequence for a curve with countable jumps. Given the jump set  $J_x$ , fix  $\varepsilon > 0$ , consider a countable set  $\{t^i\} \subseteq J_x \cup \{0, T\}$  (with  $t^i < t^{i+1}$ ) such that

$$\text{Imp}_E(x, J_x \setminus \{t^i\}) < \varepsilon, \tag{45}$$

and such that the interval  $[0, T]$  can be written as the union of disjoint subintervals

$$[0, T] = \bigcup_i \Sigma^i \quad \text{where } \Sigma^i = [t^i, t^{i+1}].$$

It is important to underline that we would like to take the set  $\{t^i\} \supset J_x$ , but then the decomposition of the interval  $[0, T]$  as a finite union of intervals would not be true in general. Due to this complication, we must take a subset, and in the end, we show how to conclude with a diagonal argument.

Then, let  $t_\beta^i = \mathbf{t}_\beta(\mathbf{s}(t^i))$ , with  $\mathbf{t}_\beta$  from (41) and  $\mathbf{s}$  from (34), we define

$$\lambda_\beta^i = \frac{t_\beta^{i+1} - t_\beta^i}{t^{i+1} - t^i},$$

and the recovery sequence is

$$x_\beta(t) := \mathbf{x} \left( \mathbf{t}_\beta^{-1} \left( \lambda_\beta^i (t - t^i) + t_\beta^i \right) \right) \quad \text{for } t \in \Sigma^i, \tag{46}$$

so that

$$\dot{x}_\beta(t) = \frac{\dot{\mathbf{x}}}{\dot{\mathbf{t}}_\beta} \left( \mathbf{t}_\beta^{-1} \left( \lambda_\beta^i (t - t^i) + t_\beta^i \right) \right) \lambda_\beta^i \quad \text{for } t \in (\Sigma^i)^\circ.$$

We have now that

$$\begin{aligned} \int_0^T \left[ \psi_\beta(\dot{x}_\beta(t)) + \psi_\beta^*(\nabla E(x_\beta(t), t)) \right] dt &= \sum_i \int_{\mathbf{s}(t^i)}^{\mathbf{s}(t^{i+1})} \left[ \psi_\beta \left( \frac{\dot{\mathbf{x}}}{\dot{\mathbf{t}}_\beta}(s) \lambda_\beta^i \right) \right. \\ &\quad \left. + \psi_\beta^* \left( \nabla E(\mathbf{x}(s), (\lambda_\beta^i)^{-1}(\mathbf{t}_\beta(s) - t_\beta^i) + t^i) \right) \right] \frac{\dot{\mathbf{t}}_\beta(s)}{\lambda_\beta^i} ds. \end{aligned}$$

Applying Lemma 1 in every subinterval  $[\mathbf{s}(t^i), \mathbf{s}(t^{i+1})]$ , we obtain the same bounds (42–44) with  $C_\beta S$  substituted by  $C_\beta |\mathbf{s}(t^{i+1}) - \mathbf{s}(t^i)|$ , and it is important to underline that  $C_\beta$  is independent of  $\varepsilon$ . Then, inequality (38) follows because

$$\sum_i C_\beta |\mathbf{s}(t^{i+1}) - \mathbf{s}(t^i)| = C_\beta S.$$

The convergence of the variation follows again by definition of  $x_\beta$ .

The pointwise convergence of  $x_\beta(t) \rightarrow x(t)$  for  $t \in [0, T] \setminus J_x$  is again trivial. The following calculations show that, by construction, the convergence holds also in the points  $\{t^i\} \subseteq J_x$

$$x_\beta(t^i) \stackrel{(46)}{=} \mathbf{x} \left( \mathbf{t}_\beta^{-1} \left( t_\beta^i \right) \right) = \mathbf{x} \left( \mathbf{t}_\beta^{-1} \left( \mathbf{t}_\beta(\mathbf{s}(t^i)) \right) \right) = \mathbf{x} \left( \mathbf{s}(t^i) \right) \stackrel{(35)}{=} x(t^i).$$

while from (45) and the convergence of the variation we have that

$$\lim_{\beta \rightarrow \infty} |x_\beta(t) - x(t)| < \varepsilon, \quad \forall t \in J_x \setminus \{t^i\}.$$

In fact the recovery sequence  $x_\beta$  has a hidden dependence on  $\varepsilon$ , then taking  $\varepsilon = \beta^{-1}$  we define a new recovery sequence, that we keep labelling  $x_\beta$ , and sending  $\beta \rightarrow \infty$  ( $\varepsilon$  to zero) we conclude.  $\square$

*Proof (Proof of Lemma 1)* First note that for any  $s' \in [0, S]$ ,

$$\begin{aligned} 0 \leq \mathbf{t}_\beta(s') - \mathbf{t}(s') &= \int_0^{s'} (\dot{\mathbf{t}}_\beta(s) - \dot{\mathbf{t}}(s)) ds \leq \int_0^S \varepsilon_\beta(s) ds \leq \int_0^S \frac{|\dot{\mathbf{x}}(s)|}{\partial \psi_\beta^*(A + \delta_\beta)} ds \\ &= \text{Var}(\mathbf{x}, [0, S]) K_\beta \leq \frac{2MK_\beta}{L}. \end{aligned}$$



Consequently,

$$0 \leq T_\beta - T = T(\lambda_\beta - 1) \leq \frac{2MK_\beta}{L} \rightarrow 0 \quad \text{as } \beta \rightarrow \infty. \quad (47)$$

Using the Lipschitz continuity of  $\nabla E$  in time, we also have

$$\begin{aligned} & |\nabla E(\mathbf{x}(s), \mathbf{t}_\beta(s)\lambda_\beta^{-1})| - |\nabla E(\mathbf{x}(s), \mathbf{t}(s))| \leq L|\mathbf{t}_\beta(s)\lambda_\beta^{-1} - \mathbf{t}(s)| \\ & \leq L|\mathbf{t}_\beta(s)\lambda_\beta^{-1} - \mathbf{t}_\beta(s)| + L|\mathbf{t}_\beta(s) - \mathbf{t}(s)| \\ & \leq MK_\beta. \end{aligned} \quad (48)$$

We now prove (42). Using the convexity of  $\psi_\beta^*$

$$\begin{aligned} \psi_\beta^* \left( |\nabla E(\mathbf{x}(s), \mathbf{t}_\beta(s)\lambda_\beta^{-1})| \right) & \stackrel{(48)}{\leq} \psi_\beta^* (|\nabla E(\mathbf{x}(s), \mathbf{t}(s))| + MK_\beta) \\ & \leq \psi_\beta^* (|\nabla E(\mathbf{x}(s), \mathbf{t}(s))|) + MK_\beta \partial \psi_\beta^* (|\nabla E(\mathbf{x}(s), \mathbf{t}(s))| + MK_\beta). \end{aligned} \quad (49)$$

Setting  $\Sigma_\beta := \{s \in [0, S] : \dot{\mathbf{t}}_\beta(s) = \dot{\mathbf{t}}(s)\}$ , we split the domain into  $\Sigma_\beta$  and  $\Sigma_\beta^c$ . On  $\Sigma_\beta$ , since  $\dot{\mathbf{t}}(s) > 0$ , the finiteness of  $\mathbb{L}$  implies that  $|\nabla E| \leq A$ ; on  $\Sigma_\beta^c$ ,  $\dot{\mathbf{t}}_\beta(s) = \varepsilon_\beta(s)$ . Therefore,

$$\begin{aligned} & \int_0^S \psi_\beta^* (|\nabla E(\mathbf{x}(s), \mathbf{t}(s))|) [\dot{\mathbf{t}}_\beta(s) - \varepsilon_\beta(s)] \, ds \\ & = \int_{\Sigma_\beta} \psi_\beta^* (|\nabla E(\mathbf{x}(s), \mathbf{t}(s))|) [\dot{\mathbf{t}}_\beta(s) - \varepsilon_\beta(s)] \, ds \\ & \leq \psi_\beta^*(A) \int_{\Sigma_\beta} \dot{\mathbf{t}}_\beta(s) \, ds \leq C_\beta S, \end{aligned} \quad (50)$$

with  $C_\beta := \psi_\beta^*(A)$ .

We also integrate the second term in (49) first on  $\Sigma_\beta$  and then on  $\Sigma_\beta^c$ . On  $\Sigma_\beta$ , again since  $|\nabla E| \leq A$

$$\begin{aligned} & \int_{\Sigma_\beta} MK_\beta \partial \psi_\beta^* (|\nabla E(\mathbf{x}(s), \mathbf{t}(s))| + MK_\beta) \dot{\mathbf{t}}_\beta(s) \, ds \\ & \leq MK_\beta \psi_\beta^*(A + MK_\beta) \int_{\Sigma_\beta} \dot{\mathbf{t}}_\beta(s) \, ds \leq C_\beta S, \end{aligned}$$

with  $C_\beta := MK_\beta \psi_\beta^*(A + MK_\beta)$ .

On the other hand, on  $\Sigma_\beta^c$ , using

$$\dot{\mathbf{t}}_\beta(s) = \varepsilon_\beta(s) = |\dot{\mathbf{x}}(s)| / \partial \psi_\beta^* (|\nabla E(\mathbf{x}(s), \mathbf{t}(s))| \vee (A + \delta_\beta)),$$

we get

$$\begin{aligned} & \int_{\Sigma_\beta^c} MK_\beta \partial \psi_\beta^* (|\nabla E(\mathbf{x}(s), \mathbf{t}(s))| + MK_\beta) \dot{\mathbf{t}}_\beta(s) \, ds \\ & \leq M \int_{\Sigma_\beta^c} K_\beta \frac{\partial \psi_\beta^* (|\nabla E(\mathbf{x}(s), \mathbf{t}(s))| + MK_\beta)}{\partial \psi_\beta^* (|\nabla E(\mathbf{x}(s), \mathbf{t}(s))| \vee (A + \delta_\beta))} |\dot{\mathbf{x}}(s)| \, ds \leq C_\beta S, \end{aligned} \quad (51)$$

where the last inequality holds because  $\text{Var}(\mathbf{x}, [0, S]) \leq S$ , with

$$C_\beta := \sup_s MK_\beta \frac{\partial \psi_\beta^* (|\nabla E(\mathbf{x}(s), \mathbf{t}(s))| + MK_\beta)}{\partial \psi_\beta^* (|\nabla E(\mathbf{x}(s), \mathbf{t}(s))| \vee (A + \delta_\beta))}.$$

Together, (50–51) prove (42) with vanishing  $C_\beta$  thanks to Hypotheses B–C.

To prove (43), we use the fact that  $\dot{\mathbf{t}}_\beta(s) \geq \varepsilon_\beta(s)$  for all  $s$ , and that the mapping  $\tau \mapsto \tau \psi_\beta(v/\tau)$  is non-increasing. Therefore,

$$\begin{aligned} \int_0^S \psi_\beta \left( \frac{\dot{\mathbf{x}}}{\dot{\mathbf{t}}_\beta}(s) \lambda_\beta \right) \frac{\dot{\mathbf{t}}_\beta(s)}{\lambda_\beta} ds &\leq \int_0^S \psi_\beta \left( \frac{\dot{\mathbf{x}}}{\varepsilon_\beta}(s) \lambda_\beta \right) \frac{\varepsilon_\beta(s)}{\lambda_\beta} ds \\ &\stackrel{(*)}{\leq} \int_0^S \left[ \psi_\beta \left( \frac{\dot{\mathbf{x}}}{\varepsilon_\beta}(s) \right) + C \partial \psi_\beta^* (|\nabla E(\mathbf{x}(s), \mathbf{t}(s))| \vee (A + \delta_\beta)) (\lambda_\beta - 1) \right] \varepsilon_\beta(s) ds. \end{aligned}$$

where  $C$  is a constant and we prove the inequality marked  $(*)$  below. Continuing with the argument, we again apply the definition (40) of  $\varepsilon_\beta$  to find

$$(\lambda_\beta - 1) \int_0^S \partial \psi_\beta^* (|\nabla E(\mathbf{x}(s), \mathbf{t}(s))| \vee (A + \delta_\beta)) \varepsilon_\beta(s) ds \leq (\lambda_\beta - 1) \int_0^S |\dot{\mathbf{x}}(s)| ds \leq C_\beta S,$$

where  $C_\beta := (\lambda_\beta - 1)$  converges to zero by (47).

We next prove (44), with  $C_\beta := \delta_\beta$ . We calculate

$$\begin{aligned} &\int_0^S \left[ \psi_\beta \left( \frac{\dot{\mathbf{x}}}{\varepsilon_\beta}(s) \right) + \psi_\beta^* (|\nabla E(\mathbf{x}(s), \mathbf{t}(s))|) \right] \varepsilon_\beta(s) ds \\ &\leq \int_0^S \left[ \psi_\beta \left( \frac{\dot{\mathbf{x}}}{\varepsilon_\beta}(s) \right) + \psi_\beta^* (|\nabla E(\mathbf{x}(s), \mathbf{t}(s))| \vee (A + \delta_\beta)) \right] \varepsilon_\beta(s) ds \\ &= \int_0^S |\dot{\mathbf{x}}(s)| (|\nabla E(\mathbf{x}(s), \mathbf{t}(s))| \vee (A + \delta_\beta)) ds \quad \text{by (39)} \\ &\leq \int_0^S |\dot{\mathbf{x}}(s)| (|\nabla E(\mathbf{x}(s), \mathbf{t}(s))| \vee A) ds + C_\beta S. \end{aligned}$$

We finally prove the inequality  $(*)$  above, as the following separate result: for each  $R > 0$ , there exists  $C > 0$  such that

$$\forall \alpha \geq 1, \forall |z| \leq R, \forall \beta : \quad \psi_\beta(\alpha \partial \psi_\beta^*(z)) \leq \psi_\beta(\partial \psi_\beta^*(z)) + C(\alpha - 1) \partial \psi_\beta^*(z).$$

To show this, note that by Hypothesis D for each  $\alpha \geq 1$  and for each  $|z| \leq R$  there exists  $\eta_\beta(z, \alpha)$  such that

$$\partial \psi_\beta^*(z + \eta_\beta(z, \alpha)) = \alpha \partial \psi_\beta^*(z),$$

and  $\eta_\beta$  is bounded uniformly in  $\alpha, \beta$ , and  $|z| \leq R$ . The following three statements follow from convexity and convex duality:

$$\begin{aligned} \psi_\beta(\alpha \partial \psi_\beta^*(z)) + \psi_\beta^*(z + \eta_\beta) - (z + \eta_\beta) \alpha \partial \psi_\beta^*(z) &= 0; \\ \psi_\beta(\partial \psi_\beta^*(z)) + \psi_\beta^*(z) - z \partial \psi_\beta^*(z) &= 0; \\ \psi_\beta^*(z + \eta_\beta) - \psi_\beta^*(z) - \eta_\beta \partial \psi_\beta^*(z) &\geq 0. \end{aligned}$$

Upon subtracting the second and third line from the first, we find

$$\psi_\beta(\alpha \partial \psi_\beta^*(z)) \leq \psi_\beta(\partial \psi_\beta^*(z)) + (z + \eta_\beta)(\alpha - 1) \partial \psi_\beta^*(z),$$

which implies the result. □

### 5 Discussion

In the introduction, we posed the question whether we could understand the distinction and the relationship between gradient-flow and rate-independent systems from the point of view of stochastic processes. The simple model in one dimension of this paper gives a very clear answer that we summarize in our words as follows:

- The continuum limit is a generalized gradient flow, with non-quadratic, non-1-homogeneous dissipation, and the large-deviations rate functional ‘is’ the corresponding generalized gradient-flow structure, in the sense of Sect. 1.4;
- Taking further limits recovers both cases: quadratic gradient-flow and rate-independent;

- At least some of the limits are robust against exchanging the order of the limits, and we conjecture that this robustness goes much further.

Therefore, the quadratic and rate-independent cases appear as natural completion in the scale of systems characterized by  $\alpha$  and  $\beta$ .

In addition, the details of the proofs show how the formulation in terms of  $\mathcal{J}$  of (a) large deviations, (b) generalized gradient flows including rate-independent systems, and (c) convergence results for these systems gives a unified view on the field and a coherent set of tools for the analysis and manipulation of the systems.

Related issues have been investigated in the case of stochastic differential equations. The two limiting processes,  $n \rightarrow \infty$  and  $\beta \rightarrow \{0, \infty\}$  can be interpreted as differently scaled combinations of two limiting processes: (a) the small-noise limit, (b) the limit of vanishing microstructure. In the case of SDEs [6, 18, 22], three regimes have been identified, corresponding to ‘microstructure smaller than noise’, ‘noise smaller than microstructure’, and the critical case. In the first of these, ‘microstructure smaller than noise’, a behaviour arises that resembles the quadratic limit of this paper, in which the microstructure is effectively swamped by the noise. The critical case resembles our original large-deviations result (Theorem 1) in that both give non-quadratic, non-one-homogeneous rate functionals. Finally, when the noise is asymptotically smaller than the microstructure, a limit similar to the rate-independent limit is obtained in [18, 22], but because the authors consider time-invariant energies and a different scaling, the behaviour of the limiting system is rather different.

The one-dimensionality of the current set-up may appear to be a significant restriction, but we believe (and in some cases we know) that the structure can be generalized to a wide class of other systems. For instance,

- The initial large-deviations result (Theorem 1) also holds in higher dimensions; other proofs of this and similar results are given in [12, 42].
- The joint large-deviations-quadratic limit (Theorem 3) can be generalized to higher dimensions with only notational changes in the proof.
- Of the proof of the convergence to a rate-independent system (Theorem 5), one part (the liminf-inequality) has been done in the generality of a metric space, with a specific functional form of the dissipation potential, in [33]. The other part, the construction of a recovery sequence, is subject of current work; here the characterization of the limiting jump term depends on the particular form of the approximating  $\psi_\beta - \psi_\beta^*$ , in a way that is not yet clear.

More generally, the results of [30] show that the connection between large-deviations principles and generalized gradient flows is robust and arises for all reversible stochastic processes and quite a few more (such as the GENERIC system in [17]).

In Fig. 3, the question mark represents an open problem: the combined large-deviation-rate-independent limit. We conjecture that, as in the combined large-deviation-quadratic limit (Theorem 3), a large-deviations principle holds in this limit, with rate functional  $\mathcal{J}_{RI}$ . Unfortunately, the framework provided by [20] does not seem to apply as-is, and the form of this functional will require a radical change in the strategy of the proof.

**Acknowledgments** We want to sincerely thank Jin Feng for the valuable comments and suggestions. During all the preparation of this work, Giuseppe Savaré provided many useful suggestions and critical remarks, and we are very grateful for his contribution. G.A.B. and M.A.P. kindly acknowledge support from the Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO) VICI Grant 639.033.008.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

1. Abeyaratne, R., Chu, C., James, R.D.: Kinetics of materials with wiggly energies: theory and application to the evolution of twinning microstructures in a Cu–Al–Ni shape memory alloy. *Philos. Mag. A* **73**(2), 457–497 (1996)
2. Adams, S., Dirr, N., Peletier, M.A., Zimmer, J.: From a large-deviations principle to the Wasserstein gradient flow: a new micro-macro passage. *Commun. Math. Phys.* **307**, 791–815 (2011)
3. Adams, S., Dirr, N., Peletier, M.A., Zimmer, J.: Large deviations and gradient flows. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **371**(2005), 20120341 (2013)
4. Ambrosio, L., Fusco, N., Pallara, D.: *Functions of Bounded Variation and Free Discontinuity Problems*, vol. 254. Clarendon Press Oxford, Oxford (2000)

5. Ambrosio, L., Gigli, N., Savaré, G.: *Gradient Flows in Metric Spaces and in the Space of Probability Measures*, 2nd ed. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel (2008)
6. Baldi, P.: Large deviations for diffusion processes with homogenization and applications. *Ann. Probab.* **19**(2), 509–524 (1991)
7. Basinski, Z.S.: Thermally activated glide in face-centred cubic metals and its application to the theory of strain hardening. *Philos. Mag.* **4**(40), 393–432 (1959). doi:[10.1080/14786435908233412](https://doi.org/10.1080/14786435908233412)
8. Becker, R.: Über die Plasticität amorpher und kristalliner fester Körper. *Phys. Z.* **26**, 919–925 (1925)
9. Berglund, N.: Kramers’ Law: Validity, Derivations and Generalisations. Arxiv preprint [arXiv:1106.5799](https://arxiv.org/abs/1106.5799) (2011)
10. Brezis, H.: *Functional Analysis, Sobolev Spaces and Partial Differential Equations*. Springer, New York (2011)
11. Cagnetti, F.: A vanishing viscosity approach to fracture growth in a cohesive zone model with prescribed crack path. *Math. Models Methods Appl. Sci.* **18**(07), 1027–1071 (2008)
12. Chen, X.: Global asymptotic limit of solutions of the Cahn–Hilliard equation. *J. Differ. Geom.* **44**, 262–311 (1996)
13. Dal Maso, G., DeSimone, A., Mora, M.G.: Quasistatic evolution problems for linearly elastic–perfectly plastic materials. *Arch. Ration. Mech. Anal.* **180**(2), 237–291 (2006)
14. Dal Maso, G., DeSimone, A., Mora, M.G., Morini, M.: A vanishing viscosity approach to quasistatic evolution in plasticity with softening. *Arch. Ration. Mech. Anal.* **189**(3), 469–544 (2008)
15. Dirr, N., Laschos, V., Zimmer, J.: Upscaling from particle models to entropic gradient flows. *J. Math. Phys.* **53**(6), 063704 (2012)
16. Duong, M.H., Laschos, V., Renger, D.R.M.: Wasserstein gradient flows from large deviations of many-particle limits. *ESAIM: Control Optim. Calc. Var.* **E-first** (2013)
17. Duong, M.H., Peletier, M.A., Zimmer, J.: GENERIC formalism of a Vlasov–Fokker–Planck equation and connection to large-deviation principles. *Nonlinearity* **26**, 2951–2971 (2013)
18. Dupuis, P., Spiliopoulos, K.: Large deviations for multiscale diffusion via weak convergence methods. *Stoch. Process. Appl.* **122**(4), 1947–1987 (2012)
19. Ethier, S.N., Kurtz, T.G.: *Markov Processes: Characterization and Convergence*, vol. 282. Wiley, Hoboken (2009)
20. Feng, J., Kurtz, T.G.: *Large Deviations for Stochastic Processes*, vol. 131. Citeseer (2006)
21. Fiaschi, A.: A vanishing viscosity approach to a quasistatic evolution problem with nonconvex energy. *Annales de l’Institut Henri Poincaré (C) Non Linear Anal.* **26**(4), 1055–1080 (2009)
22. Freidlin, M.I., Sowers, R.B.: A comparison of homogenization and large deviations, with applications to wavefront propagation. *Stoch. Process. Appl.* **82**(1), 23–52 (1999)
23. Freidlin, M.I., Wentzell, A.D.: *Random Perturbations of Dynamical Systems*, vol. 260. Springer, New York (2012)
24. Kramers, H.A.: Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica* **7**(4), 284–304 (1940)
25. Krausz, A.S., Eyring, H.: *Deformation Kinetics*. Wiley, New York (1975)
26. Mainik, A., Mielke, A.: Existence results for energetic models for rate-independent systems. *Cal. Var. Partial Differ. Equ.* **22**(1), 73–99 (2005)
27. Mielke, A.: *Handbook of Differential Equations: Evolutionary Differential Equations*, chap. Evolution in Rate-Independent Systems, pp. 461–559. North-Holland, Amsterdam (2005)
28. Mielke, A.: Emergence of rate-independent dissipation from viscous systems with wiggly energies. *Contin. Mech. Thermodyn.* **24**(4–6), 591–606 (2012)
29. Mielke, A.: On Evolutionary Gamma-Convergence for Gradient Systems. Tech. Rep. 1915, WIAS, Berlin (2014)
30. Mielke, A., Peletier, M.A., Renger, D.R.M.: On the Relation Between Gradient Flows and the Large-Deviation Principle, with Applications to Markov Chains and Diffusion. arXiv preprint [arXiv:1312.7591](https://arxiv.org/abs/1312.7591) (2013)
31. Mielke, A., Rossi, R., Savaré, G.: Modeling solutions with jumps for rate-independent systems on metric spaces. *Discrete Contin. Dyn. Syst. A* **25**(2), 585–615 (2009)
32. Mielke, A., Rossi, R., Savaré, G.: BV solutions and viscosity approximations of rate-independent systems. *ESAIM Control Optim. Calc. Var.* **18**(01), 36–80 (2012)
33. Mielke, A., Rossi, R., Savaré, G.: Variational convergence of gradient flows and rate-independent evolutions in metric spaces. *Milan J. Math.* **80**(2), 381–410 (2012)
34. Mielke, A., Rossi, R., Savaré, G.: Balanced-Viscosity (bv) Solutions to Infinite Dimensional Rate-Independent Systems. arXiv preprint [arXiv:1309.6291](https://arxiv.org/abs/1309.6291) (2013)
35. Mielke, A., Theil, F.: A mathematical model for rate-independent phase transformations with hysteresis. In: *Proceedings of the Workshop on “Models of Continuum Mechanics in Analysis and Engineering*, pp. 117–129 (1999)
36. Mielke, A., Theil, F.: On rate-independent hysteresis models. *Nonlinear Differ. Equ. Appl.* **11**(2), 151–189 (2004)
37. Mielke, A., Theil, F., Levitas, V.I.: A variational formulation of rate-independent phase transformations using an extremum principle. *Arch. Ration. Mech. Anal.* **162**(2), 137–177 (2002)
38. Mielke, A., Truskinovsky, L.: From discrete visco-elasticity to continuum rate-independent plasticity: rigorous results. *Arch. Ration. Mech. Anal.* **203**(2), 577–619 (2012)
39. Orowan, E.: Problems of plastic gliding. *Proc. Phys. Soc.* **52**, 8–22 (1940)
40. Puglisi, G., Truskinovsky, L.: Thermodynamics of rate-independent plasticity. *J. Mech. Phys. Solids* **53**(3), 655–679 (2005). doi:[10.1016/j.jmps.2004.08.004](https://doi.org/10.1016/j.jmps.2004.08.004). <http://www.sciencedirect.com/science/article/pii/S0022509604001425>
41. Renger, D.R.M.: *Microscopic Interpretation of Wasserstein Gradient Flows*. Ph.D. thesis, Technische Universiteit Eindhoven (2013). <http://alexandria.tue.nl/extra2/749143.pdf>
42. Shwartz, A., Weiss, A.: *Large Deviations for Performance Analysis: Queues, Communications, and Computing*. Chapman & Hall/CRC, London (1995)
43. Sullivan, T.J.: *Analysis of Gradient Descents in Random Energies and Heat Baths*. Ph.D. thesis, University of Warwick (2009)
44. Sullivan, T.J., Koslowski, M., Theil, F., Ortiz, M.: On the behavior of dissipative systems in contact with a heat bath: application to andrade creep. *J. Mech. Phys. Solids* **57**(7), 1058–1077 (2009)

- 
45. Trotter, H.F.: Approximation of semi-groups of operators. *Pac. J. Math.* **8**(4), 887–919 (1958)
  46. Varadhan, S.R.S.: Asymptotic probabilities and differential equations. *Commun. Pure Appl. Math.* **19**(3), 261–286 (1966)
  47. Wentzell, A.D.: Rough limit theorems on large deviations for Markov stochastic processes I. *Theory Probab. Appl.* **21**(2), 227–242 (1977)
  48. Wentzell, A.D.: *Limit Theorems on Large Deviations for Markov Stochastic Processes*, vol. 38. Springer, New York (1990)