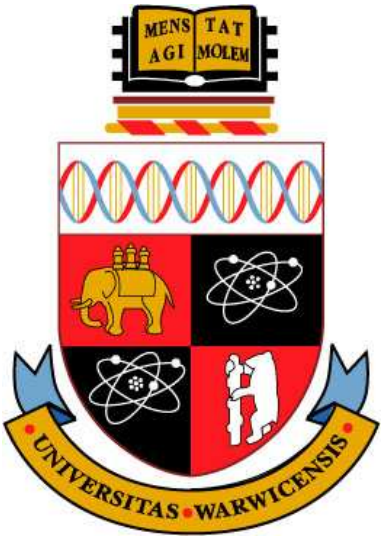


Analysis of Gradient Descents in Random Energies and Heat Baths

Timothy John Sullivan

A thesis submitted in fulfilment of the requirements
for the degree of Doctor of Philosophy

Mathematics Institute, University of Warwick

July 2009

THE UNIVERSITY OF
WARWICK

Contents

List of Figures

Acknowledgements

Thanks are due to the many friends, family, and fellow mathematicians who have helped (directly, indirectly, or by providing necessary distraction) to bring this thesis to completion. With many apologies to those whom I have missed out, thanks go to Аня, Ben G., Ben H., Dad, Dad, Elvina, Ferry, Ian, Iona, James, Jenny, John, Κώστας, Mark, McC., Mum, Настя, Nicholas, Oli, Patrick, Sensei, Sze Vone and Tom. In particular, for all her love and support, *большое спасибо Миленушке!* Since my teaching practice has also been invaluable in sharpening my understanding of many topics, I thank all my students for having put up with me for so long!

Several members of staff at Warwick have taken interest in my research and have offered kind and constructive suggestions along the way. In this connection, I would like to thank Dr Volker Betz, Prof. Martin Hairer and Dr Jochen Voß. Dr Mario Micallef and Dr Luca Sbano have also been great sources of inspiration and mathematical passion since my undergraduate days. Carole Fisher and Carol Wright also deserve a mention for their efforts in support of all the graduate students in the Warwick Mathematics Institute.

I would particularly like to thank Prof. Michael Ortiz for inviting me to the California Institute of Technology between March and May 2007 and for his guidance and hospitality during my visit.

Above all, thanks go to my advisor, Dr Florian Theil, for his patience and guidance. It was he who introduced me to the problems addressed in this work; with his experience and judgment, he has guided and occasionally prodded me down the path to their resolution. Finally, I would like to thank my examiners, Professors Valery Smyshlyaev and Andrew Stuart, for their insightful commentary and corrections to the manuscript.

This Ph.D. was funded by the Engineering and Physical Sciences Research Council. My visits to the California Institute of Technology and the Mathematisches Forschungsinstitut Oberwolfach between February and May 2007 were funded by the University of Warwick Research Development Fund. My visit to the Hausdorff Center for Mathematics, Bonn, in January 2007 was generously funded by the Hausdorff Center itself.

This thesis was typeset using LaTeX 2_ε and BibTeX 0.99c, with figures produced using Wolfram Mathematica 5.0 and PSTricks/LaTeXDraw 2.0.

Declaration

I declare that, to the best of my knowledge, the material contained in this thesis is original and my own work except where otherwise indicated, cited, or commonly known.

The material in this thesis is submitted to the University of Warwick for the degree of Doctor of Philosophy, and has not been submitted to any other university or for any other degree.

Dedication

Dedicated to the memory of
Ahmed Kotb
1974–2008

Blue Skies!

Abstract

This thesis concerns the mathematical analysis of random gradient descent evolutions as models for rate-independent dissipative systems under the influence of thermal effects. The basic notions of the theory of gradient descents (especially rate-independent evolutions) are reviewed in chapter 2. Chapters 3 and 4 focus on the scaling regime in which the microstructure dominates the thermal effects and comprise a rigorous justification of rate-independent processes in smooth, convex energies as scaling limits of rate-dependent gradient descents in energies that have rapidly-oscillating random microstructure: chapter 3 treats the one-dimensional case with quite a broad class of random microstructures; chapter 4 treats a case in which the microstructure is modeled by a sum of "dent functions" that are scattered in $\mathbb{R}^n$ using a suitable point process. Chapters 5 and 6 focus on the opposite scaling regime: a gradient descent system (typically a rate-independent process) is placed in contact with a heat bath. The method used to "thermalize" a gradient descent is an interior-point regularization of the Moreau–Yosida incremental problem for the original gradient descent. Chapter 5 treats the heuristics and generalities; chapter 6 treats the case of 1-homogeneous dissipation (rate independence) and shows that the heat bath destroys the rate independence in a controlled and deterministic way, and that the effective dynamics are a gradient descent in the original energetic potential but with respect to a different and non-trivial effective dissipation potential. The appendices contain some auxiliary definitions and results, most of them standard in the literature, that are used in the main text.

Chapter 1

Introduction and Outline

1.1 Introduction

This thesis concerns the mathematical analysis of gradient descent evolutions as models for rate-independent dissipative systems under the influence of thermal effects. A gradient descent is a very general kind of evolutionary process, one inspired by the general maxim that many systems in the real world evolve in such a way as to "move downhill" in some potential energy landscape. They do not usually do so in an instantaneous manner: the rate of descent is controlled by a dissipation functional or kinetic potential — gradient descents, therefore, fall under the more general heading of dissipative systems. Rate-independence is a special case in which the system has no intrinsic timescale of its own; it "reacts only as fast as its time-dependent inputs"; mathematically, this arises if the dissipation potential is homogeneous of degree one as a function of velocity. Whereas gradient descents exhibit a monotone decrease (or possibly conservation) of energy along trajectories, coupling a dissipative system to a heat bath injects additional disordered energy into the system, broadening the class of possible evolutions.

Dissipative systems (and, in particular, rate-independent processes) are mathematically interesting because they are multiscale systems, and the behaviours at the microscale and macroscale are both qualitatively and quantitatively different; that said, the macroscopic behaviour should arise as some "average" of the microscopic contributions. It is desired to make a mathematically rigorous passage from the microscale to the macroscale and to understand the scaling regimes involved. If noise (a heat bath) is present, then the relative strengths of the microstructural variations and

the heat bath may be particularly important. In either case, the tools of multiscale analysis and probability theory will be required, especially in the case of random microstructure, as is the case in this thesis. In the absence of noise, macroscopic rate-independent evolutions have a strongly geometric and convex-analytical flavour, a connection apparently first noticed by Moreau [Mor70] [Mor71] [Mor76].

Rate-independent systems occur throughout physics, engineering, materials science, finance and other areas; they play an important rôle in the mathematical modeling of physical phenomena such as plasticity, phase transformations in elastic solids, electromagnetism, dry friction on surfaces, and pinning problems in superconductivity. As well as being a spatial (microscale-to-macroscale) limit, they are also a temporal limit; to quote [MT04], "the evolution equations which govern [these] processes constitute the limit problems if the influence of inertia and relaxation times vanishes." This assumption of vanishing inertia and relaxation times is referred to as *quasistaticity*. Again, apparently "purer" mathematics provides useful tools: for example, there is a relationship between the Yosida regularization of nonlinear semigroup theory [Yos65] and linear viscoplasticity, a connection first noted by Ortiz [Ort81] [OR99].

As a toy model, consider a block resting on an inclined plane and suppose that the interface between the two is rough — covered with sandpaper, say. In the absence of any other forces, if the plane is flat, then the block will not move; even as the angle of inclination of the plane increases, the block will *stick* until a critical angle of inclination is reached, at which point the block will *slip*. From the microscopic point of view, the sticking occurs because of the many local energy minima (microstructural variations) possessed by the sandpaper interface; in the macroscopic world, the effective behaviour is that of dry friction. Stick-slip macroscopic behaviour is qualitatively different from that of a solution to an ordinary differential equation: the former is governed by a dissipation potential that is homogeneous of degree one, the latter by a dissipation potential that is homogeneous of degree two. In the absence of inertial effects, systems with 1-homogeneous dissipation also tend to exhibit rate-independent macroscopic behaviour.

To take a less toy-like example from materials science, the Griffith and Francfort–Marigo criteria [Gri20] [FM98] for the propagation of a crack through a material sample are examples of (discrete-time) gradient descents.

At each time, the crack grows if, and only if, the elastic energy released by the growth of the crack is greater than the energy dissipated by breaking interatomic bonds to produce a new crack segment. The crack path itself is determined by minimizing elastic energy at each time, subject to this dissipation criterion. The evolution of the crack is modeled as a quasistatic evolution: it is assumed that the time required for the crack to reach equilibrium between elastic energy release and energy dissipation is negligible. The assumption that the evolutions of study are quasistatic is common in the study of many elastoplasticity problems from materials science [DT02] [DDM06]. Again, there is an obvious separation of scales, and what happens at the microscale can be of vital importance at the macroscale: the breaking of enough atomic bonds and the propagation of a crack can amount to the failure of a building's structure or an aircraft's wing! Another paradigm of the kind of multiscale problem that arises in materials science is the formation and evolution of magnetic microstructure in ferromagnetic materials [DKMO00], including the Barkhausen effect [ABBM90] [Ber98, chapter 9].

The effective macroscopic behaviour of such a system (the friction coefficient of the block/sandpaper interface, the yield stress of a material sample, & c.) are often determined by physical experiment. However, it would be mathematically more satisfying to determine them from the microstructure via some kind of multiscale analysis. In the case of rate-*dependent* viscous systems, analysis of how the random microstructure determines the macroscopic behaviour can be found in the Green–Kubo relations and many further developments since their introduction in the 1950s [Gre54] [Kub57]. In the case of periodic microstructure, convergence results in the rate-independent include those of Abeyaratne, Chu & James [ACJ96] (in dimension one) and Menon [Men01] [Men02] (in dimension two).

Multiscale analysis is most easily performed when the microstructure is periodic (although undesirable "grid effects" may arise in dimension greater than one). Periodicity is an unsatisfying assumption to have to make: the aim of [Gru04] [Gru05] was to extend the scaling results to the case of random microstructure. In that respect, chapters 3 and 4 are a continuation of this programme; some of the results were announced in [ST07]. In summary, this thesis establishes rate-independent scaling limits for rate-dependent evolutions in two main classes of random energies: in dimension one, a broad class of admissible microstructures is identified; in dimension greater than

one, attention is confined to a prototypical energy landscape that has been "dented" by a Poisson point process.

The presence of noise adds further interest to the analysis and introduces new scaling regimes. Physically, to neglect thermal effects amounts to assuming that the system is evolving at absolute zero temperature. It may be reasonable to assume that the thermal effects are of much smaller magnitude than the microstructural variations — this is the scaling regime of the references cited above and chapters 3 and 4 of this thesis. Another scaling regime, in some sense the opposite one, would be to take the macroscopic behaviour as given and then seek to "thermalize" it in some way. This ris a stochastic scaling regime in which thermal effects dominate, and amounts to placing a dissipative system in contact with a heat bath. This scaling regime and theory of *thermalized gradient descents* are the topics of chapters 5 and 6 and [SKTO09], and fit within current efforts to develop a calculus of variations for evolutionary systems and systems that exhibit evolving microstructure — see, for example, [CO08] [MO08]. Chapter 6 focuses on the case in which the underlying gradient descent is a rate-independent evolution; one obvious question is whether or not the addition of a heat bath destroys the original rate-independence in a controlled way, and the answer to this turns out to be affirmative; indeed, the effective evolution is a gradient descent with a nonlinear dissipation potential that can be calculated as a nonlinear transformation of the original dissipation potential.

The analysis in this thesis is, for the most part, limited to the finite-dimensional case. That said, the methods of chapters 5 and 6 appear to be very robust and extension to infinite-dimensional evolutions is a natural direction for further research. It is likely that not only will stochastic partial differential equations fall under the umbrella of this approach, but also evolutions governed by nonlinear dissipation.

1.2 Outline of the Thesis

Chapter 2 introduces the classical notion of a *gradient descent* in a smooth energy on $\mathbb{R}^n$. One of the key properties of classical gradient descents is that they satisfy an evolution equation for the energy along the trajectory of the system, the so-called *energy balance*. The energy balance equation forms a key part of the theory for generalized gradient descents in spaces that lack a

differentiable or even linear structure; in the general case, the energy relation becomes an energy *inequality*. The general definition of a gradient descent in a metric space is presented in the style of the recent book [AGS08]. Of particular note is the Moreau–Yosida scheme for generating a discrete-time approximation of a gradient descent: essentially, it is a minimization problem that has as its Euler–Lagrange equations the original gradient descent. Such a discretization is typically used to prove existence and uniqueness of continuous-time gradient descents. Also, an interior-point regularization of the Moreau–Yosida approximation furnishes a way to "thermalize" a gradient descent: this is the topic of chapters 5 and 6.

Chapter 2 also introduces the basic elements of the theory of rate-independent processes. These are special cases in which the dissipation potential is homogeneous of degree one. Therefore, the system has no intrinsic timescale: the solution operator commutes with monotonic reparametrizations of time. The exposition in this thesis follows that of Mielke and collaborators [MTL02] [MT04] [Mie05] [Mie07].

Chapters 3 and 4 concern the rigorous justification of rate-independent processes as scaling limits of rate-dependent gradient descents with rapidly-oscillating random microstructure; they (implicitly) use a scaling regime in which microstructural variations dominate over thermal effects. Put another way, these chapters investigate how the microstructure of an energetic potential that is explored by a classical gradient descent (i.e. a 2-homogeneous dissipation potential) gives rise to an effective 1-homogeneous dissipation potential in the macroscopic limit. This effort is part of a well-established theory of multiscale analysis, averaging and homogenization that dates back to [BLP78], anticipated to some extent by [KK73] [Lar75]; more recent references include [PS08]. The multiscale models are of the form

$$\dot{z}_\varepsilon = -\frac{1}{\varepsilon}\big(\nabla E(t, z_\varepsilon) + G_\varepsilon(z_\varepsilon)\big), \tag{1.2.1}$$

and the sought-for rate-independent limit is of the form

$$\partial\Psi(\dot{z}) \ni -\mathrm{D}E(t, z). \tag{1.2.2}$$

The underlying energetic potential E is assumed to be (relatively) smooth and well understood; the microstructure is modeled by G_ε, which is assumed to be differentiable and to have bounded derivatives of order one that os-

cillate on a spatial scale of order ε. Such problems are well-established in the literature. For the dimension $n = 1$ problem, [ACJ96] established that if $G_\varepsilon(x) = \varepsilon G(x/\varepsilon)$ is periodic, then z_ε converges as $\varepsilon \to 0$ to the solution z of the rate-independent process in E and Ψ, where $\Psi \colon \mathbb{R} \to [0, +\infty)$ is a 1-homogeneous dissipation potential that is determined by the bounds on the derivative of G. For periodic perturbations in dimension $n \geq 2$, rather sensitive analysis is necessary because of the "grid effects" that plague periodic structures. [Men01] [Men02] considered the case $n = 2$ and established a decomposition of the macroscopic space $\mathbb{R}^2$ into a bounded, open sticking region; a transition region; and the unbounded remainder — this picture is complicated by the fact that the unbounded remainder, where there should be only slipping dynamics, contains a countable number of "resonance zones" whose structure is the complement of a Cantor set.

To confine the study to periodic perturbations is both mathematically and physically unsatisfying. However, averaging and homogenization techniques are difficult (although not impossible) to apply to random microstructure: the principal difficulty arises in the random cell problem. Therefore, this thesis takes a different approach to the study of random microstructure. The techniques used in dimension $n = 1$ (chapter 3) and $n \geq 2$ (chapter 4) are markedly different. For $n = 1$, it is possible to give an almost complete characterization of those perturbations G that give rise to the desired rate-independent limit as $\varepsilon \to 0$. The requisite criterion is labeled *property* (✠) and is satisfied by the periodic perturbations considered by [ACJ96] as well as almost every sample realization of a wide class of stochastic processes. In $n \geq 2$, the perturbation G_ε is modeled by a sum of dent functions centred on the points of a dilute Poisson point process. The convergence theorem for $n \geq 2$ has a much more probabilistic flavour, in contrast with the sample-wise application of property (✠) in dimension one.

Chapters 5 and 6 focus on a different scaling regime to chapters 3 and 4: a gradient descent system (typically a rate-independent process) is given, and the objective is to model what happens when the system is placed in contact with a heat bath. The method used to "thermalize" a gradient descent is an interior-point regularization of the Moreau–Yosida incremental problem for the original gradient descent. This regularization generates a Markov chain X on the state space $\mathbb{R}^n$ whose transition probabilities are given by

the Gibbs measure/Boltzmann distribution:

$$\mathbb{P}\big[X_{i+1} \in A \big| X_i = x_i\big] = \frac{\displaystyle\int_A \exp(-\mathscr{W}_{i+1}(x_i, x_{i+1})/\varepsilon_{i+1})\,\mathrm{d}x_{i+1}}{\displaystyle\int_{\mathbb{R}^n} \exp(-\mathscr{W}_{i+1}(x_i, x_{i+1})/\varepsilon_{i+1})\,\mathrm{d}x_{i+1}}, \qquad (1.2.3)$$

where X_i models the state of the system at time t_i; $\mathscr{W}_{i+1}(x_i, x_{i+1})$ is the classical work done (energy difference plus energy dissipated) in changing from state x_i to state x_{i+1} over the time interval $[t_i, t_{i+1}]$; $\varepsilon_{i+1} > 0$ models the temperature of the heat bath. The aim is then to identify the limiting continuous-time process as mesh of the time discretization tends to zero.

A joint paper [SKTO09] has recently been accepted on this topic. In that paper, and again in this thesis, it is shown that the limiting continuous-time evolution for $\varepsilon_{i+1} = \theta \Delta t_{i+1}$ is a deterministic gradient descent with respect to an *effective dissipation potential*, here denoted $\mathcal{F}_0$. $\mathcal{F}_0$ depends only on the original one-homogeneous dissipation potential Ψ (i.e. the geometry of the elastic region in the dual space); it can, in principle, be computed exactly, and is a "smoothing" of Ψ near its vertex at the origin. Thus, the original rate-independent behaviour is destroyed in a controlled way, and is recovered in the zero-temperature limit. Hence, exact rheological constants, time exponents & c. can be predicted and subjected to experimental validation. In particular, Andrade's $t^{1/3}$ creep law for soft metals [And10] [And14] follows as a corollary. This approach is quite robust in that it can be formulated in much more general spaces than $\mathbb{R}^n$ and the limiting results make at least formal sense if the state space is a manifold or an infinite-dimensional Banach space.

The appendices contain a table of notation, as well as a presentation of standard definitions and results referred to frequently in the main body of the thesis.

Chapter 2

Background

2.1 Introductory Remarks

This chapter consists of a review of the concepts of gradient descents and rate-independent processes as they will be needed in the sequel. The treatment follows the approaches of [AGS08] for gradient descents and [MT04] [Mie05] for rate-independent processes.

The notion of a gradient descent as an evolutionary system governed by an energetic potential and a dissipation potential is much more general than the classical case in which the evolution takes place in $\mathbb{R}^n$ and the dissipation is 2-homogeneous. Indeed, in the general picture, it is not only necessary to cater for more general dissipative potentials, but also for state spaces without a differentiable or even linear structure. A typical example of such a state space is $\mathscr{P}(\mathcal{X})$, the infinite-dimensional simplex of Borel probability measures on a Polish space $\mathcal{X}$. A recent treatment of this generalized setting is [AGS08].

Rate-independent systems fall under the umbrella of gradient descents, and the same time discretization schemes are typically used to prove existence, uniqueness and regularity results. The main technical obstacle in the rate-independent case is the one-homogeneity of the dissipation potential — in fact, it is the absence of superlinear growth at infinity that makes it hard to obtain equicontinuity estimates. As noted in the introduction, rate-independent evolutions have a strongly geometric and convex-analytical flavour, a connection apparently first noticed by Moreau [Mor70] [Mor71] [Mor76].

2.2 Classical Gradient Descents

The most elementary example of a gradient descent is the classical case of an ordinary differential equation of the form

$$\dot{z}(t) = -\nabla E(t, z(t)), \tag{2.2.1}$$

equipped with appropriate initial conditions, where $z\colon [0,T] \to \mathbb{R}^n$ and $E\colon [0,T] \times \mathbb{R}^n \to \mathbb{R} \cup \{\pm\infty\}$ is proper[2.1], bounded below and of smoothness class $\mathcal{C}^2$, say. Very often E is of the form

$$E(t,x) = V(x) - \langle \ell(t), x \rangle$$

for some $V\colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ and $\ell\colon [0,T] \to (\mathbb{R}^n)^*$, called the *applied load*. E is called an *energetic potential* and the solution z is said to be a (*classical*) *gradient descent* in E. A straightforward application of the chain rule yields the following differential equation for the energy along the trajectory followed by z:

$$\frac{\mathrm{d}}{\mathrm{d}t} E(t, z(t)) = -|\nabla E(t, z(t))|^2 + (\partial_t E)(t, z(t)) \tag{2.2.2a}$$

$$= -|\dot{z}(t)|^2 + (\partial_t E)(t, z(t)). \tag{2.2.2b}$$

If the energetic potential $E = V$ has no dependence upon time, then (2.2.2b) is often invoked to show that V is a Lyapunov function (i.e. that $V(z(t))$ is non-increasing) and thus to rule out the possibility of closed, periodic orbits for z. Indeed, the Barbašin–Krasovskiĭ–LaSalle invariance principle [BK52] [LaS76] implies that if $E = V$ is time-independent, of class $\mathcal{C}^2$, and $V(x) \to +\infty$ as $|x| \to +\infty$, then not only is every omega-limit point of the dynamics a critical point of V, but also that almost all initial conditions in $\mathbb{R}^n$ are attracted to the local minima (wells) of V.

Integration of (2.2.2b) over $[a,b] \subseteq [0,T]$ yields

$$E(b, z(b)) + \int_a^b |\dot{z}(t)|^2 \,\mathrm{d}t = E(a, z(a)) + \int_a^b (\partial_t E)(s, z(s)) \,\mathrm{d}s. \tag{2.2.3}$$

This equation, the *classical energy balance*, forms the basis of the generalized

[2.1] An extended real-valued function is said to be *proper* if it is not identically $+\infty$ and nowhere equal to $-\infty$.

theory of gradient descents and hence the global formulation of the rate-independent problem.

It is also worth noting that classical gradient descents satisfy a variational principle [BE76a] [BE76b] and [Nay76a] [Nay76b].

Theorem 2.2.1 (Brézis–Ekeland, Nayroles)**.** *Suppose that the energetic potential $V\colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is convex, proper and lower semicontinuous, that $\ell \in L^2([0,T];(\mathbb{R}^n)^*)$ and that $z_0 \in \{x \in \mathbb{R}^n \mid V(x) \neq +\infty\}$ is given. Then $z \in H^1([0,T];\mathbb{R}^n)$ solves*

$$\begin{cases} \dot{z} \in -\partial V(z) + \ell \\ z(0) = z_0 \end{cases}$$

if, and only if, z minimizes $\mathcal{J}\colon H^1([0,T];\mathbb{R}^n) \to [0,+\infty]$ defined by

$$\begin{aligned} \mathcal{J}[u] := \int_0^T \big(V(u(t)) + V^\star(\ell(t) - \dot{u}(t)) - \langle \ell(t), u(t)\rangle\big)\,\mathrm{d}t \\ + \frac{1}{2}|u(T)|^2 - \frac{1}{2}|u(0)|^2 + |u(0) - z_0|^2, \end{aligned}$$

where $V^\star$ denotes the convex conjugate of V, as defined by (C.2)*. Furthermore, the minimum of $\mathcal{J}$ is $\mathcal{J}[z] = 0$.*

Theorem 2.2.1 is also valid for Hilbert-space-valued evolutions. Generalization of theorem 2.2.1 to other nonlinear gradient-descent-type evolutions, and to the time discretizations thereof, is a topic of ongoing research; see, for example, [Ste08] [Ste09].

2.3 Gradient Descents in Metric Spaces

Classical gradient descents make obvious use of the differentiable and linear structure of n-dimensional Euclidean space. It does not take a great leap of faith to imagine that gradient descents can be easily generalized to Riemannian manifolds. Perhaps surprisingly, though, the general setting for a gradient descent is a complete metric space $(\mathcal{Q}, d)$: this is a huge improvement on $\mathbb{R}^n$ since a general metric space has no locally linear, let alone differentiable, structure. A typical example of the kind of space $\mathcal{Q}$ encountered in applications is the space $\mathrm{SBV}(\Omega;\mathbb{R}^3)$ of deformations of an elastoplastic body $\Omega \subseteq \mathbb{R}^3$ or the infinite-dimensional simplex $\mathscr{P}(\mathcal{X})$ of probability

measures on a Polish space $\mathcal{X}$; on this space, a typical metric is one of the Wasserstein distances.[2.2] For a thorough exposition and treatment of the technical aspects of gradient descents in metric spaces, see [AGS08].

In the metric space setting, as in the classical case, one of the key ingredients is a time-dependent energetic potential $E\colon [0,T] \times \mathcal{Q} \to \mathbb{R}$. What is new in the metric setting is a convex function $\psi\colon [0,+\infty) \to [0,+\infty)$ that controls the dissipation rate; actually, this is not a "new" ingredient, but rather the classical case is that of linear kinetics, in which

$$\psi(r) := \frac{1}{2}r^2 = \psi^\star(r) \text{ and so } \psi' = \mathrm{id},$$

and this choice of ψ goes without explicit mention in the classical setting. It is always required that $\psi(0) = 0$; the analysis is greatly simplified if ψ has superlinear growth at infinity; the failure of this condition complicates the analysis of rate-independent systems.

2.3.1 Gradient Descents in Metric Spaces

To describe gradient descents in metric spaces, it is first necessary to restrict the class of admissible evolutions to the class of absolutely continuous curves.

Definition 2.3.1. Let $(\mathcal{Q}, d)$ be a metric space and let $\mathcal{T} \subseteq \mathbb{R}$ be an interval. A function $u\colon \mathcal{T} \to \mathcal{Q}$ is said to be *absolutely continuous* if, for every $\varepsilon > 0$, there exists a $\delta > 0$ such that whenever $\{[a_i, b_i] \mid i \in I\}$ is a collection of subintervals of $\mathcal{T}$ whose pairwise intersections each contain at most one point,

$$\sum_{i \in I} |b_i - a_i| < \delta \implies \sum_{i \in I} d(u(b_i), u(a_i)) < \varepsilon.$$

It is also necessary to have a notion of "speed" or *metric velocity* for a

[2.2] Trying to get a consensus on what to call, and how to spell, this metric is a non-trivial and probably impossible task. In Russian, the eponymous mathematician is/was Леонид Насонович Васерштейн: phonetically speaking, this surname might be transliterated into English as "Vasershtein" or "Vasershteyn", but it clearly comes from the Germanic name "Wasserstein". This thesis will follow the etymological root, mostly because "Wasserstein distance" has a considerable (but not universal) following in the literature (it gets c. 60,200 Google hits as compared to 310 for "Vasershtein distance"). It is noteworthy that [Was69] treats only a special case of what is now a family of distances, and many other mathematicians, statisticians, economists and computer scientists have contributed to the theory of such distances: according to [Vil09, footnotes to chapter 6] the list includes Gini, Kantorovich, Wasserstein, Mallows, and Tanaka, with other contributions from Salvemini, Dall'Aglio, Hoeffding, Fréchet, Rubinstein and Ornstein.

curve $u\colon [0,T] \to \mathcal{Q}$, denoted by $|\dot{u}|\colon [0,T] \to \mathbb{R}$ and defined by

$$|\dot{u}|(t) := \lim_{h \to 0} \frac{d(u(t+h), u(t))}{|h|}.$$

If u is absolutely continuous, then it is also continuous and has a metric velocity at Lebesgue-almost every time, and, for every $[a,b] \subseteq [0,T]$,

$$d(u(a), u(b)) \leq \int_a^b |\dot{u}|(t)\,\mathrm{d}t.$$

Finally, it is also necessary to have a notion of *metric slope* for a functional $V\colon \mathcal{Q} \to \mathbb{R}$, denoted by $|\nabla V|\colon \mathcal{Q} \to [0,+\infty]$ and defined by

$$|\nabla V|(x) := \limsup_{y \to x} \frac{\max\{V(x) - V(y), 0\}}{d(x,y)}.$$

Note carefully the sign in the definition of metric slope: it is the metric analogue of "the modulus of the downward gradient" and is, therefore, a locally maximal *downward* slope.

Definition 2.3.2. An absolutely continuous curve $u\colon [0,T] \to \mathcal{Q}$ is said to be a *ψ-gradient descent in E* starting at $u_0 \in \mathcal{Q}$ if

1. $u(0) = u(0+) = u_0$;
2. $t \mapsto E(t, u(t))$ is absolutely continuous;[⟨2.3⟩]
3. the (differential) energy inequality

$$\frac{\mathrm{d}}{\mathrm{d}t} E(t, u(t)) \leq (\partial_t E)(t, u(t)) - \psi(|\dot{u}|(t)) - \psi^\star(|\nabla E|(t, u(t))) \quad (2.3.1)$$

   is satisfied for almost every $t \in [0,T]$.

The energy inequality (2.3.1) is the analogue of the classical energy equality (2.2.2b). A ψ-gradient descent u can also be characterized by the *maximal slope condition* that

$$\frac{\mathrm{d}}{\mathrm{d}t} E(t, u(t)) = (\partial_t E)(t, u(t)) - |\nabla E|(t, u(t))|\dot{u}|(t)$$

[⟨2.3⟩] By classical arguments involving the Vitali covering lemma, any absolutely continuous real-valued function on an interval is differentiable almost everywhere; see, for example, [Gor94, chapter 4]. This observation is necessary in order to ensure that (2.3.1) makes sense.

and the *velocity evolution law* that

$$\partial\psi(|\dot{u}|(t)) \ni |\nabla E|(t, u(t));$$

by theorem C.8, the velocity evolution law is equivalent to the condition that

$$|\dot{u}|(t) \in \partial\psi^\star\big(|\nabla E|(t, u(t))\big).$$

(As usual, these relations are only required to hold for Lebesgue-almost every $t \in [0, T]$.)

This thesis will not dwell on the technical aspects of gradient descent theory. Questions of existence, uniqueness and regularity are almost entirely beyond this work, with one exception: the standard method to establish the existence of generalized gradient descents is to resort to time discretization, to show that the discrete-time approximations converge in some suitable sense, and to show that the limit function is a gradient descent in the sense of definition 2.3.2. The discretization scheme is particularly noteworthy for the purposes of this thesis because a suitable regularization of it will be a fundamental object of study in chapters 5 and 6.

2.3.2 Time Discretization

Given an interval of time $[a, b]$, denote by $\mathcal{P}([a, b])$ the set of all *partitions* of $[a, b]$, i.e. finite ordered sets of the form

$$P = \{a = t_0 < t_1 < \ldots < t_N = b\}.$$

For such a partition P, let $|P| := N$, and denote its *mesh* (or *fineness*) by

$$[\![P]\!] := \max_{1 \le i \le |P|} |t_i - t_{i-1}| > 0. \tag{2.3.2}$$

$\mathcal{P}([a, b])$ is a directed set with respect to decreasing mesh size. Note that $\mathcal{P}([a, b])$ can also be directed by refinement of partitions (i.e. containment of sets), but that direction with respect to mesh size is preferable since if $(x^{(P)})_{P \in \mathcal{P}([a,b])}$ is a net in some topological space $\mathcal{X}$ that converges as $[\![P]\!] \to 0$, then it also converges in the refinement preordering, but the converse does not necessarily hold true. There are also more sensitive measures of the fineness of a partition, as used in the gauge integral of Denjoy, Perron,

Henstock and Kurzweil [Gor94].

Δ will be used to denote the backward difference operator acting on sequences (random or deterministic, finite or infinite):

$$\Delta x_i := x_i - x_{i-1},$$

and the telescopic reconstruction formula holds:

$$x_k - x_0 = \sum_{i=1}^{k} \Delta x_i.$$

Definition 2.3.3. Given a partition $P \in \mathcal{P}([0,T])$, the *Moreau–Yosida incremental formulation*[⟨2.4⟩] of the gradient descent is to solve the following sequence of minimization problems: given an initial condition $x_0^{(P)} = x_0 \in \mathcal{Q}$, find, for $i = 1, \dots, |P|$,

$$x_i^{(P)} \in \arg\min \left\{ E(t_i, y) + \psi\left(\frac{d\big(y, x_{i-1}^{(P)}\big)}{\Delta t_i}\right) \Delta t_i \,\middle|\, y \in \mathcal{Q} \right\}. \tag{2.3.3}$$

By abuse of notation, let $x^{(P)} \colon [0,T] \to \mathcal{Q}$ also denote the càdlàg piecewise constant interpolation of the finite sequence $\big(x_i^{(P)}\big)_{i=0}^{|P|}$:

$$x^{(P)}(t) := x_i^{(P)} \text{ for } t \in [t_i, t_{i+1}). \tag{2.3.4}$$

The Moreau–Yosida scheme is a variational problem for which the Euler–Lagrange equations are the equations of motion for the original gradient descent. The Moreau–Yosida scheme is essentially *the* discretization of the original gradient descent, and is often used in numerical methods; the chief difficulty, of course, lies in actually computing the minimizers.

The hope is that, in some suitable topology, $\big(x^{(P)}\big)_{P \in \mathcal{P}([0,T])}$ converges (or at least has a cluster point) as $[\![P]\!] \to 0$, and that the limit is a gradient descent in the sense of definition 2.3.2. Such a convergence theorem can indeed be proved under suitable assumptions on $(\mathcal{Q}, d)$, E and ψ; see [AGS08, chapter 3].

[⟨2.4⟩] This scheme might also be called an implicit Euler scheme, and many authors (including [MT04]) simply call it "the incremental problem". The author prefers the name "Moreau–Yosida incremental formulation" since it highlights the fact that the functional to be minimized is the Moreau–Yosida regularization of $E(t_i, \cdot)$ [Mor65] [Yos65].

2.3.3 Applications of Gradient Descents

An example of the broad reach of gradient descent theory was given by Jordan, Kinderlehrer and Otto in [JKO98]; they showed that the Fokker–Planck equation, which describes the evolution of the probability densities for solutions to a stochastic differential equation, is a gradient descent for a suitable free energy (which incorporates the energetic potential and the Gibbs–Boltzmann entropy) with respect to a Wasserstein metric.

The classical diffusion equation

$$\frac{\partial \rho}{\partial t}(t, x) = \Delta \rho(t, x) \tag{2.3.5}$$

is the Fokker–Planck equation for a standard Brownian motion (Wiener process) on $\mathbb{R}^n$. It is a classical result [CFL28] that the partial differential equation (2.3.5) is the gradient descent of the Dirichlet energy functional with respect to the square of the L^2 dissipation metric. Hence, the Moreau–Yosida discretization on a partition P is, given $\rho_i^{(P)}$, to find

$$\rho_{i+1}^{(P)} \in \arg\min \left\{ \frac{1}{2} \int_{\mathbb{R}^n} |\nabla \rho|^2 + \frac{1}{2\Delta t_{i+1}} \left\| \rho - \rho_i^{(P)} \right\|_{L^2(\mathbb{R}^n)}^2 \,\middle|\, \rho \in \mathcal{K} \right\},$$

for some appropriate class of densities $\mathcal{K} \subseteq L^1(\mathbb{R}^n, \lambda^n; \mathbb{R})$. A corollary of the results of [JKO98] was that another representation of (2.3.5) is as the gradient descent of the negative entropy functional with respect to the Wasserstein metric ℓ^2_{W}, where ℓ^p_{W} for $p \in [1, \infty)$ is defined by

$$\ell^p_{\mathrm{W}}(\mu, \nu) := \inf \left\{ \sqrt[p]{\mathbb{E}\big[|X - Y|^p\big]} \,\middle|\, \begin{array}{c} X, Y \text{ are } \mathbb{R}^n\text{-valued random} \\ \text{variables with laws } \mu, \nu \\ \text{respectively} \end{array} \right\}. \tag{2.3.6}$$

The corresponding Moreau–Yosida discretization on a partition P is, given $\rho_i^{(P)}$, to find

$$\rho_{i+1}^{(P)} \in \arg\min \left\{ \int_{\mathbb{R}^n} \rho \log \rho + \frac{1}{2\Delta t_{i+1}} \ell^2_{\mathrm{W}}\big(\rho, \rho_i^{(P)}\big) \,\middle|\, \rho \in \mathcal{K} \right\},$$

for some appropriate class of densities $\mathcal{K} \subseteq L^1(\mathbb{R}^n, \lambda^n; \mathbb{R})$. Indeed, the following result, theorem 5.1 of [JKO98], identifies a large class of Itō stochastic differential equations as gradient descents of a suitable free energy functional:

Theorem 2.3.4. *Let W denote a standard Wiener process on $\mathbb{R}^n$, defined for times $t \in [0, T]$ and fix $\theta > 0$. Let $E \in \mathcal{C}^\infty(\mathbb{R}^n; [0, +\infty))$ be an energetic potential such that, for some constant $C \geq 0$ and all $x \in \mathbb{R}^n$,*

$$|\nabla E(x)| \leq C|E(x) + 1|.$$

Define a free energy functional F on the set $\mathcal{K}$ of probability densities on $\mathbb{R}^n$ that have finite second moment by

$$F[\rho] := \int_{\mathbb{R}^n} E(x)\rho(x) + \theta\rho(x) \log \rho(x) \, \mathrm{d}x$$

and endow $\mathcal{K}$ with the Wasserstein metric ℓ^2_{W}. Let X solve the Itō stochastic gradient descent

$$\mathrm{d}X(t) = -\nabla E(X(t)) \, \mathrm{d}t + \sqrt{2\theta} \, \mathrm{d}W(t),$$

with an initial condition having a density ρ_0 with finite second moment and $F[\rho_0] < +\infty$. Let $\psi(r) := \frac{1}{2}r^2$. Then the piecewise-constant interpolants of the Moreau–Yosida discrete-time ψ-gradient descent in F on a partition P converge strongly in $L^1((0, T); \mathbb{R}^n)$ as $[\![P]\!] \to 0$ to the solution ρ of the Fokker–Planck equation for X:

$$\frac{\partial \rho}{\partial t}(t, x) = \nabla \cdot \big(\rho(t, x) \nabla E(x)\big) + \theta \Delta \rho(t, x).$$

2.4 Rate-Independent Systems

Dissipation is quite a general concept, referring merely to the loss of energy; gradient descents are examples of dissipative evolutions. Closely related to dissipation is the notion of *plasticity*, which describes a system that is subject to irreversible change. The standard example is that of a physical spring or elastic band, which extends in proportion to the applied load (i.e. elastically, in accordance with Hooke's law) up to some critical load, and then begins to extend much more in response to supercritical loadings. Furthermore, if a subcritical load is released, then the spring returns to its original state, whereas the spring will not return fully to its initial state after being supercritically loaded: this is plastic behaviour.

In situations where inertia can be neglected, a first-order approximation

to the evolution equations is valid. Solutions to such first-order systems often exhibit an additional property: *rate-independence*. Heuristically speaking, a rate-independent system is an evolutionary system that has no intrinsic timescale: its state changes "only as fast as its time-dependent inputs". The physical relevance of rate-independence is that it serves as a good mathematical model for the plastic deformation of materials, and other hysteretic behaviour. There is much literature on the topic of rate-independent processes. For a survey of the theory, modeling and analysis of rate-independent processes, see Mielke's monograph [Mie05] or lectures [Mie07]; the bibliographies therein are a veritable *who's who* of the researchers in the area.

Definition 2.4.1. Let $\mathcal{X}$ and $\mathcal{Y}$ be topological spaces. Suppose that each choice of initial condition $z_1 \in \mathcal{X}$ and each input $\ell\colon [t_1, t_2] \to \mathcal{Y}$ determines a set of outputs

$$\mathcal{O}([t_1, t_2], y_1, \ell) \subseteq \{z\colon [t_1, t_2] \to \mathcal{X} \mid z(t_1) = z_1\}.$$

The input-output relationship is said to be *rate-independent* if, whenever $\alpha\colon [t_1', t_2'] \to [t_1, t_2]$ is strictly increasing and surjective,

$$z \in \mathcal{O}([t_1, t_2], y_1, \ell) \iff z \circ \alpha \in \mathcal{O}([t_1', t_2'], y_1, \ell \circ \alpha).$$

The relationship is said to determine a (possibly multi-valued) *evolutionary system* if concatenations and restrictions of solutions are also solutions, i.e.

$$\begin{aligned} &\hat{z} \in \mathcal{O}([t_1, t_2], z_1, \ell|_{[t_1,t_2]}), \tilde{z} \in \mathcal{O}([t_2, t_3], z_2, \ell|_{[t_2,t_3]}) \text{ and } \hat{z}(t_2) = z_2 \\ &\implies z \in \mathcal{O}([t_1, t_3], z_1, \ell) \text{ where } z(t) := \begin{cases} \hat{z}(t), & \text{if } t \in [t_1, t_2], \\ \tilde{z}(t), & \text{if } t \in [t_2, t_3]; \end{cases} \end{aligned}$$

and

$$\begin{aligned} &z \in \mathcal{O}([t_1, t_2], z_1, \ell), [s_1, s_2] \subseteq [t_1, t_2] \text{ and } y_1 := z(s_1) \\ &\implies z|_{[s_1,s_2]} \in \mathcal{O}([s_1, s_2], z_1, \ell|_{[s_1,s_2]}). \end{aligned}$$

The theory of rate-independent processes is most easily framed in the case in which the state space is a "nice" Banach space and the process is generated by a *convex* energetic potential and a one-homogeneous dissipation potential. In this case, there are several equivalent formulations of the

problem. However, if the energetic potential is not convex, then these formulations cease to be equivalent. For a thorough discussion of these issues, see [MT04]. For later reference, the key points are summarized below. The assumptions on the energetic and dissipation potentials will be standard in the remainder of this thesis.

The general set-up is phrased in terms of a state space, $\mathcal{Q}$, a potential energy functional, E, and a dissipation functional, Ψ:

- The state space, $\mathcal{Q}$, is a closed and convex subset of a separable, reflexive Banach space $\mathcal{X}$. In the easiest cases, $\mathcal{Q} = \mathcal{X}$; if $\mathcal{Q} \neq \mathcal{X}$, then certain "side conditions" arise when the state of the system lies in $\partial\mathcal{Q}$. The elements of $\mathcal{X}$ are referred to as *primal variables*, while the elements of the continuous dual space $\mathcal{X}^*$ are referred to as *dual variables*; it is important not to confuse $\mathcal{X}$ with $\mathcal{X}^*$ even when $\mathcal{X}$ is a Hilbert space or even $\mathbb{R}^n$.

- The energy functional, E, is a proper (i.e. not identically $+\infty$) time-dependent energy functional

$$E\colon [0,T] \times \mathcal{Q} \to \mathbb{R} \cup \{+\infty\}.$$

  $E(t,x)$ is assumed to be bounded below and of smoothness class $\mathcal{C}^1$ in (t,x). Further, it is assumed that $E(t,\cdot)$ is weakly lower semicontinuous and that $(\partial_t E)(t,\cdot)$ is uniformly Lipschitz and weakly continuous. The prototypical situation (the *doubly nonlinear problem* of [CV90]) is that of a fixed potential V and an applied load ℓ, in which case $E\colon [0,T] \times \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ takes the form

$$E(t,x) = V(x) - \langle \ell(t), x \rangle + \chi_{\mathcal{Q}}(x).$$

- The dissipation functional, Ψ, is a convex, one-homogeneous function $\Psi\colon \mathcal{X} \to [0,+\infty)$. It is assumed that there exist constants $c_\Psi, C_\Psi > 0$ such that

$$c_\Psi \|x\| \leq \Psi(x) \leq C_\Psi \|x\| \text{ for all } x \in \mathcal{X}.$$

  This is equivalent to assuming that

$$\Psi(x) = \chi^\star_{\mathscr{E}}(x) = \sup\{\langle \ell, x \rangle \mid \ell \in \mathscr{E}\} \tag{2.4.1}$$

for some (strongly) bounded, (strongly) closed and convex set $\mathscr{E} \subseteq \mathcal{X}^*$ having 0 as an interior point. $\mathscr{E}$ is known as the *elastic region* and its frontier $\partial\mathscr{E}$ is known as the *yield surface*. The rôle of the elastic region is that it describes those loads (energy gradients) that are weak enough that they allow for an elastic change in the state of the system: as a rule of thumb (to be fine-tuned once local stability has been properly introduced in subsection 2.4.2), an elastic load is one for which $-\mathrm{D}E(t,x) \in \mathring{\mathscr{E}}$.

As they were defined in the previous section, gradient descents were required to be absolutely continuous curves. However, it is a feature of rate-independent evolutions that jump discontinuities do occur and must be treated properly in the mathematical analysis. Therefore, solutions to the rate-independent problem in E and Ψ are sought not in $\mathrm{AC}([0,T];\mathcal{X})$ but in (subspaces of) the space $\mathrm{BV}([0,T];\mathcal{X})$ of functions of *bounded variation* with the norm

$$\|z\|_{\mathrm{BV}([0,T];\mathcal{X})} := \|z\|_{L^1([0,T];\mathcal{X})} + \mathrm{Var}_{[0,T]}(z),$$

where the variation seminorm $\mathrm{Var}_{[0,T]}(\cdot)$ is defined by

$$\begin{aligned}\mathrm{Var}_{[0,T]}(z) &\equiv \int_0^T \|\mathrm{d}z\| \\ &:= \sup\left\{ \sum_{j=1}^{n} \|z(t_j) - z(t_{j-1})\| \,\middle|\, \begin{array}{c} 0 = t_0 < t_1 < \ldots < t_n = T, \\ n \in \mathbb{N} \end{array} \right\}.\end{aligned}$$

(By definition, z has bounded variation if $\mathrm{Var}_{[0,T]}(z)$ is finite.) As the alternative notation $\int_0^T \|\mathrm{d}z\|$ suggests, if $z \in W^{1,1}([0,T];\mathcal{X})$, then the variation can be calculated by

$$\mathrm{Var}_{[0,T]}(z) = \int_0^T \|\dot{z}(t)\| \,\mathrm{d}t,$$

and hence $W^{1,1}([0,T];\mathcal{X})$ is continuously embedded in $\mathrm{BV}([0,T];\mathcal{X})$. It is worth noting that the variation functional is lower semicontinuous with respect to the L^1_{loc} topology on $\mathrm{BV}([0,T];\mathcal{X})$.[⟨2.5⟩]

[⟨2.5⟩] Lower semicontinuity in L^1_{loc} also holds true if the compact interval $[0,T] \subsetneq \mathbb{R}$ is replaced by a domain $\Omega \subseteq \mathbb{R}^n$, in which case the L^1 and L^1_{loc} topologies on $\mathrm{BV}(\Omega;\mathcal{X})$ are different.

$\mathrm{BV}_{\pm}([0,T];\mathcal{X}) \subsetneq \mathrm{BV}([0,T];\mathcal{X})$ denotes the subspace of left-continuous $(-)$ or right-continuous $(+)$ functions of bounded variation:

$$z \in \mathrm{BV}_{\pm}([0,T];\mathcal{X}) \iff \begin{cases} z \in \mathrm{BV}([0,T];\mathcal{X}) \text{ and} \\ z(t) = z(t\pm) \text{ for all } t \in [0,T], \end{cases}$$

where $z(t\pm) := \lim_{s \searrow 0} z(t \pm s)$ denotes the right/left limit of z at t.

A sensitive treatment of evolutions of bounded variation needs to take account of the fact that $z \in \mathrm{BV}([0,T];\mathcal{X})$ may be discontinuous, although the only possible discontinuities are jump discontinuities, i.e. points $t \in [0,T]$ for which $z(t+) \neq z(t-)$. The set of such jump points is a countable subset of $[0,T]$, and hence has Lebesgue measure zero [Gor94, chapter 4], and this is rather inconvenient, since it is unsatisfying to have no control over the behaviour of z on its Lebesgue-null jump set. To compensate for this, more sensitive formulations of the rate-independent problem use a *differential measure* μ_z that takes proper account of jump points, and a *reduced derivative* that satisfies an appropriate fundamental theorem of calculus with respect to μ_z.

Definition 2.4.2. Let $z \in \mathrm{BV}_{\pm}([0,T];\mathcal{X})$. Define the *differential measure* $\mu_z \in \mathscr{M}([0,T])$ by

$$\mu_z([s,t)) = t - s + \int_s^t \|\mathrm{d}z\| \tag{2.4.2}$$

and the *reduced derivative* $\mathrm{rd}(z)\colon [0,T] \to \{x \in \mathcal{X} | \|x\| \leq 1\}$ by

$$z(t) - z(s) = \int_{[s,t)} \mathrm{rd}(z)(\tau)\, \mathrm{d}\mu_z(\tau). \tag{2.4.3}$$

A developed treatment of the differential measure and reduced derivative can be found in [MT04, appendix A]; a similar, but slightly different, construction is used in [Mon93, section 0.1]. In summary, the key properties are that if z is differentiable on (a,b), then

$$\mu_z((a,b)) = \int_a^b 1 + \|\dot{z}(t)\|\, \mathrm{d}t, \tag{2.4.4a}$$

$$\mathrm{rd}(z)(t) = \frac{\dot{z}(t)}{1 + \|\dot{z}(t)\|} \text{ for all } t \in (a,b); \tag{2.4.4b}$$

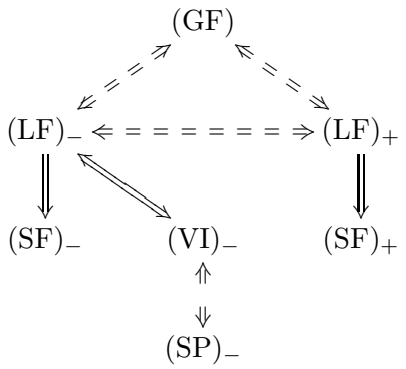


Figure 2.4.1: Implications among the various formulations of the rate-independent problem in an energetic potential E and dissipative potential Ψ. An unbroken implication arrow denotes implication for general energies E satisfying the usual assumptions of Lipschitz continuity & c., whereas a broken implication arrow indicates that the implication requires that E be convex. For (GF), see subsection 2.4.1 on page 23; for (LF) and (VI), see subsection 2.4.2 on page 25; for (SF) and (SP), see subsection 2.4.4 on page 27.

on the other hand, if t is a jump point for z, then

$$\mu_z(\{t\}) = \|z(t+) - z(t-)\|, \tag{2.4.5a}$$

$$\mathrm{rd}(z)(t) = \frac{z(t+) - z(t-)}{\|z(t+) - z(t-)\|}. \tag{2.4.5b}$$

Given appropriate initial conditions in $\mathcal{Q}$, the various formulations of the rate-independent problem are outlined in the following subsections: subscripts $\pm$ are used to denote the required direction of continuity, if any, so $(\mathrm{X})_\pm$ means "formulation X, with solutions sought in $\mathrm{BV}_\pm([0,T];\mathcal{Q})$". The relations among the various formulations are indicated in figure 2.4.1 on page 22, and the proofs can all be found in section 3 of [MT04]. To save space, the initial conditions will never be mentioned explicitly.

As is clear from the nomenclature, the two key ingredients in the global $(\mathrm{GF})_\pm$ and local $(\mathrm{LF})_\pm$ formulations of the rate-independent problem are notions of energy balance (or at least an energy inequality) and stability. The energy constraint is an infinitesimal constraint: time derivatives are

used to define both ($\mathrm{E}_{\mathrm{glob}}$) and ($\mathrm{E}_{\mathrm{loc}}$). On the other hand, the stability constraint is geometric in nature: both ($\mathrm{S}_{\mathrm{glob}}$) and ($\mathrm{S}_{\mathrm{loc}}$) require that the graph of the process should lie in a region of space-time determined solely by the properties of E and Ψ. Some simple rate-independent evolutions in smooth, uniformly convex energies in dimensions one and two are illustrated in figure 2.4.2 on page 24.

2.4.1 Global Formulation

The global formulation of the rate-independent problem is, in some sense, the most general since it requires very few structural assumptions on the state space. Although in this thesis attention is restricted to state spaces $\mathcal{Q}$ that are subsets of Banach spaces, the global formulation makes sense if $\mathcal{Q}$ is only a metric space with no linear structure whatsoever. Only cosmetic modifications to the stability and energy criteria are required, e.g. treating the dissipation Ψ as a metric (or even just a quasimetric) on $\mathcal{Q}$.

Definition 2.4.3. The *global formulation* $(\mathrm{GF})_{\pm}$ of the rate-independent problem in E and Ψ is to find $z \in \mathrm{BV}_{\pm}([0,T];\mathcal{Q})$ satisfying the stability criterion that, for λ-almost all $t \in [0,T]$, for all $y \in \mathcal{Q}$,

$$E(t,z(t)) \leq E(t,y) + \Psi(y - z(t)), \tag{$\mathrm{S}_{\mathrm{glob}}$}$$

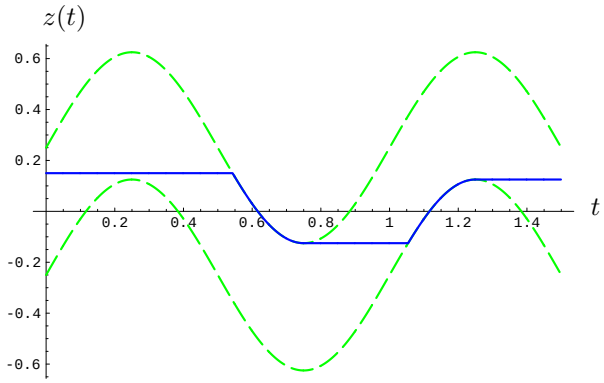
and the energy inequality that, for all $[a,b] \subseteq [0,T]$,

$$E(b,z(b)) + \int_a^b \Psi(\mathrm{d}z) \leq E(a,z(a)) + \int_a^b (\partial_t E)(s,z(s))\,\mathrm{d}s. \tag{$\mathrm{E}_{\mathrm{glob}}$}$$

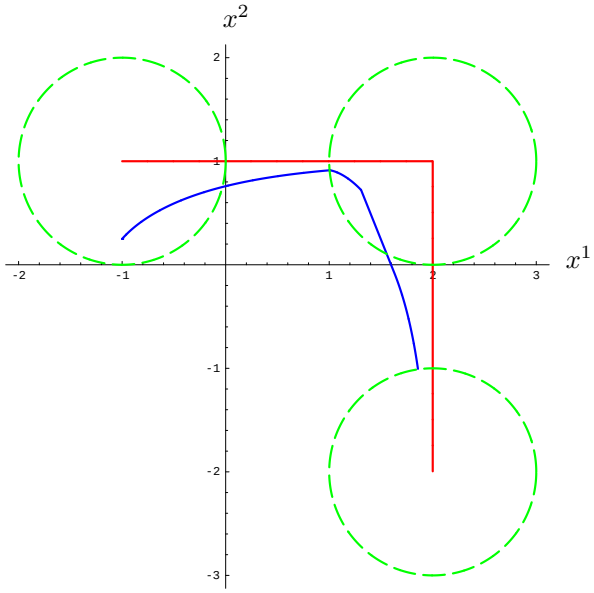
The global energy inequality ($\mathrm{E}_{\mathrm{glob}}$) is the direct analogue of the inequality (2.2.3); the dissipation term is now the Ψ-variation

$$\int_s^t \Psi(\mathrm{d}z) \equiv \sup\left\{ \sum_{j=1}^{n} \Psi(z(t_j) - z(t_{j-1})) \,\middle|\, \begin{array}{c} s = t_0 < t_1 < \ldots < t_n = t, \\ n \in \mathbb{N} \end{array} \right\}.$$

One deficiency of the global formulation is that the global stability criterion ($\mathrm{S}_{\mathrm{glob}}$) is somehow unnatural, since it allows z to explore the entire state space and make arbitrarily large jumps in search of stable states. An advantage of the global formulation is that $(\mathrm{GF})_{\pm}$ generalizes easily to spaces $\mathcal{Q}$ that have no linear structure, since neither ($\mathrm{S}_{\mathrm{glob}}$) nor ($\mathrm{E}_{\mathrm{glob}}$) make any use



(a) In dimension $n = 1$, with $E(t, x) := 4|x|^2 - \langle 3 \sin 2\pi t, x \rangle$, $\Psi(z) := 2|z|$, $z(0) := 0.15$. $\partial \mathcal{S}(t)$ is shown with green dashes.



(b) In dimension $n = 2$, with $E(t, x) := \frac{1}{2}|x|^2 - \langle \ell(t), x \rangle$, $\Psi(z) := |z|$, $z(0) := (-1, \frac{1}{4})$. The trajectory of $\ell^\sharp$ (the centre of the stable region) is shown in red. $\partial \mathcal{S}(t)$ is shown with green dashes at $t = 0, T$ and the the time at which ℓ "turns the corner".

Figure 2.4.2: Some simple rate-independent evolutions $z \colon [0, T] \to \mathbb{R}^n$ (in blue) for $n \in \{1, 2\}$ and uniformly convex energetic potentials.

of spatial derivatives of E. In contrast, the local and variational inequality formulations of the rate-independent problem use the (spatial) Gâteaux derivative $\mathrm{D}E\colon [0,T]\times\mathcal{X}\to\mathcal{X}^*$ of E.

2.4.2 Local and Variational Formulations

The local formulation of the rate-independent problem makes use of the linear structure of the ambient Banach space $\mathcal{X}$. The advantage of doing so is that it allows a more sensitive treatment of stability. Also, whereas the global formulation made use of Lebesgue measure on $[0,T]$, the local formulation uses the differential measure μ_z. The effect of this change is that jump points are treated suitably: by (2.4.5), if t is a jump point for $z\in \mathrm{BV}_{\pm}([0,T];\mathcal{Q})$, then $\mu_z(\{t\})>0$. Hence, requiring that some property holds μ_z-almost everywhere imposes the appropriate non-trivial conditions on the behaviour of z on its jump set.

Definition 2.4.4. The *local formulation* $(\mathrm{LF})_{\pm}$ of the rate-independent problem in E and Ψ is to find $z\in\mathrm{BV}_{\pm}([0,T];\mathcal{Q})$ such that, for μ_z-almost all $t\in[0,T]$, for all $v\in \mathrm{T}_{z(t)}\mathcal{Q}$,

$$\langle \mathrm{D}E(t,z(t)),v\rangle+\Psi(v)\geq 0, \tag{$\mathrm{S}_{\mathrm{loc}}$}$$

where $\mathrm{T}_x\mathcal{Q}:=\{v\in\mathcal{X}\mid \text{for some } \alpha>0, x+\alpha v\in\mathcal{Q}\}$ is the cone of vectors that point into $\mathcal{Q}$, and (again for μ_z-almost all $t\in[0,T]$)

$$\langle \mathrm{D}E(t,z(t)),\mathrm{rd}(z)(t)\rangle+\Psi(\mathrm{rd}(z)(t))\leq 0. \tag{$\mathrm{E}_{\mathrm{loc}}$}$$

Both the requirements $(\mathrm{S}_{\mathrm{loc}})$ and $(\mathrm{E}_{\mathrm{loc}})$ are essentially geometric in nature. $(\mathrm{S}_{\mathrm{loc}})$ constrains $z(t)$ to lie in some region of the state space $\mathcal{Q}$. $(\mathrm{E}_{\mathrm{loc}})$ provides a constraint on the directions in which z may move, although it does not constrain the rate of motion. For example, if $\mathcal{X}$ is a Hilbert space and $\Psi=\|\cdot\|_{\mathcal{X}}$, then $(\mathrm{E}_{\mathrm{loc}})$ is equivalent to the requirement that $\mathrm{rd}(z)(t)$ be a non-negative multiple of the Riesz representative of $-\mathrm{D}E(t,z(t))$, i.e. of $-\nabla E(t,z(t))$.

For general energies E, $(\mathrm{S}_{\mathrm{glob}})$ implies $(\mathrm{S}_{\mathrm{loc}})$, whereas the converse implication requires that E be convex. The local formulation $(\mathrm{LF})_-$ can be rewritten to yield an equivalent variational formulation:

Definition 2.4.5. The *variational inequality formulation* $(\mathrm{VI})_-$ of the rate-independent problem in E and Ψ is to find $z \in \mathrm{BV}_-([0,T];\mathcal{Q})$ such that, for μ_z-almost all $t \in [0,T]$ and for all $v \in \mathcal{X}$,

$$\begin{aligned}&\langle \mathrm{D}E(t,z(t)), v - \mathrm{rd}(z)(t)\rangle + \Psi(v) - \Psi(\mathrm{rd}(z)(t)) \\ &\quad + \chi_{\mathrm{T}_{z(t)}\mathcal{Q}}(v) - \chi_{\mathrm{T}_{z(t)}\mathcal{Q}}(\mathrm{rd}(z)(t)) \geq 0. \end{aligned} \tag{VI}$$

2.4.3 Stable States

Both the global and local formulations of the rate-independent problem make use of some notion of stability. The heuristic is that the assumption that the evolutions of study are quasistatic (and hence relaxation times are much smaller than the time scale over which E changes) enforces a requirement that the state of the process should lie in some set of stable states. Stability means that (either globally or infinitesimally) the energy dissipated by moving to any other state is greater than the energy released by moving to a lower-energy state.

Definition 2.4.6. Given E and Ψ, define $\mathcal{S}_{\mathrm{glob}}(t)$ to be the set of globally stable states at time $t \in [0,T]$:

$$\mathcal{S}_{\mathrm{glob}}(t) := \{x \in \mathcal{Q} \,|\, \forall y \in \mathcal{Q}, E(t,x) \leq E(t,y) + \Psi(y-x)\}\,;$$

similarly, define $\mathcal{S}_{\mathrm{loc}}(t)$ to be the set of locally stable states at time $t \in [0,T]$:

$$\mathcal{S}_{\mathrm{loc}}(t) := \{x \in \mathcal{Q} \,|\, \forall v \in \mathrm{T}_x\mathcal{Q}, \langle \mathrm{D}E(t,x), v\rangle + \Psi(v) \geq 0\}\,.$$

Furthermore, let

$$\begin{aligned}\mathcal{S}_{\mathrm{glob}}([0,T]) &:= \{(t,x) \in [0,T] \times \mathcal{Q} \,|\, x \in \mathcal{S}_{\mathrm{glob}}(t)\}\,; \\ \mathcal{S}_{\mathrm{loc}}([0,T]) &:= \{(t,x) \in [0,T] \times \mathcal{Q} \,|\, x \in \mathcal{S}_{\mathrm{loc}}(t)\}\,.\end{aligned}$$

Since $(\mathrm{S}_{\mathrm{glob}}) \Longrightarrow (\mathrm{S}_{\mathrm{loc}})$, it follows that $\mathcal{S}_{\mathrm{glob}}(t) \subseteq \mathcal{S}_{\mathrm{loc}}(t)$. If E is convex (in which case the two formulations of stability are equivalent and $\mathcal{S}_{\mathrm{glob}}(t) = \mathcal{S}_{\mathrm{loc}}(t)$), then the distinguishing subscript will be dropped, and the set of stable states at time t will be denoted by $\mathcal{S}(t)$. Also, if $\mathcal{Q} = \mathcal{X}$, then

$$\mathcal{S}_{\mathrm{loc}}(t) = \{x \in \mathcal{X} \mid -\mathrm{D}E(t,x) \in \mathscr{E}^\circ\};$$

in the affine case $\mathrm{D}E(t,x) = A(t)x - \ell(t)$, $\mathcal{S}_{\mathrm{loc}}(t) = A(t)^{-1}(\ell - \mathscr{E})$.

It will sometimes be convenient in the sequel to abuse notation and regard the stable set as a function of an externally-applied load instead of time. Therefore, if the energetic potential E takes the form $E(t,x) = V(x) - \langle \ell(t), x \rangle$ and $\mathcal{Q} = \mathcal{X}$, then

$$\mathcal{S}_{\mathrm{loc}}(\ell) := \{x \in \mathcal{X} \mid -\mathrm{D}V(x) + \ell \in \mathscr{E}\},$$

and similarly for $\mathcal{S}_{\mathrm{glob}}$.

Some illustrations of stable sets in the simple case $\mathcal{Q} = \mathcal{X}$ are given in figure 2.4.3 on page 28. It is important to note that the stable region need not be a convex set even if $\mathcal{Q} = \mathcal{X}$ and E is convex; in infinite-dimensional settings this is especially problematic, since the stable region then fails to be weakly closed.

2.4.4 Subdifferential and Sweeping Process Formulations

The subdifferential $(\mathrm{SF})_+$ and sweeping process $(\mathrm{SP})_+$ formulations are those closest in spirit to familiar ordinary differential equations, except that the dissipation potential is now homogeneous of degree one instead of degree two. Both formulations have the advantage of being notationally very compact; the corresponding disadvantage is that they are differential inclusions in which either the time derivative (in $(\mathrm{SF})_+$) or the energy gradient (in $(\mathrm{SP})_+$) is contained within a strong nonlinearity.

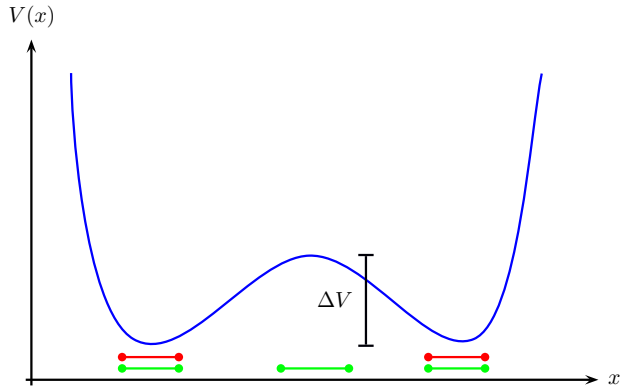
Definition 2.4.7. The *subdifferential formulation* $(\mathrm{SF})_+$ of the rate-independent problem in E and Ψ is to find $z \in \mathrm{BV}_+([0,T]; \mathcal{Q})$ such that, for μ_z-almost all $t \in [0,T]$,

$$0 \in \partial\Psi(\mathrm{rd}(z)(t)) + \mathrm{D}E(t, z(t)) + \partial\chi_{\mathcal{Q}}(z(t)). \tag{SF}$$

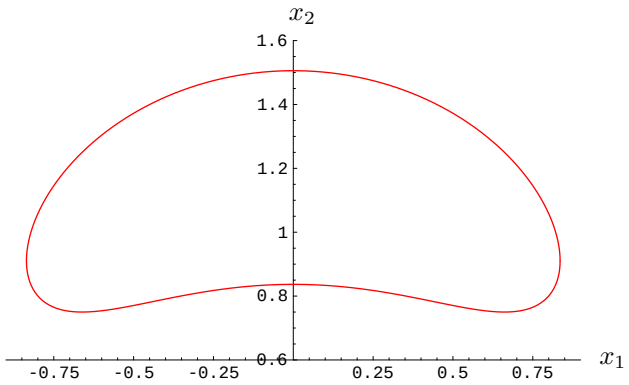
In the prototypical case that $E(t,x) = V(x) - \langle \ell(t), x \rangle + \chi_{\mathcal{Q}}(x)$, $(\mathrm{SF})_+$ is the doubly nonlinear problem of Colli & Visintin [CV90], i.e.

$$\ell(t) \in \partial\Psi(\mathrm{rd}(z)(t)) + \partial V(z(t)),$$

with the addition of the "side conditions" that arise from the constraint that $z(t) \in \mathcal{Q}$ for all $t \in [0,T]$.



(a) Schematic diagram of a non-convex double-well potential V (blue), with $\mathcal{S}_{\text{glob}}$ in red and $\mathcal{S}_{\text{loc}}$ in green. If the potential difference ΔV is large enough, then the central component of $\mathcal{S}_{\text{loc}}$ ceases to be globally stable.



(b) $V(x) := \frac{1}{4}(x_1^2 + x_2^2)^2 - 2x_2$, $\Psi(x) := 2|x|_2$, and $\partial\mathcal{S}$ in red. Note that convexity of V does not imply convexity of the stable region, cf. [MT04, example 5.5].

Figure 2.4.3: Examples of sets of locally and globally stable states in dimensions one and two.

In the case $\mathcal{Q} = \mathcal{X}$, the subdifferential formulation can be rewritten using duality arguments (theorem C.8) to place the nonlinearity on the load instead of the time derivative, which yields the sweeping process formulation of Monteiro Marques et al. [Mon93] [KM97] [KM98]; see also [Smi02].

Definition 2.4.8. The *sweeping process formulation* $(\mathrm{SP})_+$ of the rate-independent problem in E and Ψ is to find $z \in \mathrm{BV}_+([0,T];\mathcal{X})$ such that, for μ_z-almost all $t \in [0,T]$,

$$\mathrm{rd}(z)(t) \in \partial\Psi^\star(-\mathrm{D}E(t,z(t))) \equiv \partial\chi_{\mathscr{E}}(-\mathrm{D}E(t,z(t))). \tag{SP}$$

2.4.5 Incremental Formulation and Well-Posedness

The standard Moreau–Yosida incremental scheme may be used to generate approximate solutions to the rate-independent problem. That is, on a partition P of $[0,T]$, the aim is to solve the sequence of minimization problems

$$z_{i+1}^{(P)} \in \arg\min \left\{ E(t_{i+1}, y) + \Psi\left(y - z_i^{(P)}\right) \,\middle|\, y \in \mathcal{Q} \right\};$$

note that the homogeneity of Ψ eliminates the time step Δt_{i+1} from the minimization problem. As with other gradient descents, the aim is to show that the approximations (discrete-time gradient descents) converge to the continuous-time gradient descent (rate-independent process) in the limit of vanishing mesh size. The main technical difficulty is the fact that the dissipation potential does not grow superlinearly at infinity. It is also unfortunate that not every solution of the global formulation can be obtained via the Moreau–Yosida scheme, and so approximate incremental problems are often used to overcome this difficulty [MR09].

The following result, theorem 7.1 of [MT04], gives well-posedness of the rate-independent problem in the good case that the energy is uniformly convex and the state space is the entire ambient Banach space, in which case the formulations (GF), (LF), (SF), (SP) and (VI) are all equivalent. Appeals will be made to this theorem in chapters 3 and 4.

Theorem 2.4.9. *Suppose that $\mathcal{Q} = \mathcal{X}$ is a separable and reflexive Banach space, that $E(t,\cdot)$ is of smoothness class $\mathcal{C}^3$ and α-uniformly convex. Then, for every initial condition $z_0 \in \mathcal{X}$, there exists a unique solution $z \in W^{1,\infty}([0,T];\mathcal{X})$ to the rate-independent problem in E and Ψ. Furthermore, for each $t \in [0,T]$, $z(t)$ depends continuously on z_0; also, if*

$z^{(P)}\colon [0,T] \to \mathcal{X}$ denotes the càdlàg piecewise constant interpolation of the Moreau–Yosida scheme with respect to a partition $P \in \mathcal{P}([0,T])$, then there exists a constant $C > 0$ (independent of P) such that

$$\left\| z - z^{(P)} \right\|_{L^\infty([0,T];\mathcal{X})} \leq C\sqrt{[\![P]\!]}. \tag{2.4.6}$$

Two important properties of the solution to the incremental problem, used in the proof of theorem 2.4.9, are that

- for all $i = 1, \ldots, |P|$, $z_i^{(P)} \in \mathcal{S}_{\mathrm{glob}}(t_i)$; and that
- if E is α-uniformly convex, then, for all[⟨2.6⟩] $i = 2, \ldots, |P|$,

$$\left\| \Delta z_i^{(P)} \right\| \leq \frac{C}{\alpha} |\Delta t_i|, \tag{2.4.7}$$

  where $C \geq 0$ is any spatial Lipschitz constant for $\partial_t E$.

[⟨2.6⟩] Note that (2.4.7) does not hold for $i = 1$. Counterexamples arise whenever the initial condition is unstable, in which case the initial jump $\Delta z_1^{(P)}$ may be large even if $\partial_t E$ is constant. Such an initial jump is a standard feature in the study of rate-independent processes, and does not affect the temporal regularity too much: it is admissible in both $\mathrm{BV}([0,T];\mathcal{X})$ and $W^{1,\infty}([0,T];\mathcal{X})$, although obviously not in $\mathcal{C}^0([0,T];\mathcal{X})$.

Chapter 3

Wiggly Energies I: the One-Dimensional Case

3.1 Introductory Remarks

This chapter makes rigorous the intuition that a rate-independent evolution in an energetic potential $E\colon [0,T]\times\mathbb{R}\to\mathbb{R}$ ought to arise as a scaling limit of classical gradient descents in perturbed versions E_ε of E. That is, it is shown that the solutions to

$$\dot{z}_\varepsilon(t) = -\frac{1}{\varepsilon}E'_\varepsilon(t, z_\varepsilon(t)) \tag{3.1.1}$$

converge as $\varepsilon\to 0$ to the solution of

$$\partial\Psi(\dot{z}(t)) \ni -E'(t, z(t)) \tag{3.1.2}$$

for a suitable dissipation potential $\Psi\colon\mathbb{R}\to[0,+\infty)$ that is determined by the perturbations $E_\varepsilon - E$; the precise statement of this result is given in theorem 3.3.7. Note that this result is about convergence of trajectories and outside the framework of stability theory: the energies of this chapter satisfy

$$\|E_\varepsilon - E\|_\infty \leq C\varepsilon \text{ but } \lim_{\varepsilon\to 0} E'_\varepsilon \text{ does not exist.}$$

Indeed, this is an *instability* result that shows a qualitative change in the character of the solutions: $(z_\varepsilon)_{\varepsilon>0}$ is a family of rate-*dependent* processes with a rate-*independent* limit.

In the case in which the perturbation is periodic, this result follows from

the theory of averaging [ACJ96] together with a rescaling of time. In this chapter, the class of admissible perturbations is significantly wider: it includes periodic perturbations, but also a large class of well-behaved aperiodic ones, including sample paths of stochastic processes. The required properties, collectively denoted *property* (✠), are given in definition 3.3.2. Random perturbations have been considered before, for example by Grunewald [Gru04] [Gru05]. It is important to note, however, that Grunewald's scaling regime is a different one: essentially, (3.1.1) is considered without the prefactor $\frac{1}{\varepsilon}$. As a result, the limiting process as $\varepsilon \to 0$ does not show stick-slip behaviour; indeed, the velocity of the limiting motion is then positive almost everywhere.

A pointwise convergence version of theorem 3.3.7 was published without detailed proof in [ST07]. The corresponding uniform convergence result and the extension to perturbations that are not of rescaling type are new to this thesis.

3.2 Notation and Set-Up of the Problem

3.2.1 The Energetic Potential and the Perturbation

Denote by $E\colon [0,T] \times \mathbb{R} \to \mathbb{R}$ the underlying energetic potential; for simplicity, assume that E takes the form

$$E(t,x) = V(x) - \langle \ell(t), x \rangle.$$

Assume that $V \in \mathcal{C}^3(\mathbb{R};\mathbb{R})$ is uniformly convex and that $\ell\colon [0,T] \to \mathbb{R}^*$ is uniformly Lipschitz (i.e. that $\ell \in W^{1,\infty}([0,T];\mathbb{R}^*)$). It follows from these assumptions that $(t,x) \mapsto E'(t,x) \equiv V'(x) - \ell(t)$ is uniformly Lipschitz (a prime denotes a spatial derivative) and that V' is strictly increasing.

Fix $\sigma > 0$ and a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ that is rich enough to support Brownian motion (the Wiener process). For each $\varepsilon > 0$, the random perturbation $G_\varepsilon\colon \Omega \times \mathbb{R} \to \mathbb{R}$ is assumed to be $\mathbb{P}$-almost surely differentiable and such that

$$g_\varepsilon := -G_\varepsilon'\colon \Omega \times \mathbb{R} \to [-\sigma, +\sigma]$$

is sample continuous and surjective; finer properties will be specified later. The restriction to perturbations with derivatives taking values in $[-\sigma, +\sigma]$ is not significant: to consider perturbations with derivatives taking values

in $[a, b]$, simply add the appropriate constant to ℓ. The case of perturbations with unbounded gradient, or bounds about which only distributional information is known, is left for further research.

In the periodic setting, it is often assumed that each G_ε is a rescaling of a fixed perturbation G, i.e

$$G_\varepsilon(x) := \varepsilon G(x/\varepsilon), \tag{3.2.1}$$

and so, by the chain rule,

$$g_\varepsilon(x) = g(x/\varepsilon).$$

This is not a necessary restriction, however; the more general viewpoint will be adopted in this thesis, and will be particularly useful in chapter 4. To establish the sought-for convergence theorem, it is required that g_ε attain its bounds "often enough" as $\varepsilon \to 0$; in the case of a rescaling, this "often enough" criterion can be verified on the unrescaled perturbation g, and is denoted property (✠).

For $\varepsilon > 0$, let

$$E_\varepsilon(t, x) := E(t, x) + G_\varepsilon(x).$$

Hence, by the chain rule,

$$-E'_\varepsilon(t, x) = -V'(x) + g_\varepsilon(x) + \ell(t) \in -V'(x) + \ell(t) + [-\sigma, +\sigma].$$

Let $\Psi \colon \mathbb{R} \to [0, +\infty)$ be the non-degenerate one-homogeneous dissipation potential

$$\Psi(v) := \sigma|v|. \tag{3.2.2}$$

The elastic region associated to this dissipation potential is $\mathscr{E} := [-\sigma, +\sigma]$ and so the stable region in the primal space is

$$\begin{aligned} \mathcal{S}(t) &= \{x \in \mathbb{R} \mid -E'(t, x) \in \mathscr{E}\} \\ &= \{x \in \mathbb{R} \mid V'(x) \in [\ell(t) - \sigma, \ell(t) + \sigma]\}. \end{aligned}$$

In the prototypical case of a quadratic energetic potential $V(x) = \frac{\kappa}{2}x^2$, this gives

$$\mathcal{S}(t) = \left[\frac{\ell(t) - \sigma}{\kappa}, \frac{\ell(t) + \sigma}{\kappa}\right].$$

3.2.2 Evolution Equations

With the notation of the previous subsection, the initial value problem of study is

$$\begin{cases} \dot{z}_\varepsilon(t) = -\dfrac{1}{\varepsilon} E'_\varepsilon(t, z_\varepsilon(t)) \\ z_\varepsilon(0) = x_0. \end{cases} \tag{3.2.3}$$

When the perturbation G_ε is a realization of some spatial stochastic process, (3.2.3) is a random ordinary differential equation; that is, it is a deterministic evolution in a randomly chosen energy landscape that is fixed for the duration of the evolution. The purpose of the prefactor $\frac{1}{\varepsilon}$ is to force the system to equilibriate quickly for small ε, thus creating a quasistatic evolution in the limit as $\varepsilon \to 0$.

Denote by $z\colon [0,T] \to \mathbb{R}$ the solution of the rate-independent problem in E with dissipation potential Ψ given by (3.2.2): that is,

$$\begin{cases} \partial\Psi(\dot{z}(t)) \ni -E'(t, z(t)) \\ z(0) = x_0. \end{cases} \tag{3.2.4}$$

The central claim of this chapter is that solutions of (3.2.3) converge to the solution of (3.2.4) as $\varepsilon \to 0$. Before going on to prove this claim, some brief remarks on the existence and uniqueness of solutions are in order.

Since the perturbation G is only assumed to have continuous (not Lipschitz continuous) derivative G', the uniqueness of solutions to (3.2.3) cannot be guaranteed. However, this is not of great importance since solutions do exist, and can be bounded above and below by the unique solutions to

$$\dot{z}_\varepsilon(t) = -\frac{1}{\varepsilon} E'(t, z_\varepsilon(t)) \pm \frac{\sigma}{\varepsilon};$$

indeed, these bounds are used in the proof of lemma 3.5.4.

By theorem 2.4.9, there exists a unique solution $z \in W^{1,\infty}([0,T];\mathbb{R})$ to (3.2.4), and, for each $t \in [0,T]$, the state $z(t)$ depends continuously on the initial condition $x_0 \in \mathbb{R}$. Furthermore, this solution z is characterized by the conditions that

- the right limit $z(0+)$ is the unique closest point of $\mathcal{S}(0)$ to the initial condition x_0;
- z is continuous on $(0,T]$;

- $z(t) \in \mathcal{S}(t)$ for every $t \in (0, T]$;
- if $z(t_1) \in \mathring{\mathcal{S}}(t)$ for all $t \in [t_1, t_2]$, then $z(t) = z(t_1)$ for all $t \in [t_1, t_2]$.

In other words, z is static in the interior of the stable region, and is "pushed along" by the frontier of the stable region if remaining static would mean leaving the stable region. This simple characterization is only possible in dimension one because there is very little freedom in one-dimensional dynamics. In contrast, in the n-dimensional setting, there are many ways for a continuous path to move along the frontier $\partial\mathcal{S}([0, T])$ of the stable region, and this will make the analysis of the next chapter more involved and unavoidably probabilistic in nature.

3.3 Convergence Theorems

The objective is to determine conditions on the perturbation G such that the solutions z_ε to (3.2.3) converge as $\varepsilon \to 0$ to the solution z of (3.2.4). If the (gradient of the) perturbation is periodic, then the result follows from the averaging methods used in [ACJ96]:

Theorem 3.3.1. *Suppose that G' is continuous and periodic with image $[-\sigma, +\sigma]$, that the rescaling* (3.2.1) *holds, and let z_ε solve* (3.2.3)*. Then, up to a subsequence, $z_\varepsilon \to z$ uniformly and $\dot{z}_\varepsilon \overset{*}{\rightharpoonup} z$ in $L^\infty([0, T]; \mathbb{R})$, where z solves* (3.2.4) *with $\Psi(v) := \sigma|v|$.*

To extend this result to non-periodic perturbations requires the introduction of a class of "admissible" perturbations that are "wiggly enough" to ensure the required stick-slip behaviour in the limit. The requisite criterion is denoted *property* (✠). The idea will be to show that $S_\varepsilon(\ell)$, the fixed-point set of the dynamics for (3.2.3) at scale ε with $\ell \in \mathbb{R}^*$ fixed, "fills up" the stable region $\mathcal{S}(\ell)$ as $\varepsilon \to 0$; the required notion of "filling up" is Kuratowski convergence [Kur66] and rôle of property (✠) is to ensure that this convergence holds true (see subsections 3.5.1 and 3.5.2).

3.3.1 Property (✠)

Definition 3.3.2. Fix $\sigma > 0$. A function $g \colon \mathbb{R} \to \mathbb{R}$ is said to have *property* (✠) if

1. g is continuous;

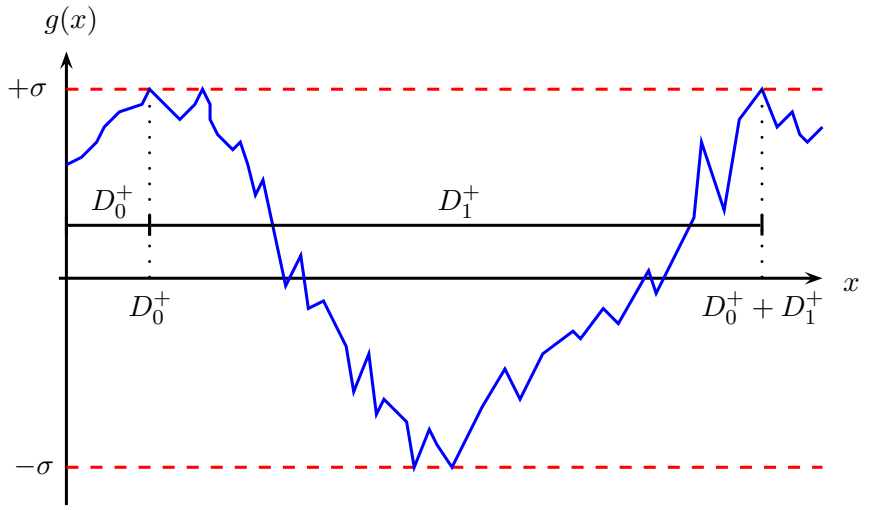


Figure 3.3.1: The distances $D_n^\pm$ as used in the definition of property (✠). A continuous, surjective function $g\colon \mathbb{R} \to [-\sigma, +\sigma]$ is shown in blue, along with the bounds $\pm\sigma$ in red and the distances D_0^+ and D_1^+.

2. the image of g is $[-\sigma, +\sigma]$;

3. define $D_0^+ \geq 0$ to be the least $x > 0$ such that $g(x) = -\sigma$; inductively define D_{n+1}^+ to be the least positive number such that g takes both values $-\sigma$ and $+\sigma$ in the interval

$$\left(\sum_{i=0}^{n} D_i^+, \sum_{i=0}^{n+1} D_i^+ \right];$$

and define $D_n^- \leq 0$ similarly. Then require that

   (a) $D_n^\pm$ exists and is finite for all n;

   (b) $\sum_{n=0}^{\infty} D_n^\pm = \pm\infty$;

   (c) $\lim_{n\to\infty} \left(D_{n+1}^\pm / \sum_{i=0}^{n} D_i^\pm \right) = 0$.

The continuity and surjectivity requirements are natural. Heuristically speaking, the requirements on the $D_n^\pm$ ensure that g has an inexhaustible supply of wiggles at infinity, and that these wiggles are not too widely spaced at infinity. The technical reasons for each of these criteria will become obvious in the proof of lemma 3.5.3.

It is quite easy to see that the class of functions having property (✠) is non-empty and includes the periodic functions:

Example 3.3.3. Any continuous, periodic function $g\colon \mathbb{R} \to \mathbb{R}$ has property (✠) on its image. This is obvious, since, for $n \geq 1$, $D_n^{\pm}$ is exactly the period of g.

Definition 3.3.4. Given $\sigma > 0$, a *doubly reflected Brownian motion* is a stochastic process $P \circ W\colon \Omega \times \mathbb{R} \to \mathbb{R}$ where $P\colon \mathbb{R} \to [-\sigma, +\sigma]$ is the 4σ-periodic extension of

$$\begin{aligned} [-\sigma, 3\sigma] &\to \mathbb{R} \\ x &\mapsto \sigma - |x - \sigma| \end{aligned}$$

to all of $\mathbb{R}$, and $\{W(x)\}_{x \geq 0}$ and $\{W(x)\}_{x \leq 0}$ are independent Brownian motions with the same (random) value at $x = 0$.

Proposition 3.3.5. *Let $\sigma > 0$ and let $g\colon \Omega \times \mathbb{R} \to [-\sigma, +\sigma]$ be a doubly-reflected Brownian motion. Then $\mathbb{P}$-almost every sample path of g has property (✠).*

Proof. It is well-known that $\mathbb{P}$-almost every sample path of W is continuous, so $\mathbb{P}$-almost every sample path of g is continuous. Note also that for g to fail to be surjective it is necessary that W never escape $[-\sigma, +\sigma]$, and that is an event of probability zero: this follows, for example, from the following inequality for a standard Brownian motion in $\mathbb{R}^d$ [DZ98, lemma 5.2.1]:

$$\mathbb{P}\left[\sup_{t \in [0,T]} |W(t)| \geq c\right] \leq 4d \exp\left(-\frac{c^2}{2dT}\right).$$

Define the random variables $D_n^{\pm}\colon \Omega \to [0, +\infty]$ as in the definition of property (✠). By the definition of g as $P \circ W$ and the law of the iterated logarithm for W [Lam77, section 22] (but originally due to Khinchin [Khi24] and Kolmogorov [Kol29]), each $D_n^{\pm}$ is $\mathbb{P}$-almost surely finite. By the strong Markov property for Brownian motion, for $n \geq 1$, $D_n^{\pm}$ are independent and identically distributed. Furthermore, standard results on the exit times of Brownian motions [KS91, p. 253] give that, for $n \geq 1$,

$$\mathbb{E}[D_n^{\pm}] = 16\sigma^2 \text{ and } \mathrm{Var}[D_n^{\pm}] = 512\sigma^4.$$

By the strong law of large numbers, $\sum_{n=1}^{\infty} D_n^{\pm} = \pm\infty$ $\mathbb{P}$-almost surely, and so $\sum_{n=0}^{\infty} D_n^{\pm} = \pm\infty$ $\mathbb{P}$-almost surely. The final part of property (✠) follows from lemma 3.5.8. ■

Indeed, the same strategy of proof can used *mutatis mutandis* to obtain:

Proposition 3.3.6. *Any $\mathbb{P}$-almost surely continuous and surjective process $g\colon \Omega \times \mathbb{R} \to [-\sigma, +\sigma]$ for which the random variables $D_n^{\pm}\colon \Omega \to [0, +\infty]$ are eventually independent and identically distributed with positive mean and finite variance $\mathbb{P}$-almost surely has property (✠).*

3.3.2 Convergence for a Rescaled Perturbation

To simplify matters, it will be assumed that the initial condition is stable, thus making it unnecessary to consider an initial jump at $t = 0$. The first convergence theorem applies in the case of a rescaling of a fixed perturbation as in (3.2.1), and is, therefore, an immediate generalization of the result of theorem 3.3.1.

Theorem 3.3.7. *Suppose that the rescaling relation* (3.2.1) *holds, and that $g := -G'$ has property (✠). Then any $z_\varepsilon\colon [0,T] \to \mathbb{R}$ solving* (3.2.3) *converges uniformly as $\varepsilon \to 0$ to $z\colon [0,T] \to \mathbb{R}$ solving* (3.2.4) *with $\Psi(v) := \sigma|v|$.*

Proof. The claim follows by application of lemmata from the final section of this chapter. By lemma 3.5.6, property (✠) implies that $\{z_\varepsilon\}_{\varepsilon>0}$ is (sequentially) compact in $\mathcal{C}^0([0,T];\mathbb{R})$; by lemma 3.5.7, any convergent subsequence of $\{z_\varepsilon\}_{\varepsilon>0}$ satisfies the rate-independent problem in E and Ψ; by theorem 2.4.9, there is only one solution, z, to the rate-independent problem in E and Ψ. Since the only cluster point of $\{z_\varepsilon\}_{\varepsilon>0}$ is z, and every subsequence of $\{z_\varepsilon\}_{\varepsilon>0}$ has a further subsequence that converges to z, it follows that $z_\varepsilon \to z$ as $\varepsilon \to 0$. ■

Theorem 3.3.7 can be applied to sample paths to yield the following corollary. Note, however, that because theorem 3.3.7 is applied to each sample path individually, no real probabilistic analysis has been done. In some sense, the limited possibilities for one-dimensional dynamics remove the need for a truly probabilistic treatment; this is in contrast to the situation in the next chapter.

Corollary 3.3.8. *Suppose that* $G\colon \Omega \times \mathbb{R} \to \mathbb{R}$ *is such that* $G'\colon \Omega \times \mathbb{R} \to [-\sigma, +\sigma]$ *is a doubly-reflected Brownian motion. Then the stochastic process* $z_\varepsilon\colon \Omega \times [0,T] \to \mathbb{R}$ *solving* (3.2.3) *converges uniformly and almost surely as* $\varepsilon \to 0$ *to the deterministic process* $z\colon [0,T] \to \mathbb{R}$ *solving* (3.2.4) *with* Ψ *as above, i.e.*

$$\mathbb{P}\left[\lim_{\varepsilon \to 0} \sup_{t \in [0,T]} |z_\varepsilon(t) - z(t)| = 0\right] = 1.$$

3.3.3 Convergence for a Family of Perturbations

It is also possible to provide a convergence result with a more probabilistic flavour in the absence of the rescaling relation (3.2.1). For a perturbation G_ε with $g_\varepsilon := -G'_\varepsilon\colon \Omega \times \mathbb{R} \to [-\sigma, +\sigma]$, define a real random variable $R_\varepsilon(\ell)$ by

$$R_\varepsilon(\ell) := \sup\{r > 0 \mid \text{for some } p \in \mathcal{S}(\ell), \mathbb{B}_r(p) \subseteq \mathcal{S}(\ell) \setminus S_\varepsilon(\ell)\},$$

where S_ε is the zero set of the vector field at scale $\varepsilon > 0$ as defined in (3.5.1). $R_\varepsilon(\ell)$ is an upper bound for the distance $|z_\varepsilon(t) - z_\varepsilon(a)|$, $t \geq a$, if $z_\varepsilon(a) \in \mathcal{S}(\ell)$ with ℓ fixed. An upper bound for the random variable $R_\varepsilon(\ell)$ is given by another random variable, M_ε, the maximum of those $D_i^\pm$ whose corresponding intervals intersect the compact set

$$K := \bigcup_{t \in [0,T]} \mathcal{S}(\ell(t)),$$

and the distance of the "last one" from ∂K.

Theorem 3.3.9. *Suppose that* $g_\varepsilon\colon \Omega \times \mathbb{R} \to [-\sigma, +\sigma]$ *is such that, for every* $\lambda \in \ell([0,T])$ $\mathbb{P}\text{-}\lim_{\varepsilon \to 0} R_\varepsilon(\lambda) = 0$. *Then* $\mathbb{P}\text{-}\lim_\varepsilon z_\varepsilon = z$ *in* $\mathcal{C}^0([0,T];\mathbb{R})$*: that is, for all* $\theta > 0$*,*

$$\lim_{\varepsilon \to 0} \mathbb{P}\left[\sup_{t \in [0,T]} |z_\varepsilon(t) - z(t)| > \theta\right] = 0.$$

Proof. The proof is entirely analogous to that of theorem 3.3.7, except that now the control over the separation of the zeroes of the random vector field $-E'_\varepsilon$, and hence the modulus of continuity of z_ε, is provided directly by the assumption on R_ε. Where the Arzelà–Ascoli theorem is used in the proof of theorem 3.3.7, the proof of theorem 3.3.9 requires Prokhorov's theorem as well, i.e. theorem D.6. ■

3.4 Directions for Further Research

There are several obvious directions for further research that would build upon the results of this chapter.

One direction would be to consider more general energetic potentials E. The uniform convexity requirement could be relaxed or dropped entirely, as could the assumption that E takes the form

$$E(t, x) := V(x) - \langle \ell(t), x \rangle.$$

If E is not convex, then the stable region can become disconnected and the solution to the rate-independent problem can exhibit jumps and be non-unique. How the limiting procedure copes with this would be interesting to investigate.

It is also noteworthy that in using property (✠), the bounds on G' were known in advance. What can be said if such bounds are not given *a priori*? What can be said if there is only probabilistic information on the bounds of G', if it has any at all?

3.5 Proofs and Supporting Results

3.5.1 Kuratowski Convergence

Kuratowski convergence, introduced by Kuratowski [Kur66], is a notion of convergence for subsets of topological spaces (although this thesis considers only the metric case); it is precisely the right notion of convergence for making precise the intuition that the zeroes of the vector field $-E'_\varepsilon(t, \cdot)$ "fill up" the stable set $\mathcal{S}(\ell(t))$ as $\varepsilon \to 0$. In the following, the standard notation

$$\operatorname{dist}(x, S) := \inf_{s \in S} d(x, S)$$

is used for the distance from a point x to a subset S in a metric space $(\mathcal{X}, d)$.

Definition 3.5.1. Let $(\mathcal{X}, d)$ be a metric space. The *Kuratowski limit inferior* of a net $(S_\alpha)_{\alpha \in A}$ of subsets of $\mathcal{X}$ is defined by

$$\text{K-}\liminf_\alpha S_\alpha := \left\{ x \in \mathcal{X} \,\middle|\, \limsup_\alpha \operatorname{dist}(x, S_\alpha) = 0 \right\}.$$

Similarly, the *Kuratowski limit superior* of $(S_\alpha)_{\alpha\in A}$ is defined by

$$\text{K-}\limsup_\alpha S_\alpha := \left\{ x \in \mathcal{X} \,\middle|\, \liminf_\alpha \operatorname{dist}(x, S_\alpha) = 0 \right\}.$$

If the Kuratowski limits inferior and superior agree, then their common value is called the *Kuratowski limit* and denoted $\text{K-lim}_\alpha S_\alpha$.

It is a standard fact, and follows quickly from the definition, that Kuratowski limits inferior and superior are always closed sets. For this reason, the Kuratowski limit inferior is sometimes called the *lower closed limit* and the Kuratowski limit superior the *upper closed limit*. The containment

$$\text{K-}\liminf_\alpha S_\alpha \subseteq \text{K-}\limsup_\alpha S_\alpha$$

obviously holds for any net of sets $(S_\alpha)_{\alpha\in A}$. As an easy example, if $S_\varepsilon \subsetneq \mathbb{R}$ denotes the zero set of $x \mapsto \sin(x/\varepsilon)$, then $\text{K-lim}_{\varepsilon\to 0} S_\varepsilon = \mathbb{R}$.

Kuratowski convergence is essentially a pointwise notion of convergence; an example of non-uniform Kuratowski convergence is given by

$$S_\varepsilon := \{x \in \mathbb{R} \mid \sin(x^2/\varepsilon) = 0\},$$

for which $\text{K-lim}_{\varepsilon\to 0} S_\varepsilon = \mathbb{R}$ as above, but

$$\sup_{x\in\mathbb{R}} \operatorname{dist}(x, S_\varepsilon) = +\infty \text{ for all } \varepsilon > 0.$$

However, Kuratowski convergence *is* uniform if the limit set is compact:

Lemma 3.5.2. *Let $(S_\varepsilon)_{\varepsilon>0}$ be a family of subsets of a metric space $(\mathcal{X}, d)$ such that $\text{K-lim}_{\varepsilon\to 0} S_\varepsilon$ is a compact set $S \subseteq \mathcal{X}$. Then*

$$\lim_{\varepsilon\to 0} \sup_{x\in S} \operatorname{dist}(x, S_\varepsilon) = 0.$$

Proof. For a contradiction, suppose not. Then

$$\limsup_{\varepsilon\to 0} \sup_{x\in S} \operatorname{dist}(x, S_\varepsilon) = L > 0.$$

Thus, there exists a sequence $(x_k)_{k\in\mathbb{N}}$ of points in S and a null sequence $(\varepsilon_k)_{k\in\mathbb{N}}$ such that

$$\operatorname{dist}(x_k, S_{\varepsilon_k}) \geq \frac{L}{2}.$$

Since S is (sequentially) compact, it has a convergent subsequence $(x_{k(i)})_{i\in\mathbb{N}}$, converging to some $x_* \in S$. The reverse triangle inequality for the distance-to-a-fixed-subset function implies that

$$\left|\operatorname{dist}\left(x_{k(i)}, S_{\varepsilon_{k(i)}}\right) - \operatorname{dist}\left(x_*, S_{\varepsilon_{k(i)}}\right)\right| \leq d\left(x_{x_{k(i)}}, x_*\right).$$

The second and third terms both tend to zero as $i \to \infty$, but the first term is bounded below by $L/2$. Hence, $x_* \notin \text{K-lim}_{i\to\infty} S_{\varepsilon_{k(i)}}$; hence, since

$$\limsup_{\varepsilon\to 0} \operatorname{dist}(x_*, S_\varepsilon) \geq \limsup_{i\to\infty} \operatorname{dist}\left(x_*, S_{\varepsilon_{k(i)}}\right) > 0,$$

it follows that $x_* \notin \text{K-lim}_{\varepsilon\to 0} S_\varepsilon =: S$. This completes the proof. ■

3.5.2 Stability

The first step in the proof of the convergence theorems of this chapter is to use Kuratowski convergence to make rigorous the intuition that the zeroes of the vector field $-E'_\varepsilon(t,\cdot)$ "fill up" the stable set $\mathcal{S}(\ell(t))$:

Lemma 3.5.3. *Let $S_\varepsilon(\ell)$ denote the fixed-point set of the dynamics for* (3.2.3) *at scale ε with $\ell \in \mathbb{R}^*$ fixed, i.e.*

$$S_\varepsilon(\ell) := \{x \in \mathbb{R} \mid E'_\varepsilon(t,x) \equiv -V'(x) + g_\varepsilon(x) + \ell = 0\}, \tag{3.5.1}$$

and assume that the rescaling (3.2.1) *holds, i.e. $g_\varepsilon(x) = g(x/\varepsilon)$. Then*

$$g \text{ has property } (\maltese) \implies \text{K-}\liminf_{\varepsilon\to 0} S_\varepsilon(\ell) = \mathcal{S}(\ell).$$

Proof. Since obviously $S_\varepsilon(\ell) \subseteq \mathcal{S}(\ell)$ for every $\varepsilon > 0$ and $\mathcal{S}(\ell)$ is closed, it follows that

$$\text{K-}\liminf_{\varepsilon\to 0} S_\varepsilon(\ell) \subseteq \mathcal{S}(\ell).$$

Since the left-hand side is closed, it suffices to show that

$$g \text{ satisfies } (\maltese) \implies \mathring{\mathcal{S}}(\ell) \subseteq \text{K-}\liminf_{\varepsilon\to 0} S_\varepsilon(\ell).$$

Let x be an interior point of $\mathcal{S}(\ell)$. Since V is strictly convex, this is equivalent to the assumption that $-V'(x) + \ell \in \mathring{\mathscr{E}}$. Fix any $r > 0$. It is required to show that property $(\maltese)$ is a sufficient condition for it to hold true that, for all sufficiently small $\varepsilon > 0$, $S_\varepsilon(\ell) \cap (x - r, x + r)$ is non-empty.

This will follow from the intermediate value theorem if it can be shown that, for all sufficiently small $\varepsilon > 0$, $(x - r, x + r)$ contains both a maximum and a minimum of $x \mapsto g_\varepsilon(x) \equiv g(x/\varepsilon)$.

Define the distances $D_n^\pm$ as in the definition of (✠); without loss of generality, assume that $x > 0$ and work with D_n^+. If g only attains its extremes within some bounded subset of $\mathbb{R}$, then $S_\varepsilon(\ell) \cap (x - r, x + r) = \emptyset$ for small enough ε; thus, conditions (a) and (b) are necessary. The strategy also fails if, having taken $\varepsilon_r, n_r > 0$ such that

$$\varepsilon_r \left[\sum_{i=0}^{n_r} D_i^+, \sum_{i=0}^{n_r+1} D_i^+ \right] \subseteq (x - r, x + r),$$

there exists some "later" interval that is "too big" in the sense that the rescaled interval does *not* fit within $(x - r, x + r)$, i.e., for some $n > n_r$,

$$\frac{\sum_{i=0}^{n+1} D_i^+}{\sum_{i=0}^{n} D_i^+} = 1 + \frac{D_{n+1}^+}{\sum_{i=0}^{n} D_i^+} \geq \frac{x + r}{x - r}.$$

Condition (c) of (✠) ensures that there is only a finite number of such "bad" intervals, and so property (✠) implies the claimed Kuratowski convergence on the positive half-line. The case $x < 0$, using D_n^-, is similar; the case $x = 0$ uses both $D_n^\pm$. ■

Lemma 3.5.4 (Almost-stability for positive time)**.** *For all $r > 0$, for all $\tau > 0$, and all $x_0 \in \mathbb{R}$, it holds true that for all small enough $\varepsilon > 0$,*

$$\operatorname{dist}(z_\varepsilon(t), \mathcal{S}(\ell(t))) \leq r \text{ for all } t \in [\tau, T] \subsetneq [0, T].$$

Furthermore, if $x_0 \in \mathcal{S}(\ell(0))$, then the same conclusion holds for $\tau \geq 0$.

Proof. The proof has two parts: showing that z_ε must get close to the stable "sausage" $\mathcal{S}(\ell(\cdot))$; and showing that, once close, it cannot later go far away. For æsthetic reasons, the second part will be shown first. For brevity, denote by $\mathcal{S}_r(\ell)$ the closed r-neighbourhood of $\mathcal{S}(\ell)$:

$$\mathcal{S}_r(\ell) := \bigcup_{x \in \mathcal{S}(\ell)} \overline{\mathbb{B}_r(x)} = \{x \in \mathbb{R} \mid \operatorname{dist}(x, \mathcal{S}(\ell)) \leq r\}.$$

It is claimed that, for sufficiently small ε, the vector field $-\varepsilon^{-1} E'_\varepsilon$ is inward-pointing on $\partial \mathcal{S}_r([0, T])$, thus ruling out the possibility that z_ε may

escape $\mathcal{S}_r([0,T])$. The strict convexity of V implies that, on $\partial\mathcal{S}_r(\ell(t))$, $|E'(t,\cdot)| > \sigma$; furthermore, since E' is continuous, the extreme value theorem implies that $|E'|$ is uniformly bounded away from σ on $\partial\mathcal{S}_r([0,T])$: there exists $\sigma_0 > \sigma$ such that

$$\begin{aligned} t \in [0,T], x \in \partial\mathcal{S}_r(\ell(t)) &\implies |E'(t,x)| \geq \sigma_0 \\ &\implies |E'_\varepsilon(t,x)| \geq \sigma_0 - \sigma. \end{aligned}$$

The two curves that comprise the two connected components of $\partial\mathcal{S}_r([0,T])$ are uniformly Lipschitz with some Lipschitz constant $L > 0$; denote the upper/lower components by $\partial_\pm\mathcal{S}_r([0,T])$ in the natural way. The Lipschitz condition implies that if $x \in \partial_\pm\mathcal{S}_r(\ell(t))$, then every vector of the form $(1, \mp y) \in \mathbb{R}^{1+1}$ with $y > L$ points into $\mathcal{S}_r([0,T])$ at (t,x).

If z_ε were to escape $\mathcal{S}_r([0,T])$ at some time t_*, then it would be impossible for $\frac{\mathrm{d}}{\mathrm{d}t}(t_*, z_\varepsilon(t_*))$ to be an inward-pointing vector. However, if

$$0 < \varepsilon < (\sigma_0 - \sigma)/L \text{ and } z_\varepsilon(t_*) \in \partial\mathcal{S}_r(t_*),$$

then the above arguments imply that

$$\frac{\mathrm{d}}{\mathrm{d}t}(t_*, z_\varepsilon(t_*)) = \left(1, -\frac{1}{\varepsilon}E'_\varepsilon(t_*, z_\varepsilon(t_*))\right)$$

points into $\mathcal{S}_r([0,T])$. Hence, z_ε cannot escape $\mathcal{S}_r([0,T])$ once inside it.

Thus, if the initial condition $x_0 \in \mathcal{S}_r(\ell(0))$, then, for small enough ε, $z_\varepsilon(t) \in \mathcal{S}_r(\ell(t))$ for all $t \in [0,T]$. Therefore, it remains only to consider the case $x_0 \notin \mathcal{S}_r(\ell(0))$. Fix $r > 0$ and, using the Lipschitz bounds on V and ℓ, choose $\tilde{\tau} > 0$ small enough that

$$t \in [0, \tilde{\tau}] \implies \mathcal{S}(\ell(t)) \subseteq \mathcal{S}_{r/2}(\ell(\tilde{\tau})).$$

Thus, $|E'_\varepsilon|$ is locally uniformly bounded away from 0: that is, there exists $\sigma_1 > \sigma$ such that

$$\begin{aligned} t \in [0, \tilde{\tau}], x \notin \mathcal{S}_r(\ell(\tilde{\tau})) &\implies |E'(t,x)| \geq \sigma_1 \\ &\implies \frac{1}{\varepsilon}|E'_\varepsilon(t,x)| \geq \frac{\sigma_1 - \sigma}{\varepsilon}. \end{aligned}$$

Thus, z_ε must lie in $\mathcal{S}_r(\ell(\tilde{\tau}))$ after a time of at most

$$\varepsilon \frac{\operatorname{dist}(x_0, \mathcal{S}_r(\ell(\tilde{\tau})))}{\sigma_1 - \sigma}$$

has elapsed. In particular, for every $\tau > 0$, $z_\varepsilon(\tau) \in \mathcal{S}_r(\ell(\tau))$ for small enough ε, and hence $z_\varepsilon(t) \in \mathcal{S}_r(\ell(t))$ for all $t \geq \tau$. ■

Since $\mathcal{S}([0,T])$ is closed, the almost-stability lemma has the following immediate consequence for every cluster point of the family z_ε, whether the limit is taken in the uniform topology or even just the topology of pointwise convergence:

Corollary 3.5.5. *If any subsequence of z_ε converges pointwise to some $z_0 \colon [0,T] \to \mathbb{R}$, then $z_0(t) \in \mathcal{S}(t)$ for all $t \in (0,T]$. If $x_0 \in \mathcal{S}(0)$, then $z_0(t) \in \mathcal{S}(t)$ for all $t \in [0,T]$.*

Of course, the existence of such cluster points has yet to be established, and to do so will require some compactness arguments. The main tool for compactness in this context is, of course, the Arzelà–Ascoli theorem. The almost-stability lemma will aid in producing the required estimates for the modulus of continuity.

3.5.3 Compactness and Properties of the Limit Process

The fact that the zeroes of the vector field $-E'_\varepsilon(t,\cdot)$ "fill up" $\mathcal{S}(\ell(t))$ in the sense of a Kuratowski limit, and that this "filling up" is uniform, furnishes the required modulus-of-continuity control for a compactness result. Here and in the sequel, ϖ_f denotes the modulus of continuity for a function $f \colon X \to Y$, where X and Y are metric spaces, defined for $\delta > 0$ by

$$\varpi_f(\delta) := \sup\{d_Y(f(x), f(y)) \mid x, y \in X \text{ and } d_X(x,y) \leq \delta\}.$$

Lemma 3.5.6. *Suppose that $z_\varepsilon(0) = x_0 \in \mathcal{S}(0)$. If g has property (✠), then the family $\{z_\varepsilon\}_{\varepsilon>0}$ is relatively compact in $\mathcal{C}^0([0,T];\mathbb{R})$ with its uniform topology.*

Proof. The initial conditions in (3.2.3) imply that $\{z_\varepsilon(0)\}_{\varepsilon>0} = \{x_0\}$, a singleton set. Thus, by the Arzelà–Ascoli theorem (theorem D.5), $\{z_\varepsilon\}_{\varepsilon>0}$

is relatively compact (i.e. has a convergent subsequence) if, and only if,

$$\lim_{\delta\to 0}\limsup_{\varepsilon\to 0}\varpi_{z_\varepsilon}(\delta)=0.$$

By property (✠), lemma 3.5.3 and lemma 3.5.2, for all $\lambda\in\mathbb{R}^*$,

$$\lim_{\varepsilon\to 0}\sup_{x\in\mathcal{S}(\lambda)}\operatorname{dist}(x,S_\varepsilon(\lambda))=0.$$

Given $\theta>0$, lemma 3.5.4 implies that, for all small enough ε and all $t\in[0,T]$, $\operatorname{dist}(z_\varepsilon(t),\mathcal{S}(\ell(t)))\leq\frac{\theta}{2}$. Let $[a,a+\delta]\subseteq[0,T]$ be such that

$$\varpi_{z_\varepsilon}(\delta)=\sup_{t\in[a,a+\delta]}|z_\varepsilon(t)-z_\varepsilon(a)|$$

Let $\ell^+:=\sup_{s\in[a,a+\delta]}\ell(s)$ and $\ell^-:=\inf_{s\in[a,a+\delta]}\ell(s)$. Then, since $z_\varepsilon(t)$ must remain within $\frac{\theta}{2}$ of $\mathcal{S}(\ell(t))$ for all $t\in[a,a+\delta]$, and cannot pass through $S_\varepsilon(\ell^+)$ and $S_\varepsilon(\ell^-)$,

$$\varpi_{z_\varepsilon}(\delta)\leq\max_{\lambda\in\{\lambda^-,\lambda^+\}}\sup_{x\in\mathcal{S}(\lambda)}\operatorname{dist}(x,S_\varepsilon(\lambda))+\frac{\theta}{2}+C\delta\varpi_\ell(\delta).$$

By hypothesis, as $\varepsilon\to 0$, the first term tends to zero; the second term is arbitrary and can be taken to be, say, $\theta/2$; as $\delta\to 0$, the third term surely tends to zero. Hence, the Arzelà–Ascoli criterion is satisfied and the proof of the lemma is complete. ■

Lemma 3.5.7. *Suppose that the initial condition $z_\varepsilon(0)=x_0\in\mathcal{S}(0)$ and let z_0 be any cluster point of the family $\{z_\varepsilon\}_{\varepsilon>0}$ in $\mathcal{C}^0([0,T];\mathbb{R})$. Then z_0 satisfies the rate-independent problem in E and Ψ.*

Proof. As observed in the section on the set-up of the problem, there is a unique solution $z\in W^{1,\infty}([0,T];\mathbb{R})\supsetneq\mathcal{C}^0([0,T];\mathbb{R})$ to the rate-independent problem in E and Ψ. Thus, it suffices to show that $z_0=z$. This will be shown by demonstrating that z_0 satisfies the characterization of z given on page 34. Let $(z_{\varepsilon_k})_{k\in\mathbb{N}}$ be a subsequence of $(z_\varepsilon)_{\varepsilon>0}$ that converges uniformly to some continuous limit function $z_0\colon[0,T]\to\mathbb{R}$.

Since z_0 is a uniform limit of a sequence of continuous functions, it is continuous, and obviously $z_0(0)=z_\varepsilon(0)=x_0$. Hence, the first and second criteria of the characterization are met. By corollary 3.5.5, $z_0(t)\in\mathcal{S}(\ell(t))$ for all $t\in[0,T]$, thus validating the third criterion. Finally, suppose that

$z_0(a) \in \mathring{\mathcal{S}}(\ell(t))$ for all $t \in [a, a+\delta]$; this is equivalent to the assumption that $z_0(a) \in \mathring{\mathcal{S}}(\ell^-) \cap \mathring{\mathcal{S}}(\ell^+)$, where $\ell^\pm$ are the upper/lower extreme values of ℓ on $[a, a+\delta]$. Then, since $\lim_{k\to\infty} z_{\varepsilon_k}(a) = z_0(a)$, for large enough k, $z_{\varepsilon_k}(a) \in \mathring{\mathcal{S}}(\ell^-) \cap \mathring{\mathcal{S}}(\ell^+)$. Then

$$\sup_{t\in[a,a+\delta]} \left|z_{\varepsilon_k}(t) - z_{\varepsilon_k}(a)\right| \leq \max_{\lambda\in\{\ell^-,\ell^+\}} \sup_{x\in\mathcal{S}(\lambda)} \operatorname{dist}(x, S_\varepsilon(\lambda))$$

which tends to 0 as $\varepsilon \to 0$. Hence, as $k \to \infty$, z_{ε_k} converges uniformly on $[a, a+\delta]$ to the constant function with value $z_0(a)$, as required. ■

3.5.4 A Lemma Concerning Random Variables

The following lemma was used in the proof of proposition 3.3.5 to show that almost every sample path of a doubly-reflected Brownian motion has property (✠). It may be seen as a result in the same vein as Slutskiĭ's theorem(s) on the weak convergence of sums and products of sequences of random variables [GS01, section 7.2] [Slu25]. Typically, however, Slutskiĭ-type results only give convergence in law or in probability.

Lemma 3.5.8. *Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and $N \in \mathbb{N}$. For $i \in \mathbb{N}$, let $X_i \colon \Omega \to [0, +\infty]$ be $\mathbb{P}$-almost surely finite Borel random variables that, for $i \geq N$, are independent and identically distributed with finite second moment, common mean $\mu > 0$ and density ρ. Then*

$$\mathbb{P}\left[\lim_{n\to\infty} \frac{X_{n+1}}{\sum_{i=1}^n X_i} = 0\right] = 1.$$

Proof. By assumption, the ratio of interest is $\mathbb{P}$-almost surely well-defined (i.e. not $\frac{\infty}{\infty}$) for every $n \in \mathbb{N}$. Note that

$$\begin{aligned}
&\mathbb{P}\left[\lim_{n\to\infty} \frac{X_{n+1}}{\sum_{i=1}^n X_i} = 0\right] \\
&\quad= \mathbb{P}\left[\lim_{n\to\infty} \frac{X_{n+1}/n}{\sum_{i=1}^n X_i/n} = 0\right] \\
&\quad\geq \mathbb{P}\left(\left[\lim_{n\to\infty} \frac{X_{n+1}}{n} = 0\right] \cap \left[\liminf_{n\to\infty} \frac{\sum_{i=1}^n X_i}{n} > 0\right]\right) \\
&\quad= \mathbb{P}\left[\lim_{n\to\infty} \frac{X_{n+1}}{n} = 0\right] \mathbb{P}\left[\liminf_{n\to\infty} \frac{\sum_{i=1}^n X_i}{n} > 0\right],
\end{aligned}$$

by the independence of the X_i. Since the X_i are $\mathbb{P}$-almost surely finite, the

different distribution of X_i, $i < N$, to that of X_i, $i \geq N$, is irrelevant, and the strong law of large numbers implies that

$$\mathbb{P}\left[\lim_{n\to\infty} \frac{\sum_{i=1}^{n} X_i}{n} = \mu > 0\right] = 1.$$

Therefore, it just remains to show that $\mathbb{P}[X_{n+1}/n \to 0] = 1$. Fix $\varepsilon > 0$ and define the event

$$E_n^\varepsilon := \{\omega \in \Omega \,|\, X_{n+1}(\omega)/n \geq \varepsilon\} \in \mathscr{F}.$$

By assumption,

$$\mathbb{E}[|X_i|^2] = \int_0^{+\infty} s^2 \rho(s)\,\mathrm{d}s < +\infty,$$

so there exist constants $\delta, C, S > 0$ such that $\rho(s) \leq \frac{C}{s^{2+\delta}}$ for $s \geq S$. For $n \geq \max\left\{N, \frac{S}{\varepsilon}\right\}$,

$$\begin{aligned}
\mathbb{P}(E_n^\varepsilon) &= \int_{n\varepsilon}^{+\infty} \rho(s)\,\mathrm{d}s \\
&\leq \int_{n\varepsilon}^{+\infty} \frac{C}{s^{2+\delta}}\,\mathrm{d}s \\
&= -\frac{C}{1+\delta}\frac{1}{s^{1+\delta}}\bigg|_{s=n\varepsilon}^{+\infty} \\
&= \frac{C}{1+\delta}\frac{1}{(n\varepsilon)^{1+\delta}}.
\end{aligned}$$

That is, $\mathbb{P}(E_n^\varepsilon) = O\big(n^{-(1+\delta)}\big)$ as $n \to \infty$ for some $\delta > 0$. Therefore, the series $\sum_{n=1}^{+\infty} \mathbb{P}(E_n^\varepsilon)$ converges (to a finite value), and so the Borel–Cantelli lemma implies that, $\mathbb{P}$-almost surely, only finitely many $n \in \mathbb{N}$ have $X_{n+1}/n \geq \varepsilon$. Since $\varepsilon > 0$ was arbitrary, $\mathbb{P}[X_{n+1}/n \to 0] = 1$, and the lemma follows. ■

Chapter 4

Wiggly Energies II: the Higher-Dimensional Case

4.1 Introductory Remarks

This chapter establishes the n-dimensional analogue of the results of chapter 3 under some additional modeling assumptions. Again, the objective is to establish that the solutions to

$$\dot{z}_\varepsilon(t) = -\frac{1}{\varepsilon}\nabla E_\varepsilon(t, z_\varepsilon(t)) \tag{4.1.1}$$

converge as $\varepsilon \to 0$ to the solution of

$$\partial\Psi(\dot{z}(t)) \ni -\mathrm{D}E(t, z(t)) \tag{4.1.2}$$

for a suitable dissipation potential $\Psi\colon \mathbb{R}^n \to [0, +\infty)$ that is determined by the perturbations $E_\varepsilon - E$. Unfortunately, an n-dimensional analogue of property (✠) that allows for a deterministic limit theorem has not yet been found. Instead, it is necessary to resort to some modeling and obtain a limit theorem with a truly probabilistic flavour; to the best of the author's knowledge, theorem 4.3.1 is the first result in the literature to establish convergence to a rate-independent limit process in high spatial dimension.

In this chapter, the underlying smooth and convex energetic potential $E\colon [0,T] \times \mathbb{R}^n \to \mathbb{R}$ will be "dented" by a spatial stochastic process G_ε. For $\varepsilon > 0$, $G_\varepsilon\colon \Omega \times \mathbb{R}^n \to \mathbb{R}$ consists of a sum of prototypical "dent functions", quadratic functions supported on balls of radius ε and with inward-pointing

gradient of known magnitude σ at the edge of their support; the centres of the dents are randomly placed using a dilute Poisson point process. Thus, z_ε solves a random ordinary differential equation, where the randomness arises from the choice of randomly dented landscape. The sought-for result is that although the law of z_ε solving (4.1.1) is non-trivial, $\lim_{\varepsilon \to 0} z_\varepsilon$ is well-defined (perhaps up to a choice of subsequence), deterministic, and equals z solving (4.1.2).

It is both interesting and important to note that a probabilistic treatment simplifies the analysis in some ways, in that it circumvents the "grid effects" that can plague periodic averaging and homogenization techniques. For example, suppose that dents of radius $\frac{\varepsilon}{2}$ are centred on the points of $\varepsilon(1,1)+2\varepsilon\mathbb{Z}^2$, and that the dent profile is such that the sought-for dissipation potential is the Euclidean norm on $\mathbb{R}^2$. Let the energetic potential E also be the Euclidean norm. Then the solution z to the rate-independent problem in E and Ψ starting at $(1,0) \in \mathcal{S}(0) \subsetneq \mathbb{R}^2$ is constant with $z(t) \equiv (1,0)$, whereas z_ε encounters no dents and simply rapidly descends along a straight-line path to the origin: z_ε converges pointwise to the discontinuous function

$$z_0(t) := \begin{cases} (1,0), & \text{if } t = 0, \\ (0,0), & \text{if } t > 0. \end{cases}$$

This example is somewhat contrived, but it illustrates how a periodic microstructure can allow the state z_ε to change "too much" for it to have a hope of remaining close to the solution of the rate-independent problem.

A critically important element of the probabilistic analysis is that the prefactor $\frac{1}{\varepsilon}$ in (4.1.1), which is there in order to make the limiting evolution quasistatic, does not cause the norm of z_ε in $\mathrm{BV}([0,T];\mathbb{R}^n)$ to blow up in an uncontrolled way; instead, the random placement of the dents is enough to "trap" z_ε before it goes "too far". The stable region has the property that z_ε, if inside a dent that is contained within the stable region, cannot leave that dent. As a result, the evolution of z_ε consists of a succession of rapid descents from one dent to another and pauses within dents; z_ε leaves a dent precisely when the dent is no longer contained in the stable region. Because the time spent waiting in a dent is inversely proportional to the distance descended, $\mathrm{Var}_{[0,T]}(z_\varepsilon)$ remains finite even as $\varepsilon \to 0$. The random distribution of dents ensures that these rapid descents do not cover too great a distance.

4.2 Notation and Set-Up of the Problem

4.2.1 The Energetic Potential and the Perturbation

In order to simplify the analysis, the energetic potential and perturbation will be quite simple examples. The energetic potential $E\colon [0,T] \times \mathbb{R}^n \to \mathbb{R}$ will be the uniformly convex moving parabola

$$E(t,x) := \frac{\kappa}{2}|x|^2 - \langle \ell(t), x \rangle$$

for some fixed $\kappa > 0$ and $\ell \in W^{1,\infty}([0,T]; (\mathbb{R}^n)^*)$. Throughout,

$$L := \|\partial_t E\|_{L^\infty([0,T];(\mathbb{R}^n)^*)} \equiv \operatorname*{ess\,sup}_{t\in[0,T]} \left|\dot{\ell}(t)\right|,$$

which is a Lipschitz constant for ℓ (possibly after redefinition on a Lebesgue-null set). The perturbation of E will require some modeling.

A common setting for rate-independent evolutions is the progression of a dislocation line through a heterogeneous medium; the heterogeneities may be modeled as small inclusions ("obstacles" or "pinning sites", i.e. local minima) in an otherwise smooth energetic potential. Therefore, in this chapter, the perturbed potential E_ε will be obtained from the underlying potential E by adding some randomly placed "dents", the profile of which is known and specified in advance. The *standard dent function* of strength $\sigma > 0$ centred on $p \in \mathbb{R}^n$ will be a clipped quadratic function like so:

$$D(x; p, \sigma) := \min\left\{0, \frac{\sigma}{2\varepsilon}|x-p|^2 - \frac{\sigma}{2}\right\}.$$

That is, the dent function is a quadratic function supported on the ε-ball about p, with an inward-pointing gradient of magnitude σ at the edge of its support.

The random placement of these dents will be modeled using a Poisson point process [DV08] [Kin93] [SKM87].

Definition 4.2.1. Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and $(S, \mathscr{S}, \mu)$ a measure space. A *Poisson point process* on S is a measurable function $N\colon \mathscr{S} \times \Omega \to [0, +\infty]$ such that

1. for each $\omega \in \Omega$, $N(\cdot, \omega)$ is a counting measure on $(S, \mathscr{S})$;

2. for each $A \in \mathscr{S}$, $N(A, \cdot) \sim \text{Poisson}(\mu(A))$, i.e.

$$\mathbb{P}\big[N(A) = k \in \mathbb{N}_0\big] = \frac{\exp(-\mu(A))(\mu(A))^k}{k!};$$

3. if $\{A_i \mid i \in I\} \subseteq \mathscr{S}$ are disjoint, then $\{N(A_i, \cdot) \mid i \in I\}$ are independent random variables.

The measure μ is called the *intensity* of the process. By abuse of notation, N can be thought of as a random set of points in S.

In this chapter, the *dent process* $\mathcal{O}_\varepsilon$ will be a Poisson point process in $\mathbb{R}^n$ with intensity $\mu_\varepsilon := \varepsilon^{-p}\lambda^n$, where λ^n denotes n-dimensional Lebesgue measure and $p \geq 0$ is a parameter that will be specified shortly. Therefore, $(\Omega, \mathscr{F}, \mathbb{P})$ will denote a fixed probability space that is rich enough to support the dent processes $\mathcal{O}_\varepsilon$. A naïve definition of G_ε would be to set $G_\varepsilon(x) := \sum_{p \in \mathcal{O}_\varepsilon} D(x; p, \sigma)$, but this definition has the disadvantage that if dents overlap, then multiple dent functions may contribute to $G_\varepsilon(x)$ for some points $x \in \mathbb{R}^n$. Therefore, to ensure that the dynamics feel the influence of at most one dent at any given time, the alternative definition

$$G_\varepsilon(x) := \begin{cases} D(x; p^!(x), \sigma), & \text{if } p^!(x) \text{ is the unique closest point of } \mathcal{O}_\varepsilon \text{ to } x, \\ 0, & \text{if there is no such unique closest point,} \end{cases} \tag{4.2.1}$$

will be used. Under both the naïve and refined definitions, the support of the perturbation G_ε is the uniform ε-neighbourhood of $\mathcal{O}_\varepsilon$:

$$\operatorname{supp} G_\varepsilon = \bigcup_{i \in \mathcal{O}_\varepsilon} \overline{\mathbb{B}_\varepsilon(i)}.$$

The purpose of the parameter $p \in [0, +\infty)$ is to control the volume fraction occupied by the support of the dents. For technical reasons that will become apparent during the course of the proof, p will be taken to lie in the range $n - 1 < p < n$. This is a *dilute limit* in which the expected volume fraction occupied by the dents tends to zero in the limit as $\varepsilon \to 0$: that is, if $K \subseteq \mathbb{R}^n$ is any compactum of positive measure, then

$$\lim_{\varepsilon \to 0} \mathbb{E}\left[\frac{\lambda^n(K \cap \operatorname{supp} G_\varepsilon)}{\lambda^n(K)}\right] = \begin{cases} 0, & \text{if } p < n, \\ 1, & \text{if } p > n. \end{cases}$$

Note that the standard dent profile is isotropic, i.e. it "looks the same in all directions". Therefore, it seems reasonable to suppose that the dissipation potential for the limiting evolution will also be isotropic, and controlled by the maximum slope of the dent profile, $\sigma > 0$. Therefore, let

$$\Psi(v) := \sigma|v|. \tag{4.2.2}$$

The corresponding stable region is given by

$$\mathcal{S}(t) = \overline{\mathbb{B}_{\sigma/\kappa}(\kappa^{-1}\ell(t))}.$$

4.2.2 Evolution Equations

As in chapter 3, define the random energy E_ε by

$$E_\varepsilon(t, x) := E(t, x) + G_\varepsilon(x),$$

where the refined definition (4.2.1) of G_ε is used. The random ordinary differential equation of study is then

$$\begin{cases} \dot{z}_\varepsilon(t) = -\dfrac{1}{\varepsilon}\nabla E_\varepsilon(t, z_\varepsilon(t)); \\ z_\varepsilon(0) = x_0. \end{cases} \tag{4.2.3}$$

Since the vector field ∇E_ε has discontinuities and the dents may, in principle, overlap, some sort of "solution convention" is called for. Note that the discontinuities of ∇E_ε are spheres centred on the points of the point process $\mathcal{O}_\varepsilon$ together with hyperplane segments that are perpendicular bisectors of lines joining nearby points of $\mathcal{O}_\varepsilon$; with probability one, this "bad set" has Lebesgue measure zero. Therefore, the solution convention for z_ε will be that it is continuous and piecewise smooth, and whenever it encounters a discontinuity, it is "restarted" on the opposite side of the discontinuity. Furthermore, since z_ε can leave a dent that overlaps with no other dents only when it becomes unstable, it will be taken as part of the convention that this property holds for overlapping dents as well: z_ε will be taken to "freeze" on the hyperplane equidistant between two dent centres, and then to move when it becomes unstable.

As in the previous chapter, denote by $z\colon [0, T] \to \mathbb{R}^n$ the solution of the

rate-independent problem in E with dissipation $\Psi := \sigma|\cdot|$:

$$\begin{cases} \partial\Psi(\dot{z}(t)) \ni -\mathrm{D}E(t, z(t)) \\ z(0) = x_0. \end{cases} \tag{4.2.4}$$

Theorem 2.4.9 ensures that this rate-independent problem has a unique solution $z \in W^{1,\infty}([0,T];\mathbb{R}^n)$ and that, for each $t \in [0,T]$, the state $z(t)$ depends continuously on the initial condition $x_0 \in \mathbb{R}^n$. Unlike the case $n = 1$, it is not possible to characterize z by continuity and stability properties alone, since for $n > 1$ there may be many continuous paths in $\partial\mathcal{S}([0,T])$, only one of which satisfies the appropriate energy constraints.

As noted above, for this E and Ψ, the stable region is a Euclidean ball, with $\mathcal{S}(t) = \overline{\mathbb{B}_{\sigma/\kappa}(\ell(t)/\kappa)}$; furthermore, the solution to the Moreau–Yosida incremental problem can be given explicitly as orthogonal convex projection onto $\mathcal{S}(t_{i+1})$:

$$z_{i+1}^{(P)} = \frac{\ell(t_{i+1})}{\kappa} + \min\left\{\frac{\sigma}{\kappa}, \left|z_i^{(P)} - \frac{\ell(t_{i+1})}{\kappa}\right|\right\} \operatorname{sgn}\left(z_i^{(P)} - \frac{\ell(t_{i+1})}{\kappa}\right)$$

For a schematic comparison of z_ε and the solution to the Moreau–Yosida incremental problem, see figure 4.2.1 on page 55.

4.2.3 Dent Exit and Entry Times

A very useful tool in the analysis will be a sequence of random times determined by the random landscape and the evolution z_ε in that landscape.

Definition 4.2.2. Given z_ε solving (4.2.3) starting at an initial condition x_0, inductively define sequences of (random) times $(\tau_i^{\text{out}})_{i\in\mathbb{N}_0}$ and $(\tau_i^{\text{in}})_{i\in\mathbb{N}}$ as follows:

- let $\tau_0^{\text{out}} := 0$;
- let τ_{i+1}^{in} be the infimum of all times t such that $\tau_i^{\text{out}} \leq t \leq T$ and $z_\varepsilon(t)$ simultaneously lies in the interior of a dent and in $\mathring{\mathcal{S}}(t)$.
- let τ_{i+1}^{out} be the infimum of all times t such that $\tau_{i+1}^{\text{in}} \leq t \leq T$ and $z_\varepsilon(t)$ lies outside the dent that it entered at time τ_{i+1}^{in}.

In order to avoid an unsightly excess of subscripts and superscripts, the obvious dependence of the τ's on the initial condition x_0, on ε, and on the

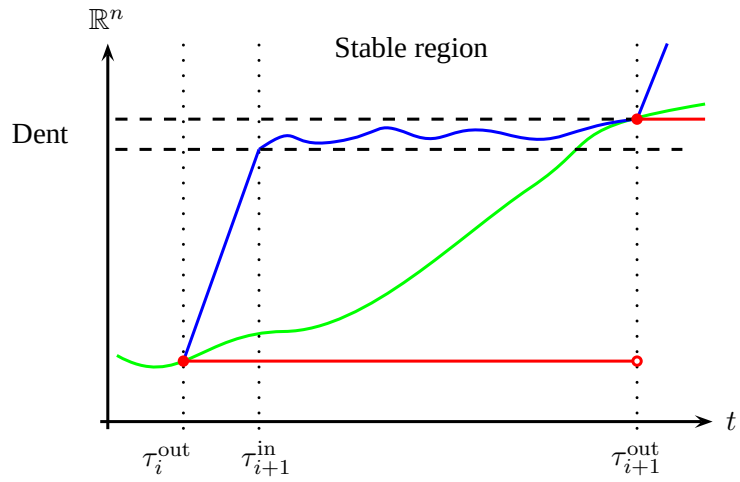


Figure 4.2.1: A "cross-sectional" schematic illustration of z_ε (blue), compared with the solution $z^{(P)}$ (red) of the Moreau–Yosida incremental problem on the random partition generated by z_ε. Everything above the green curve is considered to be stable. Note that z_ε descends rapidly to a dent and waits there, whereas $z^{(P)}$ waits in an unstable state before jumping to the frontier of the stable region. Cf. figure 4.2.2 on page 56.

landscape parameter $\omega \in \Omega$ will be suppressed. Questions of whether or not the random set

$$P(\tau) := \{0 = \tau_0^{\text{out}} \leq \tau_1^{\text{in}} \leq \tau_1^{\text{out}} \leq \tau_2^{\text{in}} \leq \tau_2^{\text{out}} \leq \dots\}$$

forms a partition of $[0, T]$ — and, if so, what the cardinality and mesh of $P(\tau)$ are — will be addressed in the course of the analysis. Key estimates in the proof of the main theorem of this chapter will be ones that establish control over $|\tau_{i+1}^{\text{out}} - \tau_i^{\text{out}}|$ and the amount by which the energy equality (E_{glob}) fails over $[\tau_i^{\text{out}}, \tau_{i+1}^{\text{out}}]$; see lemmata 4.5.5, 4.5.6 and 4.5.7 and theorem 4.5.8. A nasty pathology that must be addressed is the possibility that dents may overlap and give $\tau_i^{\text{out}} = \tau_{i+1}^{\text{in}}$.

4.3 The Convergence Theorem

The main result of this chapter is the following:

Theorem 4.3.1. *Let E, G_ε, z_ε, z be as above and suppose that $p \in (n-1, n)$. Then z_ε converges in distribution to z (with respect to the uniform topology).*

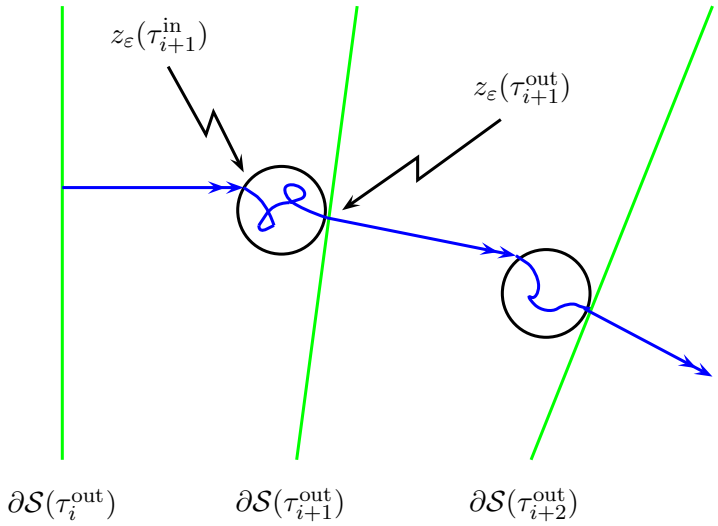


Figure 4.2.2: A "top-down" schematic illustration of z_ε (blue). The frontier of the stable region is shown in green at the three exit times; everything to the right of the green line is the stable region at that time. Dents are shown as black circles. Cf. figure 4.2.1 on page 55.

Since z is deterministic, $z_\varepsilon \to z$ in probability $\mathbb{P}$ as well.

The full proof of theorem 4.3.1 requires many supporting results and occupies the final section of this chapter. A sketch proof, which doubles as a "road map" for the actual proof and outlines the heuristics, follows:

Sketch proof of theorem 4.3.1. The first question to be resolved is one of compactness: does $(z_\varepsilon)_{\varepsilon>0}$ have any convergent subsequences? This can be established using the Arzelà–Ascoli–Prokhorov theorem (theorem D.6), the application of which calls for control of the modulus of continuity of z_ε. The proof of such control is best deferred until later in the proof, since it will follow from much stronger estimates that must be obtained anyway.

Suppose, though, that (up to a choice of subsequence) $(z_\varepsilon)_{\varepsilon>0}$ does converge in distribution to some continuous process $z_0 \colon \Omega \times [0,T] \to \mathbb{R}^n$. What else would be required in order to show that $z_0 = z$? The answer, of course, is to show that z_0 satisfies the global stability and energy criteria (S_{glob}) and (E_{glob}).

Global stability is quite easy to obtain: the prefactor $\frac{1}{\varepsilon}$ in the evolution equation (4.2.3) for z_ε and the boundedness of the perturbation G_ε make it

impossible for z_ε to escape any neighbourhood of the stable region $\mathcal{S}([0,T])$ if ε is small enough (depending on the neighbourhood, of course). This part of the proof is almost identical to the one-dimensional case.

The energy inequality requires much more work. Definition 4.5.3 introduces a numerical measure — called the "energy surplus" and denoted $\mathrm{ES}(u,[a,b])$ — of by how much the global energy inequality ($\mathrm{E}_{\mathrm{glob}}$) fails over any interval of time $[a,b]$ for any process $u\colon [a,b] \to \mathbb{R}^n$; ($\mathrm{E}_{\mathrm{glob}}$) is just the statement that $\mathrm{ES}(u,[a,b]) \leq 0$ for all $[a,b] \subseteq [0,T]$. $\mathrm{ES}(\cdot,[a,b])$ is lower semicontinuous with respect to uniform convergence (lemma 4.5.4), so

$$\liminf_{\varepsilon\to 0} \mathrm{ES}(z_\varepsilon,[0,T]) \leq 0 \implies z_0 \text{ satisfies } (\mathrm{E}_{\mathrm{glob}}).$$

In the process of showing that the left-hand side of this implication is true in a suitable probabilistic sense (theorem 4.5.8), estimates for the variation $\mathrm{Var}_{[0,T]}(z_\varepsilon)$ are obtained, and these are more than sufficient to provide the modulus-of-continuity estimates that were postponed earlier (lemma 4.5.9).

Therefore, (up to a subsequence) $(z_\varepsilon)_{\varepsilon>0}$ converges in distribution to z; since z is deterministic and unique, the convergence actually holds in probability $\mathbb{P}$. Also, since every subsequence of $(z_\varepsilon)_{\varepsilon>0}$ has a further subsequence that converges (in distribution and in probability $\mathbb{P}$) to z, $\mathbb{P}\text{-}\lim_{\varepsilon\to 0} z_\varepsilon = z$. This completes the sketch proof. □

4.4 Directions for Further Research

There are three main directions for further research that would build upon the results of this chapter.

The first direction would be to consider more general energetic potentials E. This complicates the analysis in many ways: the stable region may cease to be convex; the energy E itself may cease to be convex; the solutions to the rate-independent problem may cease to be unique; it may not even be clear which formulation of the rate-independent problem is "the right one".

The second direction would be to consider more general perturbations. It is, admittedly, somewhat unsatisfying to resort to such a specific model for the perturbation in this chapter, having obtained the nice characterization of property (✠) in chapter 3. There are many other point processes that could be used to scatter dents of a fixed and known profile throughout the state space: consider, for example, Matérn clustering and hard core processes

[Mat60]. The main technical difficulty is that exact calculations (e.g. of the nearest-neighbour distribution) are very difficult for any point process other than the Poisson point process; Markov chain Monte Carlo methods are typically used to simulate such processes, sometimes perfectly [Ken05].

In general, if scattered-dent perturbations are used, then it is to be expected that the dent profile will affect the dissipation potential of the limiting rate-independent evolution. If even more general perturbations are considered — what constitutes a "generic" bounded, gradient-type vector field on $\mathbb{R}^n$? — then the situation is even more interesting.

The third direction for generalization is to infinite-dimensional spaces, which are the natural setting for many problems in elastoplasticity. The immediate obstacle to be overcome is that the model class in this chapter is quite narrow: a dilute Poisson process with respect to Lebesgue measure. Although infinite-dimensional point processes do exist, they cannot be translation-invariant since there is no infinite-dimensional Lebesgue measure. It is not immediately clear whether Gaussian measures (the usual choice of "decent" measure on infinite-dimensional spaces) will generate the right kind of point process for the arguments of this chapter to generalize to infinite-dimensional settings.

4.5 Proofs and Supporting Results

This section contains the supporting results and proofs necessary to establish theorem 4.3.1. The route is a little roundabout, but takes in the following key points:

- the family $\{z_\varepsilon\}_{\varepsilon>0}$ is uniformly tight, and therefore has a subsequence that converges in law to some (possibly not deterministic) process z_0;
- z_0 satisfies the stability inequality;
- z_0 satisfies the energy inequality;
- since there is a unique continuous process that is both stable and satisfies the energy inequality, z_0 must be that (deterministic) process.

4.5.1 Stability

As in the one-dimensional case, stability for the limit process is established by showing that every neighbourhood of the stable region is inescapable for

sufficiently small ε.

Lemma 4.5.1 (Almost-stability for positive time)**.** *For all $r > 0$, for all $\tau > 0$, and all $x_0 \in \mathbb{R}$, it holds true that for all small enough $\varepsilon > 0$,*

$$\operatorname{dist}(z_\varepsilon(t), \mathcal{S}(t)) \leq r \text{ for all } t \in [\tau, T] \subsetneq [0, T].$$

Furthermore, if $x_0 \in \mathcal{S}(0)$, then the same conclusion holds for $\tau \geq 0$.

Proof. Note that the gradient of the perturbation is bounded above, almost everywhere in space, by σ. The remainder of the proof is simply an n-dimensional re-write of lemma 3.5.4, and is omitted for brevity's sake. ■

Since $\mathcal{S}([0, T])$ is closed, the almost-stability lemma has the following immediate consequence for every cluster point of the family z_ε, whether the limit is taken in the uniform topology or even just the topology of pointwise convergence:

Corollary 4.5.2. *If any subsequence of z_ε converges pointwise to some $z_0 \colon [0, T] \to \mathbb{R}$, then $z_0(t) \in \mathcal{S}(t)$ for all $t \in (0, T]$. If $x_0 \in \mathcal{S}(0)$, then $z_0(t) \in \mathcal{S}(t)$ for all $t \in [0, T]$.*

4.5.2 Variation Estimates and Energy Balance

Definition 4.5.3. Given $E \colon [0, T] \times \mathbb{R}^n \to \mathbb{R}$ and $\Psi \colon \mathbb{R}^n \to [0, +\infty)$, let $\mathrm{ES}(u, [a, b])$ denote the *energy surplus* of $u \colon [0, T] \to \mathbb{R}^n$ over $[a, b] \subseteq [0, T]$, defined by

$$\mathrm{ES}(u, [a, b]) := E(b, u(b)) - E(a, u(a)) + \int_a^b \Psi(\mathrm{d}u) - \int_a^b (\partial_t E)(s, u(s)) \, \mathrm{d}s. \tag{4.5.1}$$

The energy surplus is a quantitative measure of the amount by which the global energy inequality ($\mathrm{E}_{\mathrm{glob}}$) fails; ($\mathrm{E}_{\mathrm{glob}}$) is simply the statement that $\mathrm{ES}(u, [a, b]) \leq 0$ for all $[a, b] \subseteq [0, T]$. In the prototypical situation of $\Psi := \sigma |\cdot|$ and $E(t, x) := V(x) - \langle \ell(t), x \rangle$, as is the case in this chapter, the energy surplus is

$$\mathrm{ES}(u, [a, b]) = E(t, u(t))\big|_{t=a}^{\ b} + \sigma \mathrm{Var}_{[a,b]}(u) + \int_a^b \left\langle \dot{\ell}(t), u(t) \right\rangle \mathrm{d}t.$$

Note that the energy surplus is additive over non-overlapping subintervals of $[0,T]$: for all $0 \le a \le b \le c \le T$,

$$\mathrm{ES}(u,[a,c]) = \mathrm{ES}(u,[a,b]) + \mathrm{ES}(u,[b,c]).$$

In particular, if $P \in \mathcal{P}([0,T])$ is any partition and the energy surplus is non-positive on each interval of P, then the energy surplus is non-positive on $[0,T]$, i.e. ($\mathrm{E}_{\mathrm{glob}}$) holds true. One key analytical property of the energy surplus functional deserves special mention:

Lemma 4.5.4 (Lower semicontinuity of the energy surplus)**.** *Assume that E satisfies the usual assumptions (in particular, E and $(\partial_t E)(t,\cdot)$ are continuous) and that $\Psi := \sigma|\cdot|$. Then, for every $[a,b] \subseteq [0,T]$, $\mathrm{ES}(\cdot,[a,b])$ is lower semicontinuous with respect to the uniform topology.*

Proof. Suppose that $u_k \to u\colon [a,b] \to \mathbb{R}^n$ uniformly as $k \to \infty$. Since E is continuous, both the first two terms in the energy surplus converge: for $t \in \{a,b\}$,

$$E(t,u(t)) = \lim_{k\to\infty} E(t,u_k(t)).$$

Since $u_k \to u$ uniformly, $u_k \to u$ in the L^1 topology as well, and it is a standard result that the variation functional is lower semicontinuous with respect to the L^1_{loc} topology. Therefore,

$$\int_a^b \Psi(\mathrm{d}u) = \sigma \mathrm{Var}_{[a,b]}(u) \le \sigma \liminf_{k\to\infty} \mathrm{Var}_{[a,b]}(u_k) = \liminf_{k\to\infty} \int_a^b \Psi(\mathrm{d}u_k).$$

Finally, by the continuity of $(\partial_t E)(t,\cdot)$ and Lebesgue's dominated convergence theorem,

$$\lim_{k\to\infty} \int_a^b (\partial_t E)(t,u_k(t))\,\mathrm{d}t = \int_a^b (\partial_t E)(t,u(t))\,\mathrm{d}t.$$

Thus, as claimed,

$$\mathrm{ES}(u,[a,b]) \le \liminf_{k\to\infty} \mathrm{ES}(u_k,[a,b]). \qquad \blacksquare$$

This simple lower semicontinuity result is crucial to the proof of theorem 4.3.1, since it ensures that if (up to a choice of subsequence) $z_\varepsilon \to z_0$ uniformly and the z_ε "do not violate ($\mathrm{E}_{\mathrm{glob}}$) too much" (in the sense that $\liminf_{\varepsilon\to 0} \mathrm{ES}(z_\varepsilon,[0,T]) \le 0$), then z_0 satisfies ($\mathrm{E}_{\mathrm{glob}}$).

Therefore, it is clear what the next few steps must be: to establish good control over $\mathrm{ES}(z_\varepsilon, [0,T])$! To do so, the evolution of z_ε inside and outside the dents is controlled separately: control outside the dents is provided by lemma 4.5.5, and control inside the dents is provided by lemma 4.5.6. The intuition is that z_ε is well-approximated by a sequence of straight-line paths from one dent to the next.

Lemma 4.5.5 (Control on variation outside dents)**.** *Suppose that $[a,b] \subseteq [0,T]$ is such that $z_\varepsilon(t)$ lies outside the support of the dents for all $t \in [a,b]$. Then*

$$\left|\mathrm{Var}_{[a,b]}(z_\varepsilon) - |z_\varepsilon(b) - z_\varepsilon(a)|\right| \leq \left(\frac{|b-a|}{\kappa} + \frac{|b-a|^2}{\varepsilon}\right) L.$$

Proof. The proof is mostly straightforward calculation. The idea is to compare z_ε to the straight-line evolution ζ_ε that is generated by fixing $\ell(t)$ to equal $\ell(a)$.

Since, by assumption, the evolution takes place outside the dent set, z_ε satisfies the ordinary differential equation

$$\dot{z}_\varepsilon(t) = \frac{\ell(t) - \kappa z_\varepsilon(t)}{\varepsilon}.$$

For comparison purposes, let ζ_ε solve

$$\dot{\zeta}_\varepsilon(t) = \frac{\ell(a) - \kappa \zeta_\varepsilon(t)}{\varepsilon} \tag{4.5.2}$$

with $z_\varepsilon(a) = \zeta_\varepsilon(a)$. Since the orbit of ζ_ε is a straight line, $\mathrm{Var}_{[a,b]}(\zeta_\varepsilon) = |\zeta_\varepsilon(b) - \zeta_\varepsilon(a)|$. Let $\eta_\varepsilon := z_\varepsilon - \zeta_\varepsilon$, which satisfies the ordinary differential equation

$$\dot{\eta}_\varepsilon(t) = -\frac{\kappa}{\varepsilon}\eta_\varepsilon(t) + \frac{\ell(t) - \ell(a)}{\varepsilon}. \tag{4.5.3}$$

This can be solved exactly to yield

$$\eta_\varepsilon(t) = \exp(-\kappa|t-a|/\varepsilon)\frac{1}{\varepsilon}\int_a^t \exp(\kappa|s-a|/\varepsilon)(\ell(s) - \ell(a))\,\mathrm{d}s.$$

Now estimate $|\eta_\varepsilon|$ and $|\dot{\eta}_\varepsilon(t)|$:

$$\begin{aligned}
|\eta_\varepsilon(t)| &\leq \frac{\exp(-\kappa|t-a|/\varepsilon)L}{\varepsilon}\int_a^t |s-a|\exp(\kappa|s-a|/\varepsilon)\,\mathrm{d}s \\
&= \frac{\exp(-\kappa|t-a|/\varepsilon)L}{\varepsilon}\,\frac{\varepsilon^2\exp(\kappa|s-a|/\varepsilon)}{\kappa^2}\left(\frac{\kappa}{\varepsilon}|s-a|-1\right)\Big|_{s=a}^{t} \\
&= L\left(\frac{|t-a|}{\kappa}-\frac{\varepsilon}{\kappa^2}+\frac{\varepsilon}{\kappa^2}\exp(-\kappa|t-a|/\varepsilon)\right) \\
&= L\left(\frac{|t-a|}{\kappa}+\frac{\varepsilon}{\kappa^2}(\exp(-\kappa|t-a|/\varepsilon)-1)\right) \\
&\leq L\frac{|t-a|}{\kappa}.
\end{aligned}$$

Applying the just-obtained bound for $|\eta_\varepsilon(t)|$ to (4.5.3) yields

$$\begin{aligned}
|\dot{\eta}_\varepsilon(t)| &\leq \frac{|t-a|L}{\varepsilon}+\frac{|\ell(t)-\ell(a)|}{\varepsilon} \\
&\leq \frac{2|t-a|L}{\varepsilon}.
\end{aligned}$$

Therefore,

$$\mathrm{Var}_{[a,b]}(\eta_\varepsilon) \leq \frac{|b-a|^2L}{\varepsilon}.$$

The claim now follows from the inverse triangle inequality

$$\left|\mathrm{Var}_{[a,b]}(z_\varepsilon)-\mathrm{Var}_{[a,b]}(\zeta_\varepsilon)\right| \leq \mathrm{Var}_{[a,b]}(\eta_\varepsilon)$$

and the estimates for $\mathrm{Var}_{[a,b]}(\zeta_\varepsilon)$, $\mathrm{Var}_{[a,b]}(\eta_\varepsilon)$ and $|\eta_\varepsilon(b)|$. ■

Lemma 4.5.6 (Control on variation within dents)**.** *Suppose that* $[a,b] \subseteq [0,T]$ *and that* $c \in \mathcal{O}_\varepsilon$ *are such that* $z_\varepsilon(t) \in \overline{\mathbb{B}_\varepsilon(c)} \subset\subset \mathcal{S}(t)$ *for all* $t \in [a,b]$*. Then*

$$\mathrm{Var}_{[a,b]}(z_\varepsilon) \leq \frac{2\varepsilon(\kappa+\sigma)}{\sigma}+\frac{\varepsilon(\kappa+3\sigma)}{\sigma^2}|b-a|L.$$

A fortiori*, the variation is of order* ε*: there is a constant* $C \geq 0$ *depending only on* κ*,* σ*,* ℓ *and* T *such that* $\mathrm{Var}_{[a,b]}(z_\varepsilon) \leq C\varepsilon$*.*

Proof. The proof of this lemma is an exercise in heroic but ultimately straightforward calculation. The idea is to control the variation of z_ε from its velocity nullcline, and also the variation of that nullcline itself.

Since, by assumption, z_ε remains within the dent centred at c for the whole of the time interval of interest, z_ε satisfies the ordinary differential

equation

$$\dot{z}_\varepsilon(t) = -\frac{1}{\varepsilon}\left(\kappa z_\varepsilon(t) + \frac{\sigma}{\varepsilon}\big(z_\varepsilon(t) - c\big) - \ell(t)\right).$$

Rearranging this into the standard form of a linear, inhomogeneous, first-order ordinary differential equation yields

$$\dot{z}_\varepsilon(t) + \frac{\varepsilon\kappa + \sigma}{\varepsilon^2} z_\varepsilon(t) = \frac{\varepsilon\ell(t) + \sigma c}{\varepsilon^2}.$$

Now let ζ_ε be the fixed point/velocity nullcline of the equation for z_ε. Setting $\dot{z}_\varepsilon(t) = 0$ in the previous equation and algebraically solving for $z_\varepsilon(t) = \zeta_\varepsilon(t)$ yields

$$\zeta_\varepsilon(t) = \frac{\varepsilon\ell(t) + \sigma c}{\varepsilon\kappa + \sigma}.$$

Obviously, ζ_ε satisfies the ordinary differential equation

$$\dot{\zeta}_\varepsilon(t) = \frac{\varepsilon\dot{\ell}(t)}{\varepsilon\kappa + \sigma}. \tag{4.5.4}$$

Integration of the Euclidean norm of (4.5.4) yields

$$\begin{aligned}
\mathrm{Var}_{[a,b]}(\zeta_\varepsilon) &= \int_a^b \big|\dot{\zeta}_\varepsilon(t)\big|\,\mathrm{d}t \\
&= \int_a^b \frac{\varepsilon}{\varepsilon\kappa + \sigma}\big|\dot{\ell}(t)\big|\,\mathrm{d}t \\
&\leq \frac{\varepsilon}{\sigma}\mathrm{Var}_{[a,b]}(\ell) \\
&\leq \frac{\varepsilon}{\sigma}|b-a|L.
\end{aligned}$$

The next step is to show that z_ε and ζ_ε are close in the variation seminorm, since this will then provide the requisite control on z_ε itself. Let $\eta_\varepsilon := z_\varepsilon - \zeta_\varepsilon$.

Then

$$\begin{aligned}
\eta_\varepsilon(t) &= z_\varepsilon(t) - \zeta_\varepsilon(t) \\
&= z_\varepsilon(t) - \frac{\varepsilon\ell(t) + \sigma c}{\varepsilon\kappa + \sigma} \\
&= \frac{(\varepsilon\kappa + \sigma)z_\varepsilon(t) - (\varepsilon\ell(t) + \sigma c)}{\varepsilon\kappa + \sigma} \\
&= \frac{\varepsilon(\kappa z_\varepsilon(t) - \ell(t)) + \sigma(z_\varepsilon(t) - c)}{\varepsilon\kappa + \sigma} \\
&= \frac{\varepsilon^2}{\varepsilon\kappa + \sigma}\left(\frac{\kappa z_\varepsilon(t) - \ell(t)}{\varepsilon} + \frac{\sigma(z_\varepsilon(t) - c)}{\varepsilon^2}\right).
\end{aligned}$$

Therefore,

$$\begin{aligned}
\dot\eta_\varepsilon(t) &= \dot z_\varepsilon(t) - \dot\zeta_\varepsilon(t) \\
&= -\frac{1}{\varepsilon}\left(\kappa z_\varepsilon(t) + \frac{\sigma}{\varepsilon}\big(z_\varepsilon(t) - c\big) - \ell(t)\right) - \frac{\varepsilon\dot\ell(t)}{\varepsilon\kappa + \sigma} \\
&= -\frac{\kappa z_\varepsilon(t) - \ell(t)}{\varepsilon} - \frac{\sigma}{\varepsilon^2}(z_\varepsilon(t) - c) - \frac{\varepsilon\dot\ell(t)}{\varepsilon\kappa + \sigma} \\
&= -\frac{\varepsilon\kappa + \sigma}{\varepsilon^2}\big(z_\varepsilon(t) - \zeta_\varepsilon(t)\big) - \frac{\varepsilon\dot\ell(t)}{\varepsilon\kappa + \sigma} \\
&= -\frac{\varepsilon\kappa + \sigma}{\varepsilon^2}\eta_\varepsilon(t) - \frac{\varepsilon\dot\ell(t)}{\varepsilon\kappa + \sigma}.
\end{aligned}$$

That is, η_ε satisfies the ordinary differential equation

$$\dot\eta_\varepsilon(t) = -\frac{\varepsilon\kappa + \sigma}{\varepsilon^2}\eta_\varepsilon(t) - \frac{\varepsilon\dot\ell(t)}{\varepsilon\kappa + \sigma}. \tag{4.5.5}$$

(4.5.5) can be solved exactly to yield

$$\eta_\varepsilon(t) = \exp\big(-\tfrac{\varepsilon\kappa+\sigma}{\varepsilon^2}|t-a|\big)\left(\eta_\varepsilon(a) - \frac{\varepsilon}{\varepsilon\kappa + \sigma}\int_a^t \exp\big(\tfrac{\varepsilon\kappa+\sigma}{\varepsilon^2}|s-a|\big)\dot\ell(s)\,\mathrm{d}s\right).$$

A straightforward application of the triangle inequality yields

$$\begin{aligned}
&|\eta_\varepsilon(t)| \\
&\quad \le \exp\big(-\tfrac{\varepsilon\kappa+\sigma}{\varepsilon^2}|t-a|\big)\left(|\eta_\varepsilon(a)| + \frac{\varepsilon L}{\varepsilon\kappa+\sigma}\int_a^t \exp\big(+\tfrac{\varepsilon\kappa+\sigma}{\varepsilon^2}|s-a|\big)\,\mathrm{d}s\right) \\
&\quad = \exp\big(-\tfrac{\varepsilon\kappa+\sigma}{\varepsilon^2}|t-a|\big)|\eta_\varepsilon(a)| + \frac{\varepsilon^3 L}{(\varepsilon\kappa+\sigma)^2}\left(1-\exp\big(-\tfrac{\varepsilon\kappa+\sigma}{\varepsilon^2}|t-a|\big)\right) \\
&\quad \le \exp\big(-\tfrac{\varepsilon\kappa+\sigma}{\varepsilon^2}|t-a|\big)|\eta_\varepsilon(a)| + \frac{\varepsilon^3}{\sigma^2}L.
\end{aligned}$$

Applying the just-obtained bound for $|\eta_\varepsilon(t)|$ to (4.5.5) yields

$$\begin{aligned}
\big|\dot\eta_\varepsilon(t)\big| &\le \frac{\varepsilon\kappa+\sigma}{\varepsilon^2}|\eta_\varepsilon(t)| + \frac{\varepsilon}{\varepsilon\kappa+\sigma}\big|\dot\ell(t)\big| \\
&\le \frac{\kappa+\sigma}{\varepsilon^2}|\eta_\varepsilon(t)| + \frac{\varepsilon}{\sigma}\big|\dot\ell(t)\big| \\
&\le \frac{\kappa+\sigma}{\varepsilon^2}\left(\exp\big(-\tfrac{\varepsilon\kappa+\sigma}{\varepsilon^2}|t-a|\big)|\eta_\varepsilon(a)| + \frac{\varepsilon^3}{\sigma^2}L\right) + \frac{\varepsilon}{\sigma}\big|\dot\ell(t)\big| \\
&\le \frac{\kappa+\sigma}{\varepsilon^2}\exp\big(-\tfrac{\varepsilon\kappa+\sigma}{\varepsilon^2}|t-a|\big)|\eta_\varepsilon(a)| + \frac{\varepsilon(\kappa+2\sigma)}{\sigma^2}L.
\end{aligned}$$

Integration of this bound for $\big|\dot\eta_\varepsilon(t)\big|$ yields

$$\begin{aligned}
&\mathrm{Var}_{[a,b]}(\eta_\varepsilon) \\
&\quad = \int_a^b \big|\dot\eta_\varepsilon(t)\big|\,\mathrm{d}t \\
&\quad \le \int_a^b \left(\frac{\kappa+\sigma}{\varepsilon^2}\exp\big(-\tfrac{\varepsilon\kappa+\sigma}{\varepsilon^2}|t-a|\big)|\eta_\varepsilon(a)| + \frac{\varepsilon(\kappa+2\sigma)}{\sigma^2}L\right)\mathrm{d}t \\
&\quad = \frac{\kappa+\sigma}{\varepsilon^2}|\eta_\varepsilon(a)|\int_a^b \exp\big(-\tfrac{\varepsilon\kappa+\sigma}{\varepsilon^2}|t-a|\big)\,\mathrm{d}t + \frac{\varepsilon(\kappa+2\sigma)}{\sigma^2}|b-a|L \\
&\quad = \frac{\kappa+\sigma}{\varepsilon\kappa+\sigma}|\eta_\varepsilon(a)|\left(1-\exp\big(-\tfrac{\varepsilon\kappa+\sigma}{\varepsilon^2}|b-a|\big)\right) + \frac{\varepsilon(\kappa+2\sigma)}{\sigma^2}|b-a|L \\
&\quad \le \frac{\kappa+\sigma}{\sigma}|\eta_\varepsilon(a)| + \frac{\varepsilon(\kappa+2\sigma)}{\sigma^2}|b-a|L.
\end{aligned}$$

Finally, putting together the estimates for the variations of ζ_ε and η_ε and

using the triangle inequality for the variation seminorm yields

$$\begin{aligned}
&\mathrm{Var}_{[a,b]}(z_\varepsilon) \\
&\quad \le \mathrm{Var}_{[a,b]}(\zeta_\varepsilon) + \mathrm{Var}_{[a,b]}(\eta_\varepsilon) \\
&\quad \le \frac{\varepsilon}{\sigma}|b-a|L + \frac{\kappa+\sigma}{\sigma}|\eta_\varepsilon(a)| + \frac{\varepsilon(\kappa+2\sigma)}{\sigma^2}|b-a|L \\
&\quad = \frac{\kappa+\sigma}{\sigma}|\eta_\varepsilon(a)| + \frac{\varepsilon(\kappa+3\sigma)}{\sigma^2}|b-a|L,
\end{aligned}$$

and since $|\eta_\varepsilon(a)| \le 2\varepsilon$, the claim follows. ■

Putting together lemma 4.5.5 and lemma 4.5.6 yields the following estimate for the variation of z_ε between leaving one dent and leaving the next:

$$\begin{aligned}
&\left|\mathrm{Var}_{[\tau_i^{\mathrm{out}},\tau_{i+1}^{\mathrm{out}}]}(z_\varepsilon) - |z_\varepsilon(\tau_{i+1}^{\mathrm{in}}) - z_\varepsilon(\tau_i^{\mathrm{out}})|\right| \\
&\quad \overset{O(\varepsilon)}{\le} \left(\frac{|\tau_{i+1}^{\mathrm{in}} - \tau_i^{\mathrm{out}}|}{\kappa} + \frac{|\tau_{i+1}^{\mathrm{in}} - \tau_i^{\mathrm{out}}|^2}{\varepsilon}\right) L.
\end{aligned}$$

Under the natural assumption that the inter-dent distances tend to zero as $\varepsilon \to 0$, for small enough ε,

$$|\tau_{i+1}^{\mathrm{in}} - \tau_i^{\mathrm{out}}| \le \frac{\varepsilon}{2\sigma}|z_\varepsilon(\tau_{i+1}^{\mathrm{in}}) - z_\varepsilon(\tau_i^{\mathrm{out}})|.$$

Since z_ε certainly cannot perform a rapid descent over a distance greater than the diameter of the stable region, it follows that

$$\mathrm{Var}_{[\tau_i^{\mathrm{out}},\tau_{i+1}^{\mathrm{out}}]}(z_\varepsilon) \overset{O(\varepsilon)}{=} |z_\varepsilon(\tau_{i+1}^{\mathrm{in}}) - z_\varepsilon(\tau_i^{\mathrm{out}})| \overset{O(\varepsilon)}{=} |z_\varepsilon(\tau_{i+1}^{\mathrm{out}}) - z_\varepsilon(\tau_i^{\mathrm{out}})|.$$

The next step is to turn this control of the variation of z_ε into an estimate of how badly the global energy equality may fail along z_ε.

Lemma 4.5.7 (Bound on the energy surplus)**.** *There exists a constant $C \ge 0$ such that the energy surplus satisfies*

$$\mathrm{ES}(z_\varepsilon, [\tau_i^{\mathrm{out}}, \tau_{i+1}^{\mathrm{out}}]) \le C\varepsilon + \frac{\sigma|\tau_{i+1}^{\mathrm{in}} - \tau_i^{\mathrm{out}}|^2}{\varepsilon} L.$$

That is, the energy surplus is of order ε plus $\frac{1}{\varepsilon}$ times the square of the duration of the rapid descent.

Proof. As usual, the proof is mostly straightforward calculation, but two

points are noteworthy. The first is the trick of adding and subtracting $E(\tau_i^{\text{out}}, z_\varepsilon(\tau_{i+1}^{\text{in}}))$ to the left-hand side of the global energy inequality: this trick sets up the cancellation effects that are needed in order to simplify everything. The second is the *dénouement* of the proof: the cancellation that completes the proof uses the fact that the energy gradient at the frontier of the stable region is inward-pointing with magnitude σ, and so the energy loss made during the rapid descent phase is almost exactly equal to the dissipation.

Consider the sought-for global energy inequality:

$$E(s, z_\varepsilon(s))\big|_{s=\tau_i^{\text{out}}}^{\tau_{i+1}^{\text{out}}} \leq \int_{\tau_i^{\text{out}}}^{\tau_{i+1}^{\text{out}}} (\partial_t E)(s, z_\varepsilon(s))\,\mathrm{d}s - \sigma \mathrm{Var}_{[\tau_i^{\text{out}}, \tau_{i+1}^{\text{out}}]}(z_\varepsilon). \quad (4.5.6)$$

Manipulation of the left-hand side of (4.5.6) yields

$$\begin{aligned}
E(s, z_\varepsilon(s))\big|_{s=\tau_i^{\text{out}}}^{\tau_{i+1}^{\text{out}}} &= E(\tau_{i+1}^{\text{out}}, z_\varepsilon(\tau_{i+1}^{\text{out}})) - E(\tau_i^{\text{out}}, z_\varepsilon(\tau_{i+1}^{\text{in}})) \\
&\quad + E(\tau_i^{\text{out}}, z_\varepsilon(\tau_{i+1}^{\text{in}})) - E(\tau_i^{\text{out}}, z_\varepsilon(\tau_i^{\text{out}})) \\
&\overset{O(\varepsilon)}{=} E(\tau_{i+1}^{\text{out}}, z_\varepsilon(\tau_{i+1}^{\text{out}})) - E(\tau_i^{\text{out}}, z_\varepsilon(\tau_{i+1}^{\text{out}})) \\
&\quad + E(\tau_i^{\text{out}}, z_\varepsilon(\tau_{i+1}^{\text{in}})) - E(\tau_i^{\text{out}}, z_\varepsilon(\tau_i^{\text{out}})) \\
&= \int_{\tau_i^{\text{out}}}^{\tau_{i+1}^{\text{out}}} (\partial_t E)(s, z_\varepsilon(\tau_{i+1}^{\text{out}}))\,\mathrm{d}s + E(\tau_i^{\text{out}}, z_\varepsilon(s))\big|_{s=\tau_i^{\text{out}}}^{\tau_{i+1}^{\text{in}}} \\
&= \big\langle -\ell(s), z_\varepsilon(\tau_{i+1}^{\text{out}})\big\rangle\big|_{s=\tau_i^{\text{out}}}^{\tau_{i+1}^{\text{out}}} + E(\tau_i^{\text{out}}, z_\varepsilon(s))\big|_{s=\tau_i^{\text{out}}}^{\tau_{i+1}^{\text{in}}}
\end{aligned}$$

Manipulation of the right-hand side of (4.5.6) yields

$$\begin{aligned}
&\int_{\tau_i^{\text{out}}}^{\tau_{i+1}^{\text{out}}} (\partial_t E)(s, z_\varepsilon(s))\,\mathrm{d}s - \sigma \mathrm{Var}_{[\tau_i^{\text{out}}, \tau_{i+1}^{\text{out}}]}(z_\varepsilon) \\
&\quad \overset{O(\varepsilon)}{=} \int_{\tau_i^{\text{out}}}^{\tau_{i+1}^{\text{out}}} \big\langle -\dot{\ell}(s), z_\varepsilon(s))\big\rangle\,\mathrm{d}s - \sigma\big|z_\varepsilon(\tau_{i+1}^{\text{out}}) - z_\varepsilon(\tau_i^{\text{out}})\big| \\
&\quad = \int_{\tau_i^{\text{out}}}^{\tau_{i+1}^{\text{out}}} \big\langle \ell(s), \dot{z}_\varepsilon(s))\big\rangle\,\mathrm{d}s - \big\langle \ell(s), z_\varepsilon(s)\big\rangle\big|_{s=\tau_i^{\text{out}}}^{\tau_{i+1}^{\text{out}}} - \sigma\big|z_\varepsilon(\tau_{i+1}^{\text{out}}) - z_\varepsilon(\tau_i^{\text{out}})\big| \\
&\quad \overset{O(\varepsilon)}{=} \int_{\tau_i^{\text{out}}}^{\tau_{i+1}^{\text{in}}} \big\langle \ell(s), \dot{z}_\varepsilon(s))\big\rangle\,\mathrm{d}s - \big\langle \ell(s), z_\varepsilon(s)\big\rangle\big|_{s=\tau_i^{\text{out}}}^{\tau_{i+1}^{\text{out}}} - \sigma\big|z_\varepsilon(\tau_{i+1}^{\text{out}}) - z_\varepsilon(\tau_i^{\text{out}})\big|
\end{aligned}$$

Subtracting the results of these two manipulations yields that, up to order

ε, the difference between the left- and right-hand sides of (4.5.6) (i.e. the energy surplus for z_ε over $[\tau_i^{\text{out}}, \tau_{i+1}^{\text{out}}]$ is

$$\begin{aligned}
&\big\langle \ell(s), z_\varepsilon(s) - z_\varepsilon(\tau_{i+1}^{\text{out}})\big\rangle\big|_{s=\tau_i^{\text{out}}}^{\tau_{i+1}^{\text{out}}} + E(\tau_i^{\text{out}}, z_\varepsilon(s))\big|_{s=\tau_i^{\text{out}}}^{\tau_{i+1}^{\text{in}}} \\
&\qquad - \int_{\tau_i^{\text{out}}}^{\tau_{i+1}^{\text{in}}} \big\langle \ell(s), \dot z_\varepsilon(s))\big\rangle \,\mathrm{d}s + \sigma\big|z_\varepsilon(\tau_{i+1}^{\text{out}}) - z_\varepsilon(\tau_i^{\text{out}})\big| \\
&= \big\langle \ell(\tau_i^{\text{out}}), z_\varepsilon(\tau_i^{\text{out}}) - z_\varepsilon(\tau_{i+1}^{\text{out}})\big\rangle + E(\tau_i^{\text{out}}, z_\varepsilon(s))\big|_{s=\tau_i^{\text{out}}}^{\tau_{i+1}^{\text{in}}} \\
&\qquad - \int_{\tau_i^{\text{out}}}^{\tau_{i+1}^{\text{in}}} \big\langle \ell(s), \dot z_\varepsilon(s))\big\rangle \,\mathrm{d}s + \sigma\big|z_\varepsilon(\tau_{i+1}^{\text{out}}) - z_\varepsilon(\tau_i^{\text{out}})\big| \\
&\overset{O(\varepsilon)}{=} E(\tau_i^{\text{out}}, z_\varepsilon(s))\big|_{s=\tau_i^{\text{out}}}^{\tau_{i+1}^{\text{in}}} - \int_{\tau_i^{\text{out}}}^{\tau_{i+1}^{\text{in}}} \big\langle \ell(s) - \ell(\tau_i^{\text{out}}), \dot z_\varepsilon(s))\big\rangle \,\mathrm{d}s \\
&\qquad + \sigma\big|z_\varepsilon(\tau_{i+1}^{\text{in}}) - z_\varepsilon(\tau_i^{\text{out}})\big| \\
&\le -\frac{\kappa}{2}\big|z_\varepsilon(\tau_{i+1}^{\text{in}}) - z_\varepsilon(\tau_i^{\text{out}})\big|^2 + \frac{\sigma|\tau_{i+1}^{\text{in}} - \tau_i^{\text{out}}|^2}{\varepsilon} L \\
&\le \frac{\sigma|\tau_{i+1}^{\text{in}} - \tau_i^{\text{out}}|^2}{\varepsilon} L.
\end{aligned}$$

This completes the proof. ∎

As before, under natural assumptions on the inter-dent distances, lemma 4.5.7 becomes a true $O(\varepsilon)$ estimate:

$$\mathrm{ES}(z_\varepsilon, [\tau_i^{\text{out}}, \tau_{i+1}^{\text{out}}]) \overset{O(\varepsilon)}{=} 0.$$

Thus, modulo some error terms, z_ε satisfies (E_{glob}) on each $[\tau_i^{\text{out}}, \tau_{i+1}^{\text{out}}]$. The next question is this: how many of these random intervals $[\tau_i^{\text{out}}, \tau_{i+1}^{\text{out}}]$ there are in $[0, T]$? If there are too many, then the error terms may accumulate and (E_{glob}) may fail on $[0, T]$. Note that this is the first serious probabilistic question to have been asked in this analysis, and it is the probabilistic distribution of the dents and the "rapidly descend then wait in a dent" character of the evolution of z_ε that will furnish the required control.

Recall that z_ε can only leave a dent if the energetic potential E changes in such a way as for that dent to become unstable. Therefore, when z_ε leaves a dent at time τ_i^{out} and then enters the next dent at time τ_{i+1}^{in}, and in doing so travels a distance $\big|z_\varepsilon(\tau_{i+1}^{\text{in}}) - z_\varepsilon(\tau_i^{\text{out}})\big|$, it follows that the random time τ_{i+1}^{out} cannot occur until the frontier of the stable region has "caught up"

with z_ε, and passed through the $(i+1)^{\text{th}}$ dent. How long a duration of time is that? Since the frontier of the stable region moves with speed $\left|\dot{\ell}\right|/\kappa$, it follows that

$$|\tau_{i+1}^{\text{out}} - \tau_i^{\text{out}}| \geq |\tau_{i+1}^{\text{out}} - \tau_{i+1}^{\text{in}}| \geq \frac{\kappa}{L}\left|z_\varepsilon(\tau_{i+1}^{\text{in}}) - z_\varepsilon(\tau_i^{\text{out}})\right|.$$

This is almost enough to finish the proof, but for one nasty detail: it may happen that dents overlap. Recall that the solution convention is that if dents do overlap, then z_ε sticks on the piecewise-hyperplanar interface of the overlapping dents until freed by the passage of the stable region. When it is freed, it immediately enters another dent, leading to the degenerate situation $\tau_{i+1}^{\text{in}} = \tau_i^{\text{out}}$. However, since the Poisson process is dilute in the limit as $\varepsilon \to 0$, it is reasonable to suppose that this situation does not happen too often.

Indeed, the distance between dent centres is exponentially distributed with mean $c_n\varepsilon^{n-p}$. Therefore, the event that z_ε leaves one dent and then immediately enters the next is a Bernoulli random variable with occurrence probability $1 - \exp(-c_n\varepsilon^{n-p}) \approx 0$. It is possible that z_ε will go through a cluster of such overlapping dents before a non-trivial "free-fall" occurs. Hence, the number of overlapping dents encountered before free-fall is a geometric random variable with parameter $1 - \exp(-c_n\varepsilon^{n-p})$. Therefore, if there were M non-trivial free-falls during $[0, T]$, the number N of overlapping dents encountered would be distributed according to a negative binomial distribution with parameters M and $1 - \exp(-c_n\varepsilon^{n-p})$: it has mean

$$\mathbb{E}[N \mid M] = M\frac{1 - \exp(-c_n\varepsilon^{n-p})}{\exp(-c_n\varepsilon^{n-p})}$$

and variance

$$\text{Var}[N \mid M] = M\frac{1 - \exp(-c_n\varepsilon^{n-p})}{\exp(-2c_n\varepsilon^{n-p})},$$

and so, conditional on M, N converges to 0 in mean square.

Since the distance between dent centres is exponentially distributed with mean $c_n\varepsilon^{n-p}$, the number M of time intervals $[\tau_i^{\text{out}}, \tau_{i+1}^{\text{out}}] \subseteq [0, T]$ that involve non-trivial free-fall is bounded above by a Poisson random variable with mean and variance $c_n^{-1}\varepsilon^{p-n}T$. Recall, though, that whether or not $[\tau_i^{\text{out}}, \tau_{i+1}^{\text{out}}]$ has non-trivial free-fall, the energy surplus over $[\tau_i^{\text{out}}, \tau_{i+1}^{\text{out}}]$ is still of order ε.

The arguments of the last two paragraphs show that the contribution to

$\mathrm{ES}(z_\varepsilon, [0,T])$ coming from the degenerate free falls has mean at most

$$\begin{aligned}\mathbb{E}[C\varepsilon N] &= C\varepsilon \mathbb{E}\left[\mathbb{E}[N \mid M]\right] \\ &= C\varepsilon \frac{T}{c_n \varepsilon^{n-p}} \frac{1-\exp(-c_n\varepsilon^{n-p})}{\exp(-c_n\varepsilon^{n-p})} \\ &\leq C'T\varepsilon\end{aligned}$$

and variance at most

$$\begin{aligned}\mathrm{Var}[C\varepsilon N] &= (C\varepsilon)^2 \left(\mathbb{E}\left[\mathrm{Var}[N \mid M]\right] + \mathrm{Var}\left[\mathbb{E}[N \mid M]\right]\right) \\ &\leq C''T\varepsilon^2\end{aligned}$$

for all sufficiently small ε.

Putting all the arguments of this subsection together yields the following result. Heuristically, there are $O(\varepsilon^{p-n})$ "good" subintervals and $O(T)$ "bad" subintervals, each contributing $O(\varepsilon)$ to the energy surplus; the net result is an energy surplus of order ε^{p-n+1}, which goes to zero as $\varepsilon \to 0$ since $p \in (n-1, n)$.

Theorem 4.5.8. *The energy surplus over $[0,T]$ converges to zero in mean square. More precisely,*

$$\mathbb{E}\left[\mathrm{ES}(z_\varepsilon, [0,T])\right] \leq C\frac{T}{c_n}\varepsilon^{p-n+1}$$

and

$$\mathrm{Var}\left[\mathrm{ES}(z_\varepsilon, [0,T])\right] \leq C\frac{T}{c_n}\varepsilon^{p-n+2}.$$

As a corollary, the variation of z_ε over any subinterval $\mathcal{T}_\delta$ of $[0,T]$ of diameter δ satisfies

$$\lim_{\delta\to 0} \limsup_{\varepsilon\to 0} \mathrm{Var}_{\mathcal{T}_\delta}(z_\varepsilon) = 0 \text{ in } L^2(\Omega, \mathscr{F}, \mathbb{P}; \mathbb{R}),$$

and, by the Bienayme–Chebyshëv inequality, the same convergence holds in the L^0 topology (the topology of convergence in probability $\mathbb{P}$). This observation greatly simplifies the tightness proof of the next subsection.

4.5.3 Tightness

In order to establish that the family of continuous processes $(z_\varepsilon)_{\varepsilon>0}$ has a subsequence that converges in distribution, it suffices to show that the family is uniformly tight — this is the content of Prokhorov's theorem (theorem D.4). Coupled with the Arzelà–Ascoli theorem, it is enough to establish the following lemma:

Lemma 4.5.9. *Let z_ε solve* (4.2.3) *and let ϖ_{z_ε} denote the (random) modulus of continuity for z_ε. Then*

$$\text{for all } \theta > 0, \lim_{\delta\to 0} \limsup_{\varepsilon\to 0} \mathbb{P}\left[\varpi_{z_\varepsilon}(\delta) > \theta\right] = 0. \tag{4.5.7}$$

Hence, the laws of z_ε, $\varepsilon > 0$, are uniformly tight in $\mathcal{C}^0([0,T];\mathbb{R}^n)$.

Proof. Note that the modulus of continuity and variation obviously satisfy, for any function $u\colon [0,T] \to \mathbb{R}^n$ and any $0 < \delta \leq T$,

$$\varpi_u(\delta) \leq \sup\left\{\operatorname{Var}_{[a,\delta]}(u) \,\middle|\, [a, a+\delta] \subseteq [0,T]\right\} \leq \operatorname{Var}_{[0,T]}(u).$$

Hence having established good estimates of the variation of z_ε earlier, there is no need to re-invent the wheel in establishing (4.5.7): most of the hard work has already been done. Indeed, the claim follows immediately from theorem 4.5.8 and the Bienayme–Chebyshëv inequality. ■

Chapter 5

Thermalized Gradient Descent I: Generalities

5.1 Introductory Remarks

This chapter introduces a family of methods by which a (deterministic) gradient descent can be "thermalized", i.e. placed in contact with a heat bath. The idea is to perturb the Moreau–Yosida incremental formulation of the gradient descent using an interior-point regularization in such a way as to force the probability density function of the system to "spread out" in a controlled way.

In [Kos03], the finite-dimensional Euclidean case was considered, and the interior-point regularization was generated by the negative of the Gibbs entropy functional: this is essentially the principle of maximum entropy that dates back to Jaynes [Jay57a] [Jay57b]. However, such regularization problems can be posed in a much more general setting: the general formulation makes sense on any pair of Polish spaces (the natural setting for gradient descent problems) and can be phrased with respect to any f-divergence of probability measures.

In [SKTO09], the method of this chapter is referred to an an "optimal transport" procedure, albeit with a warning that it should not be confused with optimal transport in the sense of the Monge–Kantorovich problem [Mon81] [Kan42]. In this thesis, the descriptive nomenclature "thermalized gradient descent" is preferred, thus eliminating any potentially ambiguous abuse of terminology.

The presentation of the motivation and heuristics for thermalized gradient descents in this chapter owes much to the corresponding section in [SKTO09] and, in turn, to [Kos03].

5.2 Motivation and Heuristic Derivation

The heuristic motivation for thermalized gradient descents of the type used in this thesis is that they constitute a "probabilistic regularization" of the Euler–Lagrange equations for a (deterministic) dissipative system at absolute zero temperature, and thus form plausible models for dissipative systems at positive temperature. That is, they are tools used to answer the question

> What happens when a dissipative system is placed in contact with a heat bath?

As an introductory toy example, consider a gradient descent evolution $x\colon [0,T] \to \mathbb{R}^n$ governed by an energetic potential $E\colon [0,T] \times \mathbb{R}^n \to \mathbb{R}$ and a dissipative potential $\Psi\colon \mathbb{R}^n \to \mathbb{R}$. A *conservative* system is one in which Ψ is the zero function, and the state $x(t)$ at time $t \in (0,T]$ is independent of the initial condition $x(0)$ and is, in fact, given by energy minimization: that is, for each $t \in (0,T]$,

$$x(t) \in \operatorname*{arg\,min}_{y \in \mathbb{R}^n} E(t,y).$$

If the system is a *bona fide* dissipative system, then its evolution is described by the equilibrium equation/differential inclusion

$$\partial \Psi(\dot{x}(t)) + \partial E(t, x(t)) \ni 0. \tag{5.2.1}$$

The standard method for attempting to solve this problem is to resort to time discretization and the Moreau–Yosida scheme (2.3.3) for general gradient descents outlined in chapter 2. That is, given a partition

$$P = \{0 = t_0 < t_1 < \ldots < t_N = T\}$$

of $[0,T]$, the aim is to solve the causal sequence of minimization problems

$$x_{i+1} \in \operatorname*{arg\,min}_{y \in \mathbb{R}^n} \mathscr{W}_{i+1}(x_i, y), \tag{5.2.2}$$

where the *incremental work function* $\mathscr{W}_{i+1}$ for the time step from t_i to t_{i+1} is given by

$$\mathscr{W}_{i+1}(x_i, x_{i+1}) := E(t_{i+1}, x_{i+1}) - E(t_i, x_i) + \Delta t_{i+1} \Psi \left(\frac{\Delta x_{i+1}}{\Delta t_{i+1}} \right). \quad (5.2.3)$$

(The subtraction of $E(t_i, x_i)$ has no effect on the minimization problem since it is a constant.) The incremental work function $\mathscr{W}_{i+1}$ has the property that it is a potential for the net forces at time t_{i+1}; hence, the equilibrium equation (5.2.1) is the Euler–Lagrange equation for the minimization problem (5.2.2).

From an heuristic point of view, the Moreau–Yosida incremental problem is a method for "optimally transporting" the Dirac measure δ_{x_i} that describes the state of the system at time t_i to a new measure $\delta_{x_{i+1}}$. An over-complicated way of phrasing the minimization problem (5.2.2) would have been, given δ_{x_i}, to seek a minimizer among measures μ on $\mathbb{R}^n \times \mathbb{R}^n$ with first marginal δ_{x_i} for

$$\iint_{\mathbb{R}^n} \mathscr{W}_{i+1} \, \mathrm{d}\mu;$$

note that $\delta_{x_{i+1}}$ would be the second marginal of the minimizer.

To model the effect of a heat bath on (5.2.1), introduce an *interior-point regularization* that forces the probability density function of the system to "spread out": given the probability density function ρ_i at time t_i, find a joint probability density $\rho_{i,i+1}$ to minimize

$$\iint_{\mathbb{R}^n} \mathscr{W} \, \rho_{i,i+1} + \varepsilon \rho_{i,i+1} \log \rho_{i,i+1}. \quad (5.2.4)$$

The first term in the integrand is the familiar one, which is minimized by a deterministic transport of each state x_i to its successor x_{i+1} in the Moreau–Yosida scheme. The second term in the integrand, which is the negative of the Gibbs–Boltzmann entropy functional, penalizes such coherent and deterministic transport. Heuristically, the second marginal ρ_{i+1} of $\rho_{i,i+1}$ (i.e. the probability density function for the next state of the system) is the original deterministic transport of ρ_i but "spread out a bit" by the heat bath. Setting physical temperature $\Theta = \varepsilon / k_{\mathrm{B}}$ in terms of Boltzmann's constant k_{B}, ε can be interpreted as the temperature of heat bath with which the system is coupled.

As will be shown in the next section (theorem 5.3.3), the minimizer for the interior-point regularization problem (5.2.4) is Gibbsian:

$$\rho_{i,i+1}(x_i, x_{i+1}) = \rho_i(x_i)\frac{\exp(-\mathscr{W}_{i+1}(x_i, x_{i+1})/\varepsilon)}{Z(x_i)} \tag{5.2.5}$$

where the "partition function" Z is given by

$$Z(x_i) := \int_{\mathbb{R}^n} \exp(-\mathscr{W}(x_i, x_{i+1})/\varepsilon)\,\mathrm{d}x_{i+1}.$$

The idea now is to form a Markov chain $X = X^{(P)} \colon \Omega \times P \to \mathbb{R}^n$ from this sequence of problems (5.2.4), positing that it models the effect of a heat bath on (5.2.1), and to take a suitable continuous-time limit as $[\![P]\!] \to 0$.

The Chapman–Kolmogorov equation (i.e. iteration of (5.2.5)) yields the following expression for the probability density function ρ_k of the system at time t_k:

$$\rho_k(x_k) = \int \cdots \int_{\mathbb{R}^n} \rho_0(x_0) \left(\prod_{i=0}^{k-1} \frac{1}{Z(x_i)}\right) \exp\left(-\frac{1}{\varepsilon}\sum_{i=0}^{k-1} \mathscr{W}_{i+1}(x_i, x_{i+1})\right) \mathrm{d}x_0 \ldots \mathrm{d}x_{k-1}. \tag{5.2.6}$$

One possible avenue of investigation, but not the one followed in this thesis, would be to regard the above iterated integral over $(\mathbb{R}^n)^P$ as an approximation of a functional integral over $(\mathbb{R}^n)^{[0,T]}$; see [Mul87] [Wie86] for a survey of these techniques. In this approach, define an action functional $\mathcal{I}$ on paths $x \colon [0,T] \to \mathbb{R}^n$ by

$$\mathcal{I}[x] := \lim \sum_{i=0}^{k-1} \mathscr{W}_{i+1}(x_i, x_{i+1}), \tag{5.2.7}$$

where the limit is taken as $[\![P]\!] \to 0$, $k \to \infty$, $t_k = T$, to yield a path integral of the form

$$\int \exp(-\mathcal{I}[x]/\varepsilon)\,\mathcal{D}x$$

and hence a "likelihood ratio" for two paths x and y of

$$\frac{e^{-\mathcal{I}[x]/\varepsilon}}{e^{-\mathcal{I}[y]/\varepsilon}} = e^{-(\mathcal{I}[x]-\mathcal{I}[y])/\varepsilon}.$$

In particular, the minimizer of $\mathcal{I}$, if it has one, should describe the limiting behaviour of the model as $[\![P]\!] \to 0$ and $\varepsilon \to 0$. Indeed, formally,

$$\mathcal{I}[x] = E(T, x(T)) - E(0, x(0)) + \int_0^T \Psi(\dot{x}(t)) \, \mathrm{d}t - \int_0^T (\partial_t E)(t, x(t)) \, \mathrm{d}t.$$

In the examples with 2-homogeneous dissipation considered in section 5.4, $\mathcal{I}$ is the rate function of the limiting diffusion in the sense of large deviations theory [DZ98]. In the case of 1-homogeneous dissipation in chapter 6, $\mathcal{I}$ is the energy surplus of chapter 4, which equals zero for the unthermalized rate-independent process when it is a continuous process [MT04, lemma 3.7]. However, the limiting continuous-time dynamics of the thermalized system do *not* minimize $\mathcal{I}$.

5.3 Interior-Point Regularization Problems

This section considers the interior-point regularization problem (5.2.4) in greater generality. The spaces involved need not be finite-dimensional, or even have linear structure: the requirement that they be complete and separable metric spaces is desirable, since under these hypotheses the collection of Radon probability measures on the space is again a Polish space with respect to the topology of weak convergence. In addition, in the single-step problem, the "source" and "target" spaces need not be the same space: this facilitates the definition of non-Markovian (higher order) thermalized gradient descents.

5.3.1 The Single-Step Problem

There are many measures of "distance" between probability measures on a given space. One class of such "distances" are the *f-divergences* introduced independently by [AS66] and [Csi67]. Recall that, given two measures μ and ν on the same measurable space $(\Omega, \mathscr{F})$, μ is said to be *absolutely continuous* with respect to ν if

$$\text{for } A \in \mathscr{F}, \nu(A) = 0 \implies \mu(A) = 0; \tag{5.3.1}$$

this relation will be denoted by $\mu \preceq \nu$. By the celebrated Radon–Nikodým theorem, for σ-finite measures, absolute continuity is equivalent to the exis-

tence of a *density*, a function $\frac{\mathrm{d}\mu}{\mathrm{d}\nu} \in L^1(\Omega, \mathscr{F}, \nu; [0, +\infty])$ such that

$$\text{for all } A \in \mathscr{F}, \mu(A) = \int_A \frac{\mathrm{d}\mu}{\mathrm{d}\nu}\,\mathrm{d}\nu. \tag{5.3.2}$$

Definition 5.3.1. Given a measurable space Ω, two probability measures μ and ν on Ω with $\mu \preceq \nu$, and a convex function $f\colon [0, +\infty) \to \mathbb{R}$ such that $f(1) = 0$, the *f-divergence of μ with respect to ν* is defined by

$$D_f(\mu\|\nu) := \int_\Omega f \circ \frac{\mathrm{d}\mu}{\mathrm{d}\nu}\,\mathrm{d}\nu. \tag{5.3.3}$$

A frequently-used f-divergence is the *Kullback–Leibler divergence*, which is the f-divergence given by $f_{\mathrm{KL}}(\rho) := \rho \log \rho$:

$$D_{\mathrm{KL}}(\mu\|\nu) := \int_\Omega \frac{\mathrm{d}\mu}{\mathrm{d}\nu}(\omega) \log \frac{\mathrm{d}\mu}{\mathrm{d}\nu}(\omega)\,\mathrm{d}\nu(\omega). \tag{5.3.4}$$

This is none other than the negative of the classical Gibbs–Boltzmann entropy functional for the density of μ with respect to ν. Although D_{KL} is sometimes referred to as a "distance" on the space of probability measures on Ω, this is a misnomer: although D_{KL} is positive-definite (a result known as Gibbs' inequality), in general it is neither symmetric nor does it satisfy the triangle inequality; furthermore, $D_{\mathrm{KL}}(\mu\|\nu)$ is only defined if $\mu \preceq \nu$, although it makes sense to adopt the convention that $D_{\mathrm{KL}}(\mu\|\nu) = +\infty$ if μ is not absolutely continuous with respect to ν.

(Other well-known f-divergences include the *Hellinger distance*, with $f_{\mathrm{H}}(\rho) := (\sqrt{\rho} - 1)^2$, the *Pearson divergence*, with $f_{\mathrm{P}}(\rho) := (\rho - 1)^2$, and the *total variation* (or *Kolmogorov*) *distance*, with $f_{\mathrm{TV}}(\rho) := |\rho - 1|$.)

Now let $\mathcal{X}$ and $\mathcal{Y}$ be Polish spaces. By classical results that are repeated for reference in appendix D, when $\mathcal{X}$ is Polish, the space $\mathscr{P}(\mathcal{X})$ of probability measures on $\mathcal{X}$ is itself a Polish space with respect to the topology of weak convergence. Let $\pi_{\mathcal{X}}$ be a strictly positive and Radon measure on $\mathcal{X}$, which shall be called a *reference measure*. Similarly, let $\pi_{\mathcal{Y}}$ be a reference measure on $\mathcal{Y}$, and let $\pi_{\mathcal{Z}}$ be the product measure $\pi_{\mathcal{X}} \otimes \pi_{\mathcal{Y}}$ on $\mathcal{Z} := \mathcal{X} \times \mathcal{Y}$. Clearly $\pi_{\mathcal{Z}}$ is also a reference measure.

Definition 5.3.2. Given $\mathscr{W}\colon \mathcal{Z} \to \mathbb{R}$, an f-divergence D_f, $\varepsilon > 0$, and $\mu_1 \in \mathscr{P}(\mathcal{X})$, the *interior-point regularization problem* is to find

$$\mu_{1,2} \in \arg\min \left\{ \mathcal{C}^\varepsilon[\nu] \middle| \nu \in \mathscr{P}(\mathcal{Z}), (\mathrm{proj}_{\mathcal{X}})_* \nu = \mu_1 \right\}, \tag{5.3.5}$$

where the cost functional $\mathcal{C}^\varepsilon$ is defined by

$$\mathcal{C}^\varepsilon[\nu] := \int_{\mathcal{Z}} \mathscr{W} \,\mathrm{d}\nu + \varepsilon D_f(\nu \| \pi_{\mathcal{Z}}). \tag{5.3.6}$$

The second marginal $\Upsilon\mu_1 \equiv \mu_2 := (\mathrm{proj}_{\mathcal{Y}})_* \mu_{1,2} \in \mathscr{P}(\mathcal{Y})$ will be called the *thermalized gradient descent of* μ_1. (See figure 5.3.1 on page 80 for a schematic diagram.)

In general, it is not clear whether minimizers of (5.3.6) always exist or are unique; there may be no explicit representation for the minimizer even if one does exist. However, one case is eminently accessible: that of the Kullback–Leibler divergence D_{KL}, i.e. the relative entropy functional.

Theorem 5.3.3. *If $\mu_1 \preceq \pi_{\mathcal{X}}$, then the solution to the interior-point regularization problem with respect to $\mathscr{W}$ and the Kullback–Leibler divergence D_{KL} is given by*

$$\frac{\mathrm{d}\mu_{1,2}}{\mathrm{d}\pi_{\mathcal{Z}}}(x,y) = \frac{\mathrm{d}\mu_1}{\mathrm{d}\pi_{\mathcal{X}}}(x) \frac{\exp(-\mathscr{W}(x,y)/\varepsilon)}{Z(x)},$$

where the "partition function" $Z\colon \mathcal{X} \to \mathbb{R}$ is given by

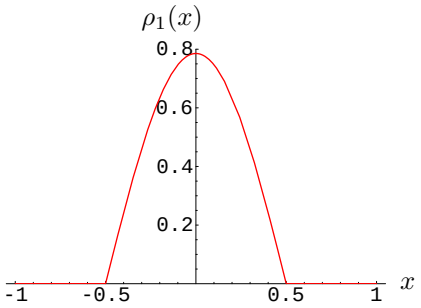
$$Z(x) := \int_{\mathcal{Y}} \exp(-\mathscr{W}(x,y)/\varepsilon) \,\mathrm{d}\pi_{\mathcal{Y}}(y).$$

Hence, $\mu_2 := (\mathrm{proj}_{\mathcal{Y}})_* \mu_{1,2} \preceq \pi$ with

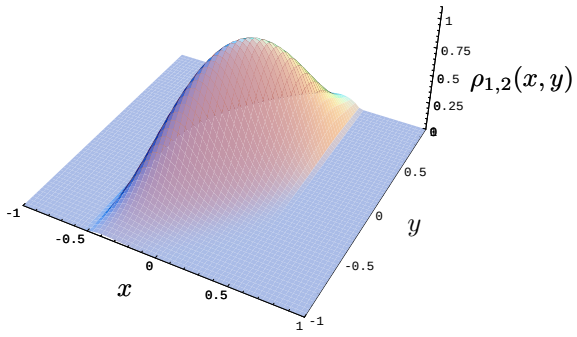
$$\frac{\mathrm{d}\mu_2}{\mathrm{d}\pi_{\mathcal{Y}}}(y) = \int_{\mathcal{X}} \frac{\mathrm{d}\mu_1}{\mathrm{d}\pi_{\mathcal{X}}}(x) \frac{\exp(-\mathscr{W}(x,y)/\varepsilon)}{Z(x)} \,\mathrm{d}\pi_{\mathcal{X}}(x). \tag{5.3.7}$$

Proof. Since this is a constrained optimization problem, the proof utilizes Lagrange multipliers on the Banach space $L^1(\mathcal{Z}, \pi_{\mathcal{Z}}; \mathbb{R})$ [Zei95]. Let $\rho_1 := \frac{\mathrm{d}\mu_1}{\mathrm{d}\pi_{\mathcal{X}}}$ be the probability density function of the given measure $\mu_1 \in \mathscr{P}(\mathcal{X})$ with respect to $\pi_{\mathcal{X}}$. It is required to find a non-negative function $u \in L^1(\mathcal{Z}, \pi_{\mathcal{Z}}; \mathbb{R})$ such that

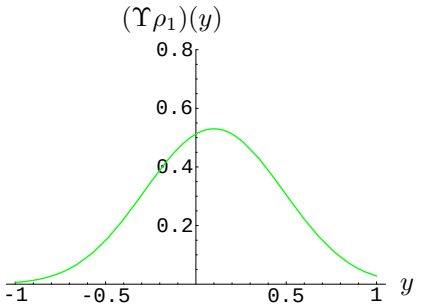
$$u \text{ minimizes } \int_{\mathcal{Z}} (\mathscr{W}\, u + \varepsilon u \log u) \,\mathrm{d}\pi_{\mathcal{Z}}$$



(a) Prior distribution



(b) Optimal joint distribution



(c) Posterior distribution

Figure 5.3.1: An example of a single step in the thermalized gradient descent scheme: (a) a prior probability density ρ_1; (b) the joint probability density $\rho_{1,2}$ minimizing the cost functional (5.3.6) and having ρ_1 as its first marginal; (c) the posterior probability density $\rho_2 = \Upsilon\rho_1$, equal to the second marginal of $\rho_{1,2}$.

subject to the constraints

$$\int_{\mathcal{Y}} u(x, \cdot) \, \mathrm{d}\pi_{\mathcal{Y}} = \rho_1(x) \text{ for } \pi_{\mathcal{X}}\text{-a.e. } x \in \mathcal{X}, \tag{5.3.8a}$$

$$\int_{\mathcal{Z}} u \, \mathrm{d}\pi_{\mathcal{Z}} = 1. \tag{5.3.8b}$$

The first constraint induces a one-parameter family of Lagrange multipliers $\ell \colon \mathcal{X} \to \mathbb{R}$; the second induces a single Lagrange multiplier $m \in \mathbb{R}$; together, these yield the Lagrangian

$$\mathcal{L}[u] := \int_{\mathcal{Z}} (\mathscr{W} u + \ell \, u + mu + \varepsilon u \log u) \, \mathrm{d}\pi_{\mathcal{Z}}.$$

Suppose that u is a critical point for this Lagrangian, i.e. $\delta\mathcal{L}[u] = 0$. That is, for any $v \in L^1(\mathcal{Z}, \pi_{\mathcal{Z}}; \mathbb{R})$,

$$\frac{\mathrm{d}}{\mathrm{d}t} \mathcal{L}[u + tv] \bigg|_{t=0} = 0.$$

A straightforward but tedious calculation of the derivative at $t = 0$ yields the requirement that

$$\int_{\mathcal{Z}} (\mathscr{W} v + \ell \, v + mv + \varepsilon v \log u) \, \mathrm{d}\pi_{\mathcal{Z}} = 0,$$

and it is clear that this equality is satisfied by

$$u(x, y) = \exp\left(-\frac{\mathscr{W}(x, y) + \ell(x) + m + \varepsilon}{\varepsilon} \right).$$

Application of the constraints (5.3.8) yields that $m = -\varepsilon$ and that

$$\ell(x) = -\varepsilon \log \frac{\rho_1(x)}{Z(x)},$$

as required. ■

There is an analogue of theorem 5.3.3 for more general f-divergences, but it requires an additional hypothesis on f and does not yield an explicit formula for the minimizer. Under the assumption that f is differentiable and f' is invertible, the steps in the proof of theorem 5.3.3 can be followed

mutatis mutandis to yield that

$$u(x,y) = (f')^{-1}\left(-\frac{\mathscr{W}(x,y)+\ell(x)+m}{\varepsilon}\right).$$

It is not clear, in general, how to apply the constraints (5.3.8) to eliminate the Lagrange multipliers ℓ and m. Furthermore, there are classical f-divergences for which the hypothesis of invertibility of the derivative of f fails, the total variation distance being one example. For these reasons, attention will henceforth be confined to the Kullback–Leibler (negative entropy) case.

Theorem 5.3.3 suggests that the thermalized gradient descent of an arbitrary probability measure $\mu \in \mathscr{P}(\mathcal{X})$ to a measure $\Upsilon\mu \in \mathscr{M}(\mathcal{Y})$ should be defined by

$$\mathrm{d}(\Upsilon\mu)(y) := \left(\int_{\mathcal{X}} \frac{\exp(-\mathscr{W}(x,y)/\varepsilon)}{Z(x)}\,\mathrm{d}\mu(x)\right)\mathrm{d}\pi_{\mathcal{Y}}(y). \tag{5.3.9}$$

At the moment, this is just a formal definition that happens to agree with (5.3.7) when $\mu \preceq \pi_{\mathcal{X}}$. However, if it can be shown that the map Υ is continuous with respect to the topology of weak convergence of measures, then Υ extends uniquely from $\mathscr{P}_{\preceq\pi_{\mathcal{X}}}(\mathcal{X})$ to its completion in $\mathscr{P}(\mathcal{X})$. Fortunately, this holds under quite mild assumptions on $\mathscr{W}$:

Lemma 5.3.4. *Suppose that $\mathscr{W}\colon \mathcal{Z} \to \mathbb{R}$ is continuous and that, for every $x \in \mathcal{X}$, there exists a neighbourhood $U[x]$ of x such that*

$$\int_{\mathcal{Y}} \left(\sup_{x' \in U[x]} \exp(-\mathscr{W}(x',y)/\varepsilon)\right)\mathrm{d}\pi_{\mathcal{Y}}(y) < +\infty.$$

Then $\Upsilon\colon \mathscr{P}(\mathcal{X}) \to \mathscr{P}(\mathcal{Y})$ is continuous with respect to the weak topology.

The proof of this lemma is given on page 92.

Hence, since Υ is weakly continuous on $\mathscr{M}(\mathcal{X})$, it is weakly continuous when restricted to $\mathscr{P}(\mathcal{X})$. Since $\mathcal{X}$ is Polish, so is $\mathscr{P}(\mathcal{X})$, and so Υ is Cauchy continuous on both $\mathscr{P}(\mathcal{X})$ and the subset $\mathscr{P}_{\preceq\pi_{\mathcal{X}}}(\mathcal{X})$. However, Υ agrees with (5.3.7) on $\mathscr{P}_{\preceq\pi_{\mathcal{X}}}(\mathcal{X})$, so Υ must be the unique weakly continuous extension of (5.3.7) to the completion of $\mathscr{P}_{\preceq\pi_{\mathcal{X}}}(\mathcal{X})$, and, by lemma 5.5.1, that completion is $\mathscr{P}(\mathcal{X})$.

In particular, and again by lemma 5.5.1, Dirac measures lie in the domain

of the thermalized gradient descent operator Υ, and so it is now an easy matter to specify the transition kernels for Markov chains (and higher-order stochastic processes) associated to interior-point regularization problems.

Before moving on to examine stochastic processes generated in this way and relevant to the study of dissipative evolutions, one important general note should be made: even in the most basic cases, the choice of reference measure makes a difference. As the following change-of-measure formula shows, a change of reference measure is equivalent to the addition of a bias to the cost function $\mathscr{W}$ that is given by the logarithm of the Radon–Nikodým derivative of the two reference measures.

An example of the change-of-measure formula as applied to Markov chains is given by considering the state space $\mathbb{R}^n$ with a "flat" energy $E(t, x) \equiv 0$ and the dissipation corresponding to linear kinetics, i.e. $\Psi(v) = \frac{1}{2}|v|^2$. Let X denote the Markov chain computed with respect to Lebesgue measure; let Y denote the Markov chain computed with respect to a centred Gaussian measure. It will be shown in the next section that a suitable continuous interpolation of X converges in mean square to a standard Wiener process (Brownian motion) (see theorem 5.4.1); on the other hand, the same interpolation of Y converges in mean square to an Ornstein–Uhlenbeck process (see theorem 5.4.2).

Theorem 5.3.5 (Change-of-measure formula)**.** *With the same notation as theorem 5.3.3, suppose that $\tilde{\pi}_{\mathcal{X}}$ and $\tilde{\pi}_{\mathcal{Y}}$ are also reference measures on $\mathcal{X}$ and $\mathcal{Y}$, equivalent to $\pi_{\mathcal{X}}$ and $\pi_{\mathcal{Y}}$ respectively, and let Υ and $\tilde{\Upsilon}$ be the corresponding thermalized gradient descent maps. Then "$\tilde{\Upsilon}$ with respect to $\mathscr{W}$ is Υ with respect to $\tilde{\mathscr{W}}$", where*

$$\tilde{\mathscr{W}}(x, y) := \mathscr{W}(x, y) - \varepsilon \log \frac{\mathrm{d}\tilde{\pi}_{\mathcal{Y}}}{\mathrm{d}\pi_{\mathcal{Y}}}(y) + \varepsilon \log \frac{\mathrm{d}\tilde{\pi}_{\mathcal{X}}}{\mathrm{d}\pi_{\mathcal{X}}}(x). \tag{5.3.10}$$

Note that, as usual, the "$+\varepsilon \log \frac{\mathrm{d}\tilde{\pi}_{\mathcal{X}}}{\mathrm{d}\pi_{\mathcal{X}}}(x)$" term can be included or omitted at will in most expressions.

Proof. Let $\mu \in \mathscr{P}(\mathcal{X})$. By theorem 5.3.3,

$$\begin{aligned}
&\frac{\mathrm{d}(\tilde{\Upsilon}\mu)}{\mathrm{d}\tilde{\pi}_{\mathcal{Y}}}(y) \\
&\quad = \frac{\mathrm{d}(\tilde{\Upsilon}\mu)}{\mathrm{d}\pi_{\mathcal{Y}}}(y)\frac{\mathrm{d}\pi_{\mathcal{Y}}}{\mathrm{d}\tilde{\pi}_{\mathcal{Y}}}(y) \\
&\quad = \int_{\mathcal{X}} \frac{\frac{\mathrm{d}\mu}{\mathrm{d}\pi_{\mathcal{X}}}(x)\exp(-\mathscr{W}(x,y)/\varepsilon)}{\int_{\mathcal{Y}} \exp(-\mathscr{W}(x,y)/\varepsilon)\,\mathrm{d}\tilde{\pi}_{\mathcal{Y}}(y)}\,\mathrm{d}\pi_{\mathcal{X}}(x) \\
&\quad = \int_{\mathcal{X}} \frac{\frac{\mathrm{d}\mu}{\mathrm{d}\tilde{\pi}_{\mathcal{X}}}(x)\frac{\mathrm{d}\tilde{\pi}_{\mathcal{X}}}{\mathrm{d}\pi_{\mathcal{X}}}(x)\exp(-\mathscr{W}(x,y)/\varepsilon)}{\int_{\mathcal{Y}} \exp(-\mathscr{W}(x,y)/\varepsilon)\frac{\mathrm{d}\tilde{\pi}_{\mathcal{Y}}}{\mathrm{d}\pi_{\mathcal{Y}}}(y)\,\mathrm{d}\pi_{\mathcal{Y}}(y)}\frac{\mathrm{d}\pi_{\mathcal{X}}}{\mathrm{d}\tilde{\pi}_{\mathcal{X}}}(x)\,\mathrm{d}\tilde{\pi}_{\mathcal{X}}(x) \\
&\quad = \int_{\mathcal{X}} \frac{\frac{\mathrm{d}\mu}{\mathrm{d}\tilde{\pi}_{\mathcal{X}}}(x)\exp(-\mathscr{W}(x,y)/\varepsilon)}{\int_{\mathcal{Y}} \exp(-\mathscr{W}(x,y)/\varepsilon)\frac{\mathrm{d}\tilde{\pi}_{\mathcal{Y}}}{\mathrm{d}\pi_{\mathcal{Y}}}(y)\,\mathrm{d}\pi_{\mathcal{Y}}(y)}\,\mathrm{d}\tilde{\pi}_{\mathcal{X}}(x).
\end{aligned}$$

Hence,

$$\begin{aligned}
\frac{\mathrm{d}(\tilde{\Upsilon}\mu)}{\mathrm{d}\pi_{\mathcal{Y}}}(y) &= \frac{\frac{\mathrm{d}\tilde{\pi}}{\mathrm{d}\pi}(y)\exp(-\mathscr{W}(x,y)/\varepsilon)}{\int_{\mathcal{X}} \frac{\mathrm{d}\tilde{\pi}}{\mathrm{d}\pi}(y)\exp(-\mathscr{W}(x,y)/\varepsilon)\,\mathrm{d}\pi(y)} \\
&= \frac{\exp(-\tilde{\mathscr{W}}(x,y)/\varepsilon)}{\int_{\mathcal{X}} \exp(-\tilde{\mathscr{W}}(x,y)/\varepsilon)\,\mathrm{d}\pi(y)},
\end{aligned}$$

where $\tilde{\mathscr{W}}$ is as in (5.3.10). ■

5.3.2 Markov Chains

It is now an easy matter to define a Markov chain using a sequence of interior-point regularization problems:

Definition 5.3.6. Let $\mathcal{X}$ be a Polish space and let π be a reference measure on $\mathcal{X}$. Given a sequence of cost functions $(\mathscr{W}_i)_{i\in\mathbb{N}_0}$ and a positive sequence $(\varepsilon_i)_{i\in\mathbb{N}_0}$, the associated *thermalized gradient descent Markov chain* $X\colon \Omega \times \mathbb{N}_0 \to \mathcal{X}$ is the Markov chain with single-step transition probabilities given

by

$$\mathbb{P}\big[X_{i+1} \in A \big| X_i = x_i\big] = \frac{1}{Z(x_i)} \int_A \exp(-\mathscr{W}_{i+1}(x_i, x_{i+1})/\varepsilon_{i+1}) \, \mathrm{d}\pi(x_{i+1}).$$

Very often, this will simply be called *the Markov chain*.

In the sequel, attention will be confined to cost functions $\mathscr{W}$ that are of energetics-and-dissipation form: that is, given an energetic potential $E \colon [0,T] \times \mathcal{X} \to \mathbb{R}$, a dissipation potential $\Psi \colon \mathcal{X} \to \mathbb{R}$ and a partition $P = \{0 = t_0 < t_1 < \ldots < t_N = T\}$ of $[0,T]$, $\mathscr{W}_{i+1}$ takes the form

$$\mathscr{W}_{i+1}(x_i, x_{i+1}) := E(t_{i+1}, x_{i+1}) - E(t_i, x_i) + \Delta t_{i+1} \Psi \left(\frac{\Delta x_{i+1}}{\Delta t_{i+1}} \right)$$

and the Markov chain $X = X^{(P)}$ is defined on $\Omega \times P$. By the usual abuse of notation, $X^{(P)}$ will also stand for its own càdlàg piecewise-constant interpolation, defined on $\Omega \times [0,T]$, i.e.

$$X^{(P)}(t) := X_i^{(P)} \text{ for } t \in [t_i, t_{i+1}).$$

Note that the "prior energy" $E(t_i, x_i)$ has no effect on minimization problems for the t_{i+1} time step, plays no part in determining the transition probabilities for the t_{i+1} time step, and can be omitted or included according to taste. Another important observation is that the parameters $\varepsilon_i > 0$, which model the temperature of the heat bath, need not be constant — indeed, this is of vital importance in chapter 6, in which $\varepsilon_i := \theta \Delta t_{i+1}$ for a fixed $\theta > 0$.

The stage is now set to phrase the original question of what happens when a dissipative system is placed in contact with a heat bath in mathematical terms:

> Taking a suitable continuous-time interpolation of $X^{(P)}$, what is the limiting law on the space of all functions from $[0,T]$ into $\mathcal{X}$ as $[\![P]\!] \to 0$?

5.3.3 Higher-Order Processes

If the only intended application of the interior-point regularization problem were to generate Markov chains on a state space $\mathcal{X}$, then the generalized

framework with two spaces $\mathcal{X}$ and $\mathcal{Y}$ would have been surplus to requirements. The strength of the two-space approach, though, is that it allows not only the definition of Markov chains (which model the effects of heat baths on first-order evolution equations that neglect inertial effects) but also higher-order processes (which can model, for example, the effects of heat baths on second-order evolution equations that take account of inertia).

For example, consider a particle with inertial mass $m > 0$, the state x of which evolves according to Newton's second law:

$$m\ddot{x} \in -\mathrm{D}E(t, x) + \partial\Psi(\dot{x}),$$

where, as before, E is a time-dependent energetic potential and Ψ is a dissipation potential. On a partition $P \in \mathcal{P}([0, T])$, this equation of motion has the incremental work function

$$\begin{aligned}&\mathscr{W}_{i+1}(x_{i-1}, x_i, x_{i+1})\\ &\quad := \frac{m}{2}\frac{|\Delta x_{i+1} - \Delta x_i|^2}{\Delta t_{i+1}\Delta t_i} + E(t_{i+1}, x_{i+1}) - E(t_i, x_i) + \Delta t_{i+1}\Psi\left(\frac{\Delta x_{i+1}}{\Delta t_{i+1}}\right).\end{aligned}$$

Application of the thermalized gradient descent scheme to this incremental work function yields a second-order stochastic process in which the distribution of the random state X_{i+1} depends on the previous *two* states X_i and X_{i-1}. In theory, it ought to be possible to make the appropriate calculations and extract a limiting law on path space as $[\![P]\!] \to 0$.

A one-dimensional prototypical example of such an evolution would be the classical mechanical problem of an elastically restrained mass on a moving belt [Fid06, chapter 2]; see figure 5.3.2 on page 87 for a schematic illustration. Consider a box of mass $m > 0$, resting in contact with a rough belt that moves with constant velocity $v \in \mathbb{R}$. The box/belt interface exhibits dry friction with coefficient of friction $\mu > 0$ and the box is restrained by an elastic spring with spring constant $\kappa > 0$. Let $x(t) \in \mathbb{R}$ be the displacement of the box from the rest length of the spring at time t. The equation of motion is

$$m\ddot{x}(t) \in -\kappa x(t) + \partial\Psi(\dot{x}(t) - v);$$

the dissipation potential is $\Psi(x) := \mu m g|x|$, where g is the local acceleration due to gravity, $9.80665\,\mathrm{m\,s^{-2}}$ at sea level on Earth.

However, the analysis of even a simple example such as this is very com-

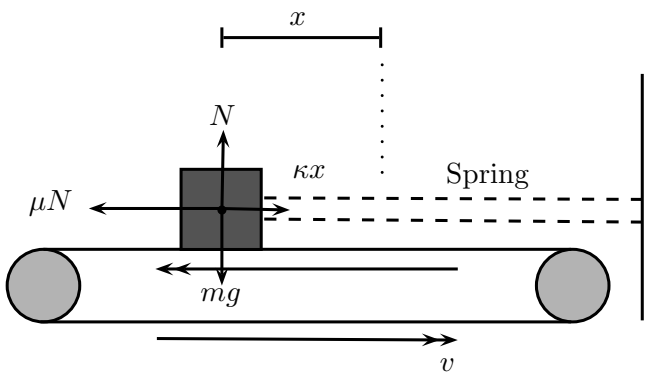


Figure 5.3.2: A schematic diagram of the classical mechanical problem of an elastically restrained mass on a moving belt. The block (dark grey), of mass m, rests on a taut rough belt which moves around the two spindles (light grey) at constant speed v. It is restrained by a spring of elasticity κ. x denotes the displacement from the rest length of the spring. As drawn, the box is in vertical equilibrium ($N = -mg$) but the extension of the spring is insufficient to overcome the frictional resistance μmg.

plicated. Even a back-of-an-envelope calculation of the mean and variance of X_{i+1} given X_i and X_{i-1} yields long and unwieldy expressions — although, promisingly, coarse approximations are normally distributed and incorporate the effect of the spring reversing the direction of evolution as the spring is extended/compressed. The non-Markovian character of the processes further compounds the difficulties of the analysis. Therefore, evolutions like this are left as a topic for future research.

5.4 Examples with 2-Homogeneous Dissipation

In this section, the claim that thermalized gradient descents provide a reasonable model for subjecting the gradient descent in E and Ψ to the effects of a heat bath is justified by means of some examples in which the dissipation potential Ψ is homogeneous of degree two, i.e. linear kinetics. It turns out that the thermalizing procedure generates (discretizations of) stochastic differential equations with Itō noise. With the exception of the first example, detailed convergence proofs will not be given.

5.4.1 Brownian Motion

As a simple first example, consider a trivial energetic potential ($E \equiv 0$) and the dissipation potential $\Psi(v) := \frac{1}{2}|v|^2$. It seems intuitively reasonable to hypothesize that the effect of a heat bath on such a set-up should be to generate a Brownian motion (a Wiener process). Indeed, this is what does happen.

Theorem 5.4.1. *Let $X^{(P)} \colon \Omega \times P \to \mathbb{R}^n$ be the Markov chain calculated with respect to n-dimensional Lebesgue measure, the trivial energetic potential $E \equiv 0$, the dissipation potential $\Psi(v) := \frac{1}{2}|v|^2$ and a partition P of $[0, T]$. Then, as $[\![P]\!] \to 0$, $X^{(P)}$ converges in law in path space $\mathcal{C}^0([0, T]; \mathbb{R}^n)$ to $Y := x_0 + \sqrt{\varepsilon} W$, where W is a standard Wiener process (Brownian motion) on $\mathbb{R}^n$.*

Proof. A direct calculation gives that

$$\begin{aligned} \rho_{i+1}(x_{i+1} \mid x_i) &\propto \exp\left(-\frac{\Delta t_{i+1}}{\varepsilon} \Psi\left(\frac{\Delta x_{i+1}}{\Delta t_{i+1}}\right)\right) \\ &= \exp\left(-\frac{|x_{i+1} - x_i|^2}{2\varepsilon \Delta t_{i+1}}\right). \end{aligned}$$

That is, given $X_i^{(P)}$, the state $X_{i+1}^{(P)}$ is normally distributed with expected value $X_i^{(P)}$ and covariance operator $\varepsilon \Delta t_{i+1} \mathbb{1}$. Thus, on any compact interval of time $[0, T]$, the claimed weak convergence follows from the Donsker–Prokhorov invariance principle [Don51] [Pro54]. ∎

This weak convergence result can be improved upon somewhat. For example, by [Ött96], the càdlàg piecewise-constant interpolants converge strongly with order $1/4$ at every time, i.e. for all $t \geq 0$, there exists a constant $C_t > 0$ such that

$$\mathbb{E}\left[\left|\bar{X}^{(P)}(t) - Y(t)\right|^2\right] \leq C_t [\![P]\!]^{1/2}.$$

If a suitable stochastic interpolation scheme (to be defined in a moment) is used, then L^2 convergence with respect to the uniform norm on path space is obtained: by [HMS02], $\tilde{X}$ converges to Y in mean square; [YM08] implies that the order of convergence is one half, i.e. there exists a constant $C_T > 0$

such that

$$\mathbb{E}\left[\sup_{t\in[0,T]}\left|\tilde{X}^{(P)}(t)-Y(t)\right|^2\right]\leq C_T[\![P]\!].$$

Given that the claimed limit process satisfies an Itō stochastic differential equation of the form

$$\dot{Y}(t)=f(Y(t))+g(Y(t))\,\dot{W}(t),$$

the corresponding stochastic interpolation $\tilde{X}^{(P)}$ of $X^{(P)}$ is given by

$$\begin{aligned}&\tilde{X}^{(P)}(t)\\&\quad:=X_i^{(P)}+(t-t_i)f\big(X_i^{(P)}\big)+g\big(X_i^{(P)}\big)(W(t)-W(t_i))\text{ for }t\in[t_i,t_{i+1})\\&\quad\equiv X_0+\int_0^t f\big(X^{(P)}(s)\big)\,\mathrm{d}s+\int_0^t g\big(X^{(P)}(s)\big)\,\mathrm{d}W(s)\text{ for }t\in[0,T].\end{aligned}$$

This simple example is one of the cases in which the path-integral intuition is a valid way to describe the behaviour of the limiting system as $[\![P]\!]\to 0$. That is, the exponential term in the Chapman–Kolmogorov equation (5.2.6) is

$$\begin{aligned}\exp\left(-\sum_{i=1}^k\frac{1}{2\varepsilon\Delta t_{i+1}}\left|\Delta x_{i+1}\right|^2\right)&=\exp\left(-\frac{1}{\varepsilon}\sum_{i=1}^k\Delta t_{i+1}\frac{1}{2}\left|\frac{\Delta x_{i+1}}{\Delta t_{i+1}}\right|^2\right)\\&\to\exp\left(-\frac{1}{\varepsilon}\int_0^T\frac{1}{2}\left|\dot{x}(t)\right|^2\mathrm{d}t\right)\end{aligned}$$

in the limit as $[\![P]\!]\to 0$ and $k\to\infty$ with $t_{k+1}=\sum_{i=1}^k\Delta t_{i+1}=T$. Let $\mathcal{I}$ denote the integral functional on path space so arrived at:

$$\mathcal{I}[u]:=\begin{cases}\dfrac{1}{2}\displaystyle\int_0^T\left|\dot{u}(t)\right|^2\mathrm{d}t, & \text{if } u\in H^1([0,T];\mathbb{R}^n),\\ +\infty, & \text{otherwise.}\end{cases}$$

By Schilder's theorem [DZ98, theorem 5.2] [Sch66], $\mathcal{I}$ is the rate function for standard Brownian motion. This thesis will not dwell on these issues, but it might be hoped that connections can be made between the thermalized gradient descent method and the Freĭdlin–Wentzell theory for stochastic diffusions [FW98].

5.4.2 An Applied Load

A mild generalization of the previous example is the case of a time-dependent applied load, i.e. $E(t,x) := -\langle \ell(t), x\rangle$, where $\ell\colon [0,T] \to (\mathbb{R}^n)^*$ is uniformly Lipschitz. Another mild generalization is to assume that the dissipation potential is a non-degenerate quadratic form: let $A \in \mathbb{R}^{n\times n}$ be a symmetric and positive-definite matrix (the "viscosity matrix") and let $\Psi(v) := \frac{1}{2}\langle Av, v\rangle$. (Hence, both A and A^{-1} have square roots.) Direct calculation gives

$$\begin{aligned}\rho_{i+1}(x_{i+1} \mid x_i) &\propto \exp\left(-\frac{1}{\varepsilon}\left(-\langle \ell(t_{i+1}), x_{i+1}\rangle + |\Delta t_{i+1}|\Psi\left(\frac{x_{i+1}-x_i}{\Delta t_{i+1}}\right)\right)\right)\\ &= \exp\left(-\frac{1}{\varepsilon\Delta t_{i+1}}\Psi\big(\Delta x_{i+1} - A^{-1}\ell(t_{i+1})\Delta t_{i+1}\big)\right).\end{aligned}$$

That is, given X_i, the state X_{i+1} is normally distributed with mean $X_i + A^{-1}\ell(t_{i+1})\Delta t_{i+1}$ and covariance operator $\varepsilon\Delta t_{i+1}A^{-1}$. This gives a limiting stochastic differential equation

$$\begin{cases}\mathrm{d}Y(t) = A^{-1}\ell(t)\,\mathrm{d}t + \sqrt{\varepsilon A^{-1}}\,\mathrm{d}W,\\ Y(0) = x_0,\end{cases}$$

where W is a standard Wiener process (Brownian motion) on $\mathbb{R}^n$.

5.4.3 Itō Gradient Descents

Similarly, if $V\colon \mathbb{R}^n \to [0,+\infty)$ is a scalar potential satisfying appropriate smoothness and growth conditions, then the Itō stochastic gradient descent equation

$$\dot{Y}(t) = -\nabla V(Y(t)) + \sqrt{\varepsilon}\,\dot{W}(t)$$

has a strong solution. The thermalized gradient descent Markov chain in V with respect to $\Psi(v) := \frac{1}{2}|v|^2$ has transition probabilities with densities

$$\rho_{i+1}(x_{i+1} \mid x_i) \propto \exp\left(-\frac{1}{\varepsilon}\left(V(x_{i+1}) - V(x_i) + \frac{|\Delta x_{i+1}|^2}{2\Delta t_{i+1}}\right)\right).$$

To a first-order approximation, this is

$$\rho_{i+1}(x_{i+1} \mid x_i) \propto \exp\left(-\frac{|x_{i+1} - (x_i - \Delta t_{i+1}\nabla V(x_i))|^2}{2\varepsilon\Delta t_{i+1}}\right),$$

i.e. $X_{i+1}^{(P)}$ is normally distributed with mean $X_i^{(P)} - \Delta t_{i+1} \nabla V(X_i^{(P)})$ and covariance operator $\varepsilon \Delta t_{i+1} \mathbb{1}$; generating a sequence of random variables according to such a scheme is the usual Euler–Maruyama method for approximating Y.

5.4.4 A Gaussian Reference Measure

As a final example, consider the state space $\mathbb{R}^n$ equipped with a non-degenerate Gaussian measure instead of Lebesgue measure. Let γ be the centred Gaussian probability measure on $\mathbb{R}^n$ with invertible covariance operator $\Gamma \colon \mathbb{R}^n \to \mathbb{R}^n$; that is, γ is given in terms of its Radon–Nikodým derivative with respect to Lebesgue measure by

$$\frac{\mathrm{d}\gamma}{\mathrm{d}x}(x) = \frac{1}{(2\pi \det \Gamma)^{n/2}} \exp\left(-\frac{x \cdot \Gamma^{-1} x}{2} \right). \tag{5.4.1}$$

Consider the trivial energetic potential $E(t, x) \equiv 0$ and the dissipation potential $\Psi(v) := \frac{1}{2}|v|^2$. Then

$$\mathbb{P}\left[X_{i+1}^{(P)} \in A \,\middle|\, X_i^{(P)} = x_i \right] \propto \int_A \exp\left(-\left(\frac{|\Delta x_{i+1}|^2}{2\varepsilon \Delta t_{i+1}} + \frac{x \cdot \Gamma^{-1} x}{2} \right) \right) \mathrm{d}x.$$

Completing the square in the exponent yields that the law of $X_{i+1}^{(P)}$ given $X_i^{(P)}$ is Gaussian with mean $A^{-1} X_i^{(P)}$ and covariance operator $\varepsilon h A^{-1}$, where $A := \mathbb{1} + \varepsilon h \Gamma^{-1}$. Up to leading order, this is the usual discretization of an Ornstein–Uhlenbeck process, and so it suggests the following (the proof of which will be omitted):

Theorem 5.4.2. *Let $X^{(P)}$ be the Markov chain calculated with respect to the Gaussian measure γ (5.4.1), the trivial energetic potential $E \equiv 0$, the dissipation potential $\Psi(v) := \frac{1}{2}|v|^2$ and a partition P of $[0, T]$. Then, as $[\![P]\!] \to 0$, $X^{(P)}$ converges in distribution to the solution U of the Ornstein–Uhlenbeck equation*

$$\dot{U}(t) = -\varepsilon \Gamma^{-1} U(t) + \sqrt{\varepsilon}\, \dot{W}(t),$$

where W is a standard n-dimensional Wiener process.

As before, if the continuous stochastic interpolation $\tilde{X}^{(P)}$ is used, then [YM08] gives global convergence in mean square of order one half: there

exists a constant $C \geq 0$ such that

$$\mathbb{E}\left[\sup_{0\leq t\leq T}\left|\tilde{X}^{(P)}(t) - U(t)\right|^2\right] \leq C[\![P]\!].$$

As a concluding remark on directions for possible future research, note that theorem 5.4.2 suggests that the interior-point regularization and thermalized gradient descent method may provide a tool for the approximation of stochastic partial differential equations. A plausible conjecture would be that the Markov chain on an infinite-dimensional separable Hilbert space $\mathcal{H}$ with respect to the trivial energetic potential and the dissipation potential $\Psi(v) := \frac{1}{2}\|v\|_{\mathcal{H}}^2$ converges in a suitable sense to an infinite-dimensional Ornstein–Uhlenbeck process, i.e. the solution to a stochastic heat equation.

5.5 Proofs and Supporting Results

Proof of lemma 5.3.4. Since the weak topology on the space of probability measures on a Polish space is metrizable, it suffices to check continuity in terms of sequential continuity. Suppose that μ_n is a sequence in $\mathscr{P}(\mathcal{X})$ such that $\mu_n \rightharpoonup \mu$; it is required to show that $\Upsilon\mu_n \rightharpoonup \Upsilon\mu$. Let $\phi\colon \mathcal{Y} \to \mathbb{R}$ be bounded and continuous. Define $\Phi\colon \mathcal{X} \to \mathbb{R}$ by

$$\Phi(x) := \frac{1}{Z(x)}\int_{\mathcal{Y}} \phi(y)e^{-\mathscr{W}(x,y)/\varepsilon}\,\mathrm{d}\pi_{\mathcal{Y}}(y),$$

where, as usual,

$$Z(x) := \int_{\mathcal{Y}} e^{-\mathscr{W}(x,y)/\varepsilon}\,\mathrm{d}\pi_{\mathcal{Y}}(y).$$

Since, by assumption, the integral in the denominator is finite, Φ is bounded with $\|\Phi\|_\infty \leq \|\phi\|_\infty$.

Since $\mathscr{W}$ is continuous, so are both $(x,y) \mapsto e^{-\mathscr{W}(x,y)/\varepsilon}$ and $(x,y) \mapsto \phi(y)e^{-\mathscr{W}(x,y)/\varepsilon}$; hence, these functions are measurable as well. The assumption in the statement of the lemma ensures that Lebesgue's dominated convergence theorem in the form of theorem B.2 applies. Thus, both $Z\colon \mathcal{X} \to (0,+\infty)$ and

$$x \mapsto \int_{\mathcal{Y}} \phi(y)e^{-\mathscr{W}(x,y)/\varepsilon}\,\mathrm{d}\pi_{\mathcal{Y}}(y)$$

are continuous, and so Φ is continuous.

Hence,

$$\begin{aligned}
\int_{\mathcal{Y}} \phi \,\mathrm{d}(\Upsilon\mu_n) &= \int_{\mathcal{Y}} \phi(y) \int_{\mathcal{X}} \frac{e^{-\mathscr{W}(x,y)/\varepsilon}}{Z(x)} \,\mathrm{d}\mu_n(x)\mathrm{d}\pi_{\mathcal{Y}}(y) \text{ by definition of } \Upsilon, \\
&= \int_{\mathcal{X}} \int_{\mathcal{Y}} \frac{\phi(y)e^{-\mathscr{W}(x,y)/\varepsilon}}{Z(x)} \,\mathrm{d}\pi_{\mathcal{Y}}(y)\mathrm{d}\mu_n(x) \text{ by Fubini's theorem}, \\
&= \int_{\mathcal{X}} \Phi(x) \,\mathrm{d}\mu_n(x) \text{ by definition of } \Phi, \\
&\xrightarrow[n\to\infty]{} \int_{\mathcal{X}} \Phi(x) \,\mathrm{d}\mu(x) \text{ since } \mu_n \rightharpoonup \mu, \\
&= \int_{\mathcal{Y}} \phi \,\mathrm{d}(\Upsilon\mu) \text{ reversing lines 1–3},
\end{aligned}$$

and so $\Upsilon\mu_n \rightharpoonup \Upsilon\mu$, as claimed. ■

Lemma 5.5.1. *Let $(\mathcal{X}, d)$ be a metric space and π a strictly positive and locally finite Borel measure on $\mathcal{X}$.*

1. *For every $x \in \mathcal{X}$, there exists a sequence of probability measures μ_n on $\mathcal{X}$ such that $\mu_n \preceq \pi$ and $\mu_n \rightharpoonup \delta_x$.*

2. *If $\mathcal{X}$ is separable, then $\mathscr{P}_{\preceq\pi}(\mathcal{X})$ is dense in $\mathscr{P}(\mathcal{X})$ in the weak topology.*

Proof. 1. Let μ_n be the probability measure defined by

$$\mu_n(A) := \frac{\pi(A \cap \mathbb{B}_{1/n}(x))}{\pi(\mathbb{B}_{1/n}(x))};$$

since π is strictly positive, the denominator is never 0, and since π is locally finite, the numerator is finite for large enough n. Let U be any open subset of $\mathcal{X}$. If $x \in U$, then, for all large enough n, $\mathbb{B}_{1/n}(x) \subseteq U$, and so $\mu_n(U) = 1$. On the other hand, if $x \notin U$, then, for all large enough n, $\mathbb{B}_{1/n}(x) \cap U = \emptyset$, and so $\mu_n(U) = 0$. That is, for all large enough n (depending on x),

$$\mu_n(U) = \begin{cases} 0, & \text{if } x \notin U, \\ 1, & \text{if } x \in U. \end{cases}$$

Hence,

$$\lim_{n\to\infty} \mu_n(U) = \liminf_{n\to\infty} \mu_n(U) = \delta_x(U),$$

and the claim follows from the portmanteau theorem (theorem D.1).

2. Let D be any countable dense subset of $\mathcal{X}$. By the previous part, for every $x \in D$, δ_x lies in the closure of $\mathscr{P}_{\preceq\pi}(\mathcal{X})$. By theorem D.2, the closure of $\{\delta_x \mid x \in D\}$ is $\mathscr{P}(\mathcal{X})$. Hence, the closure of $\mathscr{P}_{\preceq\pi}(\mathcal{X})$ is $\mathscr{P}(\mathcal{X})$. ■

Chapter 6

Thermalized Gradient Descent II: 1-Homogeneous Dissipation

6.1 Introductory Remarks

This chapter makes rigorous the heuristics of chapter 5 in the case that the state space is $\mathbb{R}^n$, the energetic potential E is sufficiently well-behaved and the dissipation potential Ψ is homogeneous of degree one. That is, this chapter comprises the analysis of a rate-independent system

$$\partial\Psi(\dot{z}(t)) \ni -\mathrm{D}E(t, z(t))$$

in contact with a heat bath, where the effect of the heat bath is modeled by the interior-point regularization procedure. The main ingredients of this analysis were introduced in [SKTO09]. Of particular note is an effective dual dissipation potential, $\mathcal{F}_0^\star$, that controls the dynamics of the limiting continuous-time evolution and is determined purely by the dissipation potential Ψ. $\mathcal{F}_0^\star$ is a convex, extended-real-valued function defined on the dual space[6.1] of $\mathbb{R}^n$:

$$\mathcal{F}_0^\star(w) := \log \int_{\mathbb{R}^n} \exp(-(\langle w, z\rangle + \Psi(z)))\,\mathrm{d}z.$$

[6.1] In fact, in full generality, $\mathcal{F}_0^\star$ is defined on the cotangent bundle of the state space, and its dual is defined on the tangent bundle. See the remarks on manifold-type state spaces in subsection 6.6.1 for a derivation that explains this.

The limiting evolution is a deterministic ordinary differential equation of the form

$$\dot{x} = -\theta \mathrm{D}\mathcal{F}_0^\star(\mathrm{D}E(t,x)).$$

However, it would be closer to the "gradient descent spirit" of this thesis to describe the limiting evolution as a gradient descent with a smooth, convex, nonlinear effective dissipation potential $\mathcal{F}_0$ given by the convex conjugate of $\mathcal{F}_0^\star$:

$$\mathrm{D}\mathcal{F}_0(-\dot{x}(t)/\theta) = \mathrm{D}E(t,x), \tag{6.1.1}$$

where

$$\mathcal{F}_0(v) := \sup\{\langle \ell, v\rangle - \mathcal{F}_0^\star(\ell) \mid \ell \in (\mathbb{R}^n)^*\}.$$

In most cases, Ψ and hence $\mathcal{F}_0^\star$ and $\mathcal{F}_0$ are even, and so $\mathrm{D}\mathcal{F}_0$ is odd, and (6.1.1) can be rearranged into the more familiar form

$$\mathrm{D}\mathcal{F}_0(\dot{x}(t)/\theta) = -\mathrm{D}E(t,x).$$

Many of the results of this chapter have been published in [SKTO09], although that work considered only the cases in which the Hessian of E was either identically zero or a constant, symmetric, and positive-definite operator. The remarks on the mean-field approximation of the Koslowski–Cuitiño–Ortiz phase-field model in section 6.5 owe a great deal to discussions with the other authors of the joint paper [SKTO09], and their contributions are gratefully acknowledged.

6.2 Notation and Set-Up of the Problem

The following assumptions on the energetic and dissipation potentials will be hold for the rest of this chapter unless otherwise noted:

- The energetic potential $E\colon [0,T] \times \mathbb{R}^n \to \mathbb{R}$ is assumed to be bounded below, smooth in space with all derivatives uniformly bounded, and such that $(t,x) \mapsto \mathrm{D}E(t,x)$ is uniformly Lipschitz. It is also assumed that E is convex, and hence that the Hessian of E is a non-negative operator.

- The dissipation potential $\Psi = \chi_{\mathscr{E}}^\star\colon \mathbb{R}^n \to \mathbb{R}$ is continuous, positive-definite and homogeneous of degree one. There will be cause to con-

sider degenerate dissipation potentials, but this will be pointed out explicitly at the time.

The prototypical situation is the case in which E is a time-dependent quadratic form:

$$E(t, x) := \frac{1}{2}\langle Ax, x\rangle - \langle \ell(t), x\rangle,$$

where $A\colon \mathbb{R}^n \to (\mathbb{R}^n)^*$ is symmetric and positive-definite and $\ell\colon [0, T] \to (\mathbb{R}^n)^*$ is uniformly Lipschitz, i.e. $\ell \in W^{1,\infty}([0, T]; (\mathbb{R}^n)^*)$. A physical setting for this example is to consider a block resting on a rough table and restrained by some springs. The roughness of the block/table interface is encoded in the dissipation potential Ψ; the operator ("matrix") A is the elasticity matrix for the restraining springs; it is assumed that inertial effects can be neglected. The interest is in what happens if the dissipation is stronger than the load imparted by the springs (so the block is in a stable state and ought not to move), and then the table is shaken randomly. Under the influence of the shaking (the heat bath), does the block move, and can a quantitative description of its dynamics be given? The analysis of this chapter answers both questions in the affirmative.

6.3 Effective (Dual) Dissipation Potential

The central object in the analysis of continuous-time limit of the thermalized gradient descent Markov chain X is the effective dual dissipation potential alluded to at the start of this chapter. To motivate its introduction and definition, it helps to first make an explicit calculation of the statistics for the Markov chain and appeal to some heuristics.

6.3.1 Heuristics and Calculation of Moments

For simplicity, suppose that E is of the prototypical quadratic type and consider the following calculation for the conditional expectation of the next

state of the Markov chain $X^{(P)}$ given the current state:

$$\begin{aligned}
&\mathbb{E}\left[X_{i+1}^{(P)} \,\middle|\, X_i^{(P)} = x_i\right] \\
&\quad= \int_{\mathbb{R}^n} x_{i+1} \rho_{i+1}^{h,\varepsilon}(x_{i+1} \mid x_i)\,\mathrm{d}x_{i+1} \\
&\quad= \frac{\displaystyle\int_{\mathbb{R}^n} x_{i+1} \exp\big(-(E(t_{i+1},x_{i+1}) - E(t_i,x_i) + \Psi(x_{i+1}-x_i))/\varepsilon\big)\,\mathrm{d}x_{i+1}}{\displaystyle\int_{\mathbb{R}^n} \exp\big(-(E(t_{i+1},x_{i+1}) - E(t_i,x_i) + \Psi(x_{i+1}-x_i))/\varepsilon\big)\,\mathrm{d}x_{i+1}} \\
&\quad= \frac{\displaystyle\int_{\mathbb{R}^n} x_{i+1} \exp\big(-(E(t_{i+1},x_{i+1}) + \Psi(x_{i+1}-x_i))/\varepsilon\big)\,\mathrm{d}x_{i+1}}{\displaystyle\int_{\mathbb{R}^n} \exp\big(-(E(t_{i+1},x_{i+1}) + \Psi(x_{i+1}-x_i))/\varepsilon\big)\,\mathrm{d}x_{i+1}}
\end{aligned}$$

setting $y := x_{i+1} - x_i$ yields

$$\begin{aligned}
&= x_i + \frac{\displaystyle\int_{\mathbb{R}^n} y \exp\big(-(E(t_{i+1},x_i+y) + \Psi(y))/\varepsilon\big)\,\mathrm{d}y}{\displaystyle\int_{\mathbb{R}^n} \exp\big(-(E(t_{i+1},x_i+y) + \Psi(y))/\varepsilon\big)\,\mathrm{d}y} \\
&= x_i + \frac{\displaystyle\int_{\mathbb{R}^n} y \exp\big(-(\langle Ax_i - \ell(t_{i+1}), y\rangle + \tfrac{1}{2}\langle Ay, y\rangle + \Psi(y))/\varepsilon\big)\,\mathrm{d}y}{\displaystyle\int_{\mathbb{R}^n} \exp\big(-(\langle Ax_i - \ell(t_{i+1}), y\rangle + \tfrac{1}{2}\langle Ay, y\rangle + \Psi(y))/\varepsilon\big)\,\mathrm{d}y}
\end{aligned}$$

and setting $z := y/\varepsilon = (x_{i+1} - x_i)/\varepsilon$ yields

$$= x_i + \varepsilon \frac{\displaystyle\int_{\mathbb{R}^n} z \exp\big(-(\langle Ax_i - \ell(t_{i+1}), z\rangle + \tfrac{\varepsilon}{2}\langle Az, z\rangle + \Psi(z))\big)\,\mathrm{d}z}{\displaystyle\int_{\mathbb{R}^n} \exp\big(-(\langle Ax_i - \ell(t_{i+1}), z\rangle + \tfrac{\varepsilon}{2}\langle Az, z\rangle + \Psi(z))\big)\,\mathrm{d}z}.$$

Let

$$\mathcal{F}_\varepsilon^\star(w) := \log \int_{\mathbb{R}^n} \exp\big(-(\langle w, z\rangle + \tfrac{\varepsilon}{2}\langle Az, z\rangle + \Psi(z))\big)\,\mathrm{d}z.$$

Then the result of the above calculation may be summarized as

$$\mathbb{E}\left[\Delta X_{i+1}^{(P)} \,\middle|\, X_i^{(P)} = x_i\right] = -\varepsilon \mathrm{D}\mathcal{F}_\varepsilon^\star(w)|_{w=Ax_i - \ell(t_{i+1})},$$

i.e.

$$\mathbb{E}\left[\Delta X_{i+1}^{(P)} \,\middle|\, X_i^{(P)} = x_i\right] = -\varepsilon \mathrm{D}\mathcal{F}_\varepsilon^\star(\mathrm{D}E(t_{i+1}, x_i)).$$

Furthermore, the same change of variables $z := (x_{i+1} - x_i)/\varepsilon$ gives an estimate for the p^{th} moment of the increments of the Markov chain:

$$\begin{aligned}&\mathbb{E}\left[\left|\Delta X_{i+1}^{(P)}\right|^p \,\middle|\, X_i^{(P)} = x_i\right] \\ &\quad\leq \varepsilon^p \frac{\displaystyle\int_{\mathbb{R}^n} |z|^p \exp\left(-\left(\langle Ax_i - \ell(t_{i+1}), z\rangle + \tfrac{\varepsilon}{2}\langle Az, z\rangle + \Psi(z)\right)\right) \mathrm{d}z}{\displaystyle\int_{\mathbb{R}^n} \exp\left(-\left(\langle Ax_i - \ell(t_{i+1}), z\rangle + \tfrac{\varepsilon}{2}\langle Az, z\rangle + \Psi(z)\right)\right) \mathrm{d}z},\end{aligned}$$

For later reference, these calculations are summarized in the following lemma:

Lemma 6.3.1. *Let $E(t,x) = \frac{1}{2}\langle Ax, x\rangle - \langle \ell(t), x\rangle$ with $A\colon \mathbb{R}^n \to (\mathbb{R}^n)^*$ symmetric and non-negative. Suppose also that $\Psi = \chi_{\mathscr{E}}^\star\colon \mathbb{R}^n \to [0, +\infty)$ is 1-homogeneous and non-degenerate. Let $X^{(P)}$ denote the Markov chain in E and Ψ on a partition P of $[0,T]$. Then*

$$\mathbb{E}\left[\Delta X_{i+1}^{(P)} \,\middle|\, X_i^{(P)} = x_i\right] = -\varepsilon \mathrm{D}\mathcal{F}_\varepsilon^\star(Ax_i - \ell(t_{i+1})).$$

and, for $p > 0$,

$$\begin{aligned}&\mathbb{E}\left[\left|\Delta X_{i+1}^{(P)}\right|^p \,\middle|\, X_i^{(P)} = x_i\right] \\ &\quad\leq \varepsilon^p \frac{\displaystyle\int_{\mathbb{R}^n} |z|^p \exp\left(-\left(\langle Ax_i - \ell(t_{i+1}), z\rangle + \tfrac{\varepsilon}{2}\langle Az, z\rangle + \Psi(z)\right)\right) \mathrm{d}z}{\displaystyle\int_{\mathbb{R}^n} \exp\left(-\left(\langle Ax_i - \ell(t_{i+1}), z\rangle + \tfrac{\varepsilon}{2}\langle Az, z\rangle + \Psi(z)\right)\right) \mathrm{d}z}.\end{aligned}$$

The above calculations go also through, at least formally, even if E is not a quadratic form. The non-Ψ terms in the exponent are the Taylor series expansion of $E(t_{i+1}, x_{i+1}) - E(t_{i+1}, x_i)$ and, thus, the corresponding expression for $\mathcal{F}_\varepsilon^\star$ is

$$\mathcal{F}_\varepsilon^\star(w) := \log \int_{\mathbb{R}^n} \exp\left(-\left(\langle w, z\rangle + \sum_{k=2}^{\infty} \tfrac{\varepsilon^{k-1}}{k!}\langle \mathrm{D}^k E(t_{i+1}, x_i), z^{\otimes k}\rangle + \Psi(z)\right)\right) \mathrm{d}z.$$

Note, however, that in none of these expressions does the time increment appear explicitly. This is, of course, to be expected, since the original evolution was a rate-independent one. Therefore, in order to obtain a Markov chain that takes any account of time, it will be necessary to use the flexibility built into definition 5.3.6 and take ε to be proportional to the time

step.

The potential $\mathcal{F}_\varepsilon^\star \colon (\mathbb{R}^n)^* \to [0, +\infty]$ clearly encodes a great deal of information about the Markov chain X. Most of the terms in the exponent of $\mathcal{F}_\varepsilon^\star$ are of order ε or higher, and so can reasonably be expected to have no influence in the limit as $[\![P]\!]$ tends to zero in proportion to ε. The limiting dynamics of the Markov chain are expected to be controlled by an *effective dual dissipation potential* $\mathcal{F}_0^\star$, which is $\mathcal{F}_\varepsilon^\star$ with these higher-order terms omitted. Furthermore, the strong similarity to the Euler method for an ordinary differential equation and the fact that the variances are of order $\varepsilon^2 \ll \varepsilon$ suggest that the limiting evolution takes the form of a deterministic ordinary differential equation

$$\dot{x}(t) = -\theta \mathrm{D}\mathcal{F}_0^\star(\mathrm{D}E(t, x(t))),$$

where $\theta = \varepsilon_i / \Delta t_i$.

6.3.2 Effective (Dual) Dissipation Potential

Having gone through some motivational calculations and heuristics, the definition of the effective dual dissipation potential is as follows:

Definition 6.3.2. Given $\Psi \colon \mathbb{R}^n \to [0, +\infty)$ homogeneous of degree one, define the associated *effective dual dissipation potential* $\mathcal{F}_0^\star \colon (\mathbb{R}^n)^* \to [0, +\infty]$ by

$$\mathcal{F}_0^\star(w) := \log \int_{\mathbb{R}^n} \exp(-(\langle w, z \rangle + \Psi(z)))\, \mathrm{d}z. \tag{6.3.1}$$

The associated *effective dissipation potential* $\mathcal{F}_0 \colon \mathbb{R}^n \to [0, +\infty]$ is defined by convex conjugation: $\mathcal{F}_0 := (\mathcal{F}_0^\star)^\star$, i.e.

$$\mathcal{F}_0(x) := \sup\{\langle w, x \rangle - \mathcal{F}_0^\star(w) \mid w \in (\mathbb{R}^n)^*\}. \tag{6.3.2}$$

Note that $\mathcal{F}_0$ and $\mathcal{F}_0^\star$ are objects that are entirely intrinsic to the dissipation, and that the energetic structure plays no part. They are determined entirely by Ψ and the duality between $\mathbb{R}^n$ and $(\mathbb{R}^n)^*$; put another way, they are determined by the geometry of the elastic region $\mathscr{E}$. (A correct, but not necessarily enlightening, description of $\mathcal{F}_0^\star$ is that it is the Laplace transform of the exponential of the negative of the convex conjugate of the convex characteristic function of $\mathscr{E}$!) $\mathcal{F}_0^\star$ appears to be a novel contribution to the literature, having been introduced in [SKTO09].

As noted above, heuristics suggest that the limiting evolution takes the form

$$\dot{x}(t) = -\theta \mathrm{D}\mathcal{F}_0^\star(\mathrm{D}E(t, x(t))),$$

and rigorous justification of this limit will be provided in the next section. A standard convex conjugation argument (see theorem C.8) can be used to rephrase this claimed limiting evolution equation in terms of the convex conjugate $\mathcal{F}_0$ of $\mathcal{F}_0^\star$:

$$\mathrm{D}\mathcal{F}_0(-\dot{x}(t)/\theta) = \mathrm{D}E(t, x(t)).$$

In other words, the dual of $\mathcal{F}_0^\star$ acts as a dissipation potential for the effective (limiting) dynamics. Plots of $\mathcal{F}_0^\star$ and $\mathcal{F}_0$ in a simple case are given in figure 6.3.1 on page 102: $\mathcal{F}_0$ exhibits linear growth at infinity, but is smooth at the origin, and it is for these reasons that the thermalized gradient descent destroys the original rate-independence in a controlled way. The rôle of the temperature-like parameter θ is also interesting: if θ is large, then even moderate velocities are subject to rate-dependent dissipation, and so the thermalized and the original rate-independent dynamics are quite different; if θ is small, then "most" velocities are subject to a rate-independent dissipation, and so the thermalized dynamics and original rate-independent dynamics are quite similar.

It will sometimes be convenient to consider the "partition function" that corresponds to $\mathcal{F}_0^\star$, i.e.

$$\mathcal{Z}_0^\star(w) := \int_{\mathbb{R}^n} \exp(-(\langle w, z\rangle + \Psi(z)))\,\mathrm{d}z. \tag{6.3.3}$$

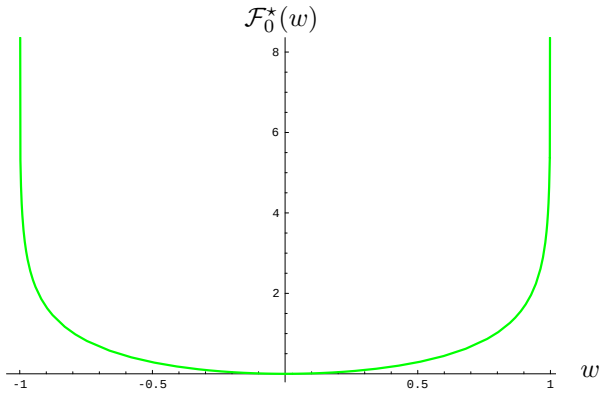
Using the homogeneity of the exponent in the integrand, (6.3.1) can be re-written as an integral over the Euclidean unit sphere $\mathbb{S}^{n-1} \subset \mathbb{R}^n$:

$$\mathcal{F}_0^\star(w) = \log \int_{\mathbb{S}^{n-1}} \frac{(n-1)!}{(\langle w, \omega\rangle + \Psi(\omega))^n}\,\mathrm{d}\mathscr{H}^{n-1}(\omega), \tag{6.3.4}$$

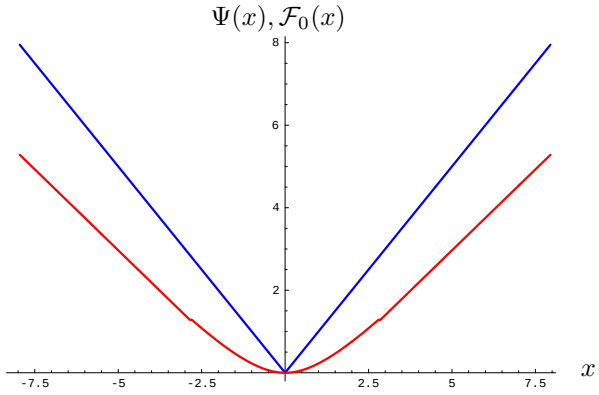
and similarly for $\mathcal{Z}_0^\star$.

Theorem 6.3.3. *Suppose that $\Psi = \chi_{\mathscr{E}}^\star$ for some compact, convex set $\mathscr{E} \subsetneq (\mathbb{R}^n)^*$ having $0 \in \mathring{\mathscr{E}}$. Then $\mathcal{F}_0^\star \colon (\mathbb{R}^n)^* \to [0, +\infty]$ defined as in (6.3.1) satisfies*

1. *$\mathcal{F}_0^\star(w) > 0$ for all $w \in (\mathbb{R}^n)^*$;*



(a) In green, the effective dual dissipation potential $\mathcal{F}_0^\star$.



(b) In blue, the original dissipation potential Ψ. In red, the effective dissipation potential $\mathcal{F}_0$: note the linear growth at infinity and smoothness near the origin.

Figure 6.3.1: The effective (dual) dissipation potential in dimension one, with dissipation potential $\Psi(x) := |x|$.

2. $\mathcal{F}_0^\star(w) < +\infty \iff -w \in \mathring{\mathscr{E}}$;

3. $\mathcal{F}_0^\star$ *is convex on* $(\mathbb{R}^n)^*$;

4. $\mathcal{F}_0^\star$ *is smooth on* $-\mathring{\mathscr{E}}$;

5. $\mathcal{F}_0^\star(w)$ *and* $|\mathrm{D}\mathcal{F}_0^\star(w)| \to +\infty$ *as* $-w \to \partial\mathscr{E}$.

The proof of this result may be found on page 128. The most important property of $\mathcal{F}_0^\star$ is its convexity: convexity/monotonicity properties are essential in the proof of theorem 6.4.2.

6.3.3 $\mathcal{Z}_0^\star$ and $\mathcal{F}_0^\star$ in Some Special Cases

In some special cases of mathematical or physical interest, the effective dual dissipation potential $\mathcal{F}_0^\star$ can be calculated in closed form. The most accessible cases are those in which

1. the elastic region $\mathscr{E}$ is a rectangular box with faces perpendicular to the coordinate axes in $(\mathbb{R}^n)^*$, so the dissipative potential Ψ is a weighted ℓ^1 "Manhattan" norm of the form $\Psi(z) = \sum_{j=1}^n \sigma_j |z^j|$ for some weights $\sigma_1, \dots, \sigma_n > 0$;

2. the elastic region $\mathscr{E}$ is a Euclidean ball $\overline{\mathbb{B}_\sigma(0; |\cdot|_2)}$ about the origin in $(\mathbb{R}^n)^*$, so the dissipative potential Ψ is exactly σ times the usual Euclidean norm.

Pedagogically speaking, of course, investigation of these cases comes first and inspires the more general results like theorem 6.3.3.

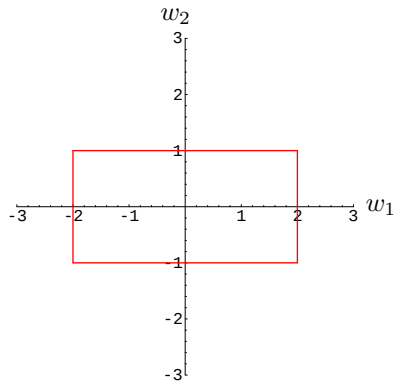
Proposition 6.3.4. *Given weights* $\sigma_1, \dots, \sigma_n > 0$, *let the elastic region* $\mathscr{E}$ *be the cuboid*

$$\mathscr{E} := \big\{ w = (w_1, \dots, w_n) \in (\mathbb{R}^n)^* \big| |w_i| \leq \sigma_i \text{ for } i = 1, \dots, n \big\},$$

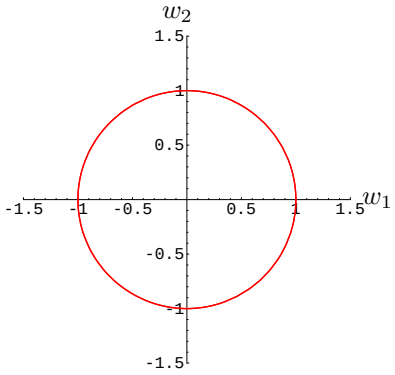
for which the associated dissipative potential $\Psi \colon \mathbb{R}^n \to \mathbb{R}$ *is the weighted* ℓ^1 *"Manhattan" norm* $\Psi(z) := \sigma_1 |z^1| + \cdots + \sigma_n |z^n|$. *Then, for* $w \in \mathscr{E}$, *up to gauge,*

$$\mathcal{Z}_0^\star(w) = \prod_{i=1}^n \frac{1}{\sigma_i^2 - |w_i|^2}, \tag{6.3.5}$$

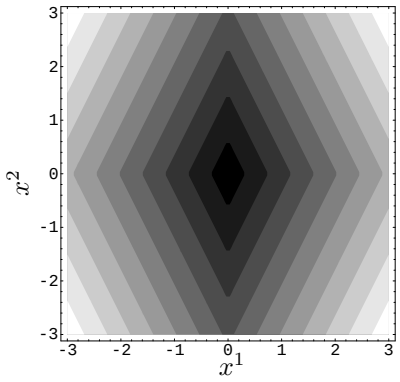
$$\mathcal{F}_0^\star(w) = -\sum_{i=1}^n \log\big(\sigma_i^2 - |w_i|^2\big). \tag{6.3.6}$$



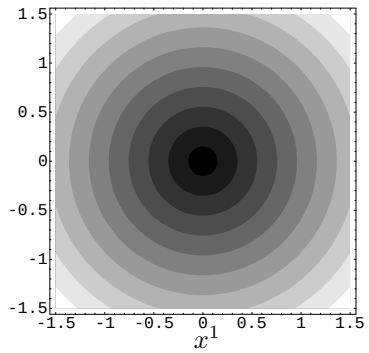
(a) Manhattan $\mathscr{E}$



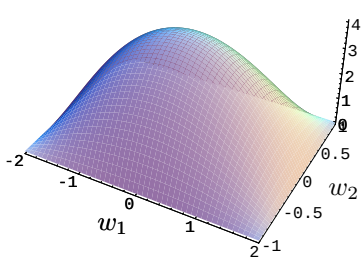
(b) Euclidean $\mathscr{E}$



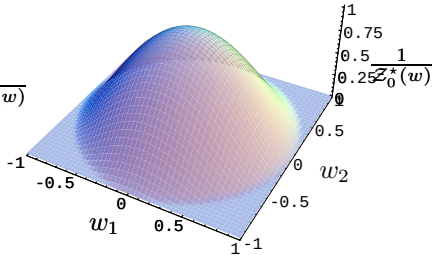
(c) Manhattan Ψ



(d) Euclidean Ψ



(e) Manhattan $1/\mathcal{Z}_0^\star$



(f) Euclidean $1/\mathcal{Z}_0^\star$

Figure 6.3.2: The elastic region $\mathscr{E}$, the associated dissipation potential Ψ and the effective dual dissipation potential $\mathcal{F}_0^\star$ in two cases: on the left, the case in which Ψ is a weighted Manhattan norm, cf. proposition 6.3.4; on the right, the case in which Ψ is the Euclidean norm, cf. proposition 6.3.5. The frontier $\partial\mathscr{E}$ is shown in red; the contour plots of Ψ assign darker colours to lower values. For clarity, $1/\mathcal{Z}_0^\star = e^{-\mathcal{F}_0^\star}$ is plotted.

Proof. Direct calculation using iterated integrals yields that

$$\begin{aligned}
\mathcal{Z}_0^\star(w) &= \int_{\mathbb{R}^n} \exp\big(-\big(\langle w, z\rangle + \Psi(z)\big)\big)\,\mathrm{d}z \\
&= \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} \exp\left(-\left(\sum_{i=1}^n w_i z^i + \sigma_i |z^i|\right)\right)\,\mathrm{d}z^1 \ldots \mathrm{d}z^n \\
&= \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} \prod_{i=1}^n \exp\big(-\big(w_i z^i + \sigma_i |z^i|\big)\big)\,\mathrm{d}z^1 \ldots \mathrm{d}z^n \\
&= \prod_{i=1}^n \int_{\mathbb{R}} \exp\big(-\big(w_i z^i + \sigma_i |z^i|\big)\big)\,\mathrm{d}z^i \\
&= \prod_{i=1}^n \frac{2\sigma_i}{\sigma_i^2 - |w_i|^2}.
\end{aligned}$$

Eliminating gauge constants and using the relation $\mathcal{F}_0^\star = \log \mathcal{Z}_0^\star$ completes the proof. ∎

Proposition 6.3.5. *Let the elastic region $\mathscr{E}$ be the Euclidean ball*

$$\mathscr{E} := \overline{\mathbb{B}_\sigma(0; |\cdot|)},$$

for which the associated dissipative potential $\Psi\colon \mathbb{R}^n \to \mathbb{R}$ is a multiple of the Euclidean norm, $\Psi := \sigma|\cdot|$. Then, for $w \in \mathscr{E}$, up to gauge,

$$\mathcal{Z}_0^\star(w) = \big(\sigma^2 - |w|^2\big)^{-\frac{n+1}{2}}, \tag{6.3.7}$$

$$\mathcal{F}_0^\star(w) = -\frac{n+1}{2} \log\big(\sigma^2 - |w|^2\big). \tag{6.3.8}$$

Proof. Direct calculation using spherical polar coordinates yields that

$$\begin{aligned}
\mathcal{Z}_0^\star(w) &= \int_{\mathbb{R}^n} \exp\big(-\big(\langle w, z\rangle + \Psi(z)\big)\big)\,\mathrm{d}z \\
&= \int_{\mathbb{S}^{n-1}} \left(\int_0^{+\infty} \exp\big(-r\big(\langle w, \omega\rangle + \sigma\big)\big) r^{n-1}\,\mathrm{d}r\right) \mathrm{d}\mathscr{H}^{n-1}(\omega) \\
&= \int_{\mathbb{S}^{n-1}} \frac{(n-1)!}{(\langle w, \omega\rangle + \sigma)^n}\,\mathrm{d}\mathscr{H}^{n-1}(\omega)
\end{aligned}$$

and, since the integrand is constant on "lines of latitude" on $\mathbb{S}^{n-1}$ with w regarded as the "north pole",

$$\begin{aligned} &= \int_0^\pi \frac{(n-1)!\mathscr{H}^{n-2}(\mathbb{S}^{n-2}_{\sin\theta})}{(|w|\cos\theta+\sigma)^n}\,\mathrm{d}\theta \\ &= \frac{2\pi^{n/2}\sigma\Gamma(n)}{\Gamma(n/2)}\big(\sigma^2-|w|^2\big)^{-\frac{n+1}{2}}, \end{aligned}$$

where Γ denotes the Gamma function

$$\Gamma(z) := \int_0^{+\infty} t^{z-1}e^{-t}\,\mathrm{d}t \text{ for } z\in\mathbb{C}, \mathrm{Re}(z)>0.$$

Eliminating gauge constants and using the relation $\mathcal{F}_0^\star = \log \mathcal{Z}_0^\star$ completes the proof. ■

6.4 Convergence Theorems in $\mathbb{R}^n$

6.4.1 Flat Energetic Potential

The simplest case of study is that in which the energetic potential E is determined purely by an applied external load $\ell\colon [0,T]\to(\mathbb{R}^n)^*$, i.e.

$$E(t,x) := -\langle \ell(t), x\rangle.$$

This is the case of a flat energetic potential, since the Hessian of E vanishes (as do all higher-order derivatives). To ensure the existence and uniqueness of solutions to the rate-independent problem and the limiting ordinary differential equation, ℓ will be assumed to be uniformly Lipschitz, i.e. $\ell\in W^{1,\infty}([0,T];(\mathbb{R}^n)^*)$. The dissipation potential Ψ is assumed to be continuous, 1-homogeneous and non-degenerate, i.e. $\Psi=\chi^\star_{\mathscr{E}}$ for some compact, convex set $\mathscr{E}\subsetneq(\mathbb{R}^n)^*$ having 0 in its interior. The effective dual dissipation potential $\mathcal{F}_0^\star\colon(\mathbb{R}^n)^*\to[0,+\infty]$ is defined as in (6.3.1); note that

$$\mathcal{F}_0^\star(\mathrm{D}E(t,x)) = \log\int_{\mathbb{R}^n} e^{\langle \ell(t),z\rangle-\Psi(z)}\,\mathrm{d}z$$

and that this expression, since it is independent of x, is convex in x. Also, the approximate dual dissipation potential $\mathcal{F}_\varepsilon^\star$ is independent of ε, so $\mathcal{F}_0^\star=\mathcal{F}_\varepsilon^\star$,

and so

$$\mathbb{E}\left[\Delta X_{i+1}^{(P)} \,\middle|\, X_i^{(P)} = x_i\right] = -\varepsilon \mathrm{D}\mathcal{F}_0^\star(\mathrm{D}E(t_{i+1}, x_i)).$$

Obviously, this expression makes sense only if $\mathrm{D}E(t,x)$ lies in the effective domain of $\mathcal{F}_0^\star$, i.e. $\ell(t_{i+1}) \in \mathring{\mathscr{E}}$. Furthermore, since variance estimates for $\Delta X_{i+1}^{(P)}$ are also given in terms of $\mathcal{F}_0^\star$ and blow up as the applied load approaches the frontier of the elastic region, it is sensible to impose a *uniform stability criterion* on ℓ:

$$\inf_{t\in[0,T]} \operatorname{dist}(\ell(t), \partial\mathscr{E}) = \delta > 0 \tag{6.4.1}$$

Note that (6.4.1) is equivalent to the statement that

$$\inf_{t\in[0,T]} \inf_{|z|=1} \big(\Psi(z) - \ell(t)\cdot z\big) > 0.$$

Theorem 6.4.1 (Flat case)**.** *Let E, Ψ be as defined at the start of this subsection, and suppose that ℓ satisfies the uniform stability criterion* (6.4.1)*. Let $\theta > 0$. Then, as $[\![P]\!] \to 0$, the piecewise constant càdlàg interpolants of $X^{(P)}$ with $\varepsilon_i = \theta\Delta t_i$ converge in probability in the uniform norm to the differentiable, deterministic process $y = y^\theta\colon [0,T] \to \mathbb{R}^n$ satisfying*

$$\dot{y} = -\theta \mathrm{D}\mathcal{F}_0^\star(\mathrm{D}E(t, y(t))), \tag{6.4.2}$$

with the same initial condition $X_0^{(P)} = y(0) = x_0$. More precisely, for any $T > 0$, $\lambda > 0$, there exists a constant $C \geq 0$ such that, for all small enough $[\![P]\!]$,

$$\mathbb{P}\left[\sup_{t\in[0,T]} \left|X^{(P)}(t) - y(t)\right| \geq \lambda\right] \leq C[\![P]\!]. \tag{6.4.3}$$

A fortiori*, the limiting probability measure on path space is a Dirac measure supported on y^θ.*

The proof of theorem 6.4.1 is deferred to the end of the chapter; see page 130.

Since the Markov chain $X^{(P)}$ was defined using a minimization principle, it is natural to ask whether the limiting evolution (6.4.2) also satisfies a minimization principle. The answer is affirmative: it follows from Stefanelli's generalization of the Brézis–Ekeland theorem (apply [Ste08, theorem 1.2] with $\psi = \mathcal{F}_0$ and $\phi = 0$) that the solution to (6.4.2) for $\theta = 1$, say, is the

unique minimizer (with value 0) of the Brézis–Ekeland-type functional

$$\mathcal{J}[u] := \max\left\{0, \int_0^T \mathcal{F}_0(\dot{u}(t)) + \mathcal{F}_0^\star(\ell(t)) - \langle \ell(t), \dot{u}(t)\rangle \,\mathrm{d}t\right\} + |u(0) - x_0|^2. \tag{6.4.4}$$

In view of the convex duality theorem (theorem C.8), one interpretation of the integrand in (6.4.4) is that it quantitatively measures the failure of the relation (6.4.2): assuming that Ψ (and hence $\mathcal{F}_0^\star$ and $\mathcal{F}_0$) is an even function, in which case $\mathrm{D}\mathcal{F}_0$ is odd, (6.4.2) is equivalent to

$$\mathrm{D}\mathcal{F}_0(\dot{x}(t)) = \ell(t).$$

Note, however, that the action $\mathcal{J}$ in (6.4.4) is inequivalent to the path integral exponent $\mathcal{I}$ in (5.2.7); integration by parts yields that

$$\mathcal{I}[u] = \int_0^T \big(\Psi(\dot{u}(t)) - \langle \ell(t), \dot{u}(t)\rangle\big) \,\mathrm{d}t.$$

Since, by assumption, $z \mapsto \Psi(z) - \langle \ell(t), z\rangle$ is positive-definite, $\mathcal{I}[u] \geq 0$ and

$$\mathcal{I}[u] = 0 \iff \langle \ell(t), \dot{u}(t)\rangle = \Psi(\dot{u}(t)) \text{ for } \lambda\text{-almost all } t \in [0,T].$$

This condition is not generally satisfied by minimizers of $\mathcal{J}$: for $n = 1$ with $\Psi(z) = \sigma|z|$ and $\theta = 1$, the minimizer x of $\mathcal{J}$ (i.e. the solution of (6.4.2)) satisfies

$$\dot{x}(t) = \frac{2\ell^\sharp(t)}{\sigma^2 - |\ell(t)|^2}$$

but

$$\begin{aligned}
\langle \ell(t), \dot{x}(t)\rangle &= \langle \ell(t), -\mathrm{D}\mathcal{F}_0^\star(-\ell(t))\rangle \\
&= \frac{2|\ell(t)|^2}{\sigma^2 - |\ell(t)|^2} \\
&\neq \frac{2\sigma|\ell(t)|}{\sigma^2 - |\ell(t)|^2} \\
&= \Psi(\dot{x}(t)).
\end{aligned}$$

Note also that the deterministic limiting system responds almost linearly to the applied load ℓ when ℓ is small. However, it exhibits a strongly nonlinear yielding response to applied loads close to the yield surface $\partial\mathscr{E}$. The

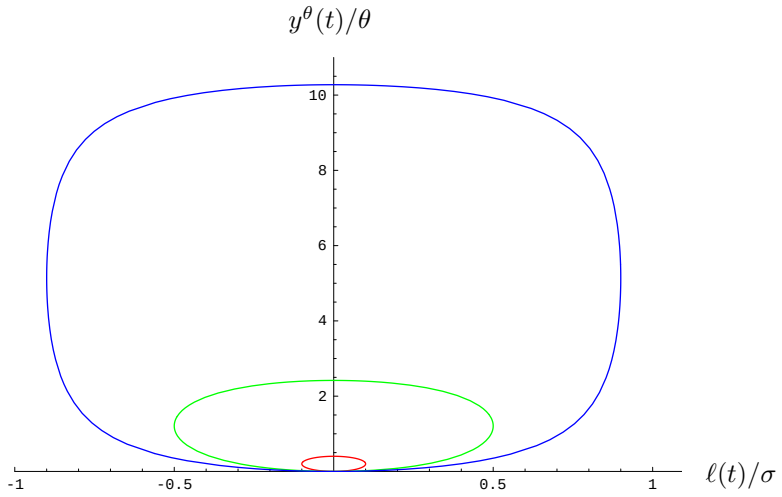


Figure 6.4.1: The nonlinear yielding response of the effective dynamics y^θ given by (6.4.2) to an applied cyclic load $\ell(t) = \ell_0 \sin t$ for $\ell_0 = 0.1\sigma$ (red), 0.5σ (green) and 0.9σ (blue), where $\Psi(v) = \sigma|v|$.

effect of the "temperature-like" parameter θ is to soften the system response in proportion to θ as θ increases. See figure 6.4.1 on page 109 for an illustration of the response of y to a cyclic applied load ℓ in the one-dimensional case.

6.4.2 Convex Energetic Potential

In this subsection, the energetic potential is assumed to be convex, along with a further structural assumption that will be addressed shortly. As before, the dissipation potential $\Psi = \chi^\star_{\mathscr{E}}$ is assumed to be 1-homogeneous and positive-definite.

It is an unfortunate æsthetic deficiency in the results obtained to date that the order of certain error terms in the proof of theorem 6.4.2 has not yet been controlled properly. Instead, a monotonicity assumption is used to ensure that these terms have the right sign regardless of their magnitude. The requisite assumption is that

$$\text{for all } t \in [0,T], x \mapsto \mathcal{F}_0^\star(\mathrm{D}E(t,x)) \text{ is convex.} \tag{6.4.5}$$

By Kachurovsiĭ's theorem, this is equivalent to the statement that

$$x \mapsto -\mathrm{D}\mathcal{F}_0^\star(\mathrm{D}E(t,x))$$

is a monotone vector field for every $t \in [0,T]$. This is a non-trivial assumption even if E is strictly convex, as the example illustrated in figure 6.4.2 on page 111 shows. Note that the assumption (6.4.5) includes an additional "hidden" assumption that the set $\mathcal{S}(t) \subseteq \mathbb{R}^n$ of stable states is convex for every $t \in [0,T]$.

A prototypical example for this subsection is

$$E(t,x) := \frac{1}{2}\langle Ax, x\rangle - \langle \ell(t), x\rangle,$$

where $\ell \in W^{1,\infty}([0,T];(\mathbb{R}^n)^*)$ as before, and $A\colon \mathbb{R}^n \to (\mathbb{R}^n)^*$ is symmetric and non-negative. The assumptions on the "spring constant" A imply that the stable region at time $t \in [0,T]$ is given by

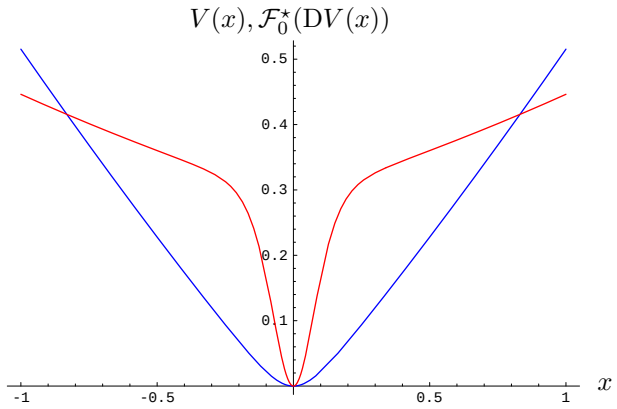
$$\mathcal{S}(t) = A^{-1}(\ell(t) - \mathscr{E})$$

and is convex and closed for every t; if A is positive-definite, then $\mathcal{S}(t)$ is also bounded, and hence compact. Also, since $\mathrm{D}E$ is an affine function and the composition of convex function with an affine one always yields a convex function [Roc70, section 4], (6.4.5) is satisfied.

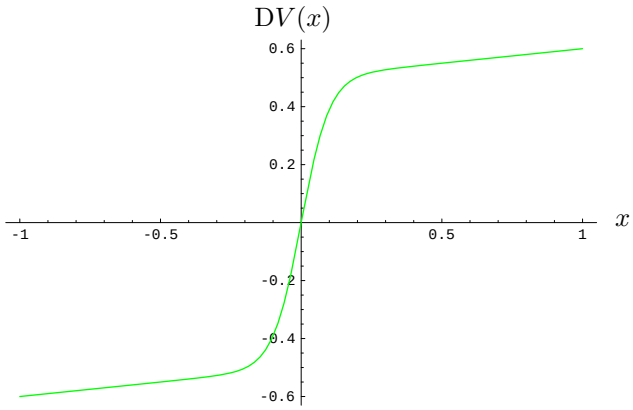
The question of what, if any, constraints need to be imposed on the time-dependency of E is more subtle in this case. As before, the problem is that all the estimates for the moments of the increments $\Delta X_{i+1}^{(P)}$ blow up as $-\mathrm{D}E(t_i, X_i^{(P)})$ approaches the yield surface $\partial\mathscr{E}$. In the course of proving theorem 6.4.2 below, it becomes clear that the situation to be avoided can be expressed neatly in terms of the proposed limiting deterministic process and the effective dual dissipation potential. The desideratum is a kind of "finite energy criterion":

$$\left.\begin{array}{l}\mathcal{T} \text{ an interval of time starting at } 0,\\ y(0) \text{ such that } -\mathrm{D}E(0,y(0)) \in \mathring{\mathscr{E}},\\ \quad \dot{y} = -\theta\mathrm{D}\mathcal{F}_0^\star(\mathrm{D}E(t,y)) \text{ on } \mathcal{T}\end{array}\right\} \implies \sup_{t\in\mathcal{T}} \mathcal{F}_0^\star(\mathrm{D}E(t,y(t))) < +\infty. \tag{6.4.6}$$

Admittedly, this is somewhat implicit, but appears to be unavoidable if energies that have neither identically zero nor positive-definite Hessian are



(a) Comparison of V (convex, blue) and $\mathcal{F}_0^\star \circ \mathrm{D}V$ (non-convex, red).



(b) The potential gradient $\mathrm{D}V$ (green).

Figure 6.4.2: An example of a strictly convex potential V for which $\mathcal{F}_0^\star \circ \mathrm{D}V$ is not convex. In dimension 1, consider $\Psi(x) := |x|$ and $V(x) := \frac{x^2}{20} + \frac{1}{20}\log\cosh(10x)$. In this case, $\mathcal{F}_0^\star(w) = -\log(1-w^2)$ and $\mathrm{D}V(x) = \frac{1}{2}\tanh(10x) + \frac{1}{10}x$. The composition $x \mapsto \mathcal{F}_0^\star(\mathrm{D}V(x))$ is evidently non-convex, although it is quasiconvex (i.e. it has convex sublevel sets).

to be considered. If the second-and-higher-order derivatives of E do vanish, then the above criterion is the same as the stability criterion of the previous subsection. If E is of the prototypical quadratic form and A is positive definite, then the criterion above always holds whenever $y(0) \in \mathring{\mathcal{S}}(0)$.

Theorem 6.4.2. *Suppose that E, Ψ satisfy the usual hypotheses and that* (6.4.5) *and* (6.4.6) *hold. Let $\theta > 0$. Then, as $[\![P]\!] \to 0$, the piecewise constant càdlàg interpolants of $X^{(P)}$ with $\varepsilon_i = \theta \Delta t_i$ converge in probability in the uniform norm to the differentiable, deterministic process $y = y^\theta \colon [0,T] \to \mathbb{R}^n$ satisfying*

$$\dot{y}(t) = -\theta \mathrm{D}\mathcal{F}_0^\star(\mathrm{D}E(t, y(t))), \tag{6.4.7}$$

with the same initial condition $X_0 = y(0) = x_0 \in \mathring{\mathcal{S}}(0)$. More precisely, for any $T > 0$, $\lambda > 0$, there exists a constant $C \geq 0$ such that, for all small enough $[\![P]\!]$,

$$\mathbb{P}\left[\sup_{t\in[0,T]} \left|X^{(P)}(t) - y^\theta(t)\right| \geq \lambda\right] \leq C[\![P]\!]^{1/2}. \tag{6.4.8}$$

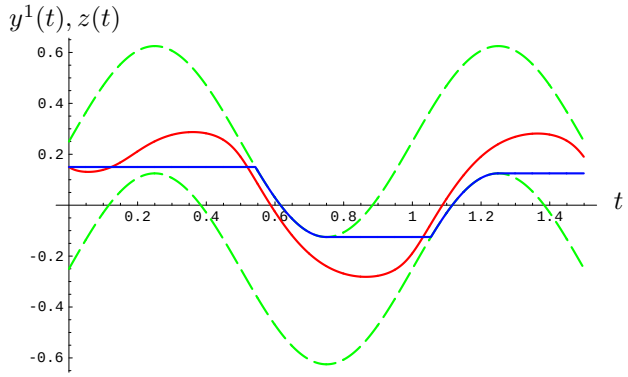
A fortiori*, the limiting probability measure on path space is a Dirac measure supported on y^θ.*

Proof. The claim follows from the standard $O([\![P]\!])$ error bound for Euler schemes, the $O(\varepsilon^{1/2})$ estimate of lemma 6.7.4, and the $O([\![P]\!])$ estimate of lemma 6.7.5. ■

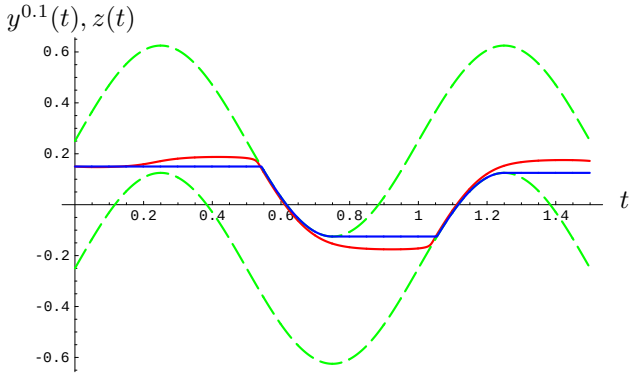
An illustrative comparison of the original rate-independent evolution and the effect of the heat bath is given in figure 6.4.3 on page 113.

In the case that $E(t,x) = V(x) - \langle \ell(t), x\rangle$ with all the usual assumptions on V and ℓ, a Brézis–Ekeland-type variational principle for the limiting thermalized dynamics can be obtained with ease. To simplify the notation a little, take $\theta = 1$ and assume that Ψ is even so that $\mathrm{D}\mathcal{F}_0$ is odd. The initial value problem for the limiting thermalized dynamics is then a Colli–Visintin-type doubly nonlinear problem:

$$\begin{cases} \mathrm{D}\mathcal{F}_0(\dot{x}(t)) + \mathrm{D}V(x(t)) = \ell(t), \\ x(0) = x_0. \end{cases}$$



(a) "Hot" thermalized gradient descent with $\theta = 1$.



(b) "Cold" thermalized gradient descent with $\theta = 0.1$.

Figure 6.4.3: A comparison of the original rate-independent evolution and the effect of the heat bath. The original rate-independent process z is shown in blue and the heated process y^θ in red. The frontier of the stable region is shown with green dashes. Parameters: $\Psi(z) = 2|z|$, $E(t, x) = 4|x|^2 - \langle 3 \sin 2\pi t, x \rangle$, initial condition 0.15.

(Since, in this case, the energetic and dissipation potentials are smooth, classical derivatives and equalities can be used instead of subdifferentials and containments.) This problem is equivalent to the following auxiliary problem for $(x, y) \in W^{1,p}([0,T];\mathbb{R}^n) \times L^q([0,T];(\mathbb{R}^n)^*)$ (where $\frac{1}{p} + \frac{1}{q} = 1$):

$$\begin{cases} y(t) = \mathrm{D}\mathcal{F}_0(\dot{x}(t)), \\ y(t) + \mathrm{D}V(x(t)) = \ell(t), \\ x(0) = x_0. \end{cases}$$

It then follows as a special case of [Ste08, theorem 1.2] that (x, y) solves the auxiliary problem if, and only if, it minimizes (with value 0) the following functional:

$$\begin{aligned} \mathcal{J}[u,v] := \max\left\{0, \int_0^T \big(\mathcal{F}_0(\dot{u}) + \mathcal{F}_0^\star(v) - \langle \ell, \dot{u}\rangle\big) + V(u(T)) - V(x_0)\right\} \\ + \int_0^T \big(V(u) - V^\star(\ell - v) - \langle \ell - v, u\rangle\big) + |u(0) - x_0|^2. \end{aligned}$$

As noted in the previous chapter, the path integral exponent $\mathcal{I}$ in (5.2.7) is exactly the energy surplus of chapter 4:

$$\begin{aligned} \mathcal{I}[u] &= \mathrm{ES}(u, [0,T]) \\ &= E(T, u(T)) - E(0, u(0)) + \int_0^T \Psi(\mathrm{d}u) - \int_0^T (\partial_t E)(s, u(s))\,\mathrm{d}s. \end{aligned}$$

The original unthermalized rate-independent process is a minimizer for $\mathcal{I}$, although the thermalized process is not.

6.4.3 An Example with Degenerate Dissipation

The analysis so far has treated the case in which the energetic potential E is a strictly convex and the dissipation potential Ψ is non-degenerate. The effect of the non-degeneracy of Ψ has been to make the effective dual dissipation potential $\mathcal{F}_0^\star$ finite somewhere. On the other hand, if Ψ has a non-trivial nullspace, then $\mathcal{F}_0^\star$ may be infinite everywhere.

In general, and in particular in infinite-dimensional contexts, the zero set of Ψ may have quite complicated structure: the most that can be said for it is that it is a union of pointed convex cones. However, in the case in which the zero set of Ψ is a linear subspace, the method of Schur complementation

[Zha05] offers a way forward.

For definiteness, consider a quadratic energy $E(t,x) := \frac{1}{2}\langle A_0 x, x\rangle - \langle \ell(t), x\rangle$ with A_0 symmetric and non-negative — A_0 may have a non-trivial kernel, and so E may fail to be strictly convex. Let the dissipation potential $\Psi\colon \mathbb{R}^n \to [0,+\infty)$ be the weighted ℓ^1 "Manhattan" *seminorm*

$$\Psi(v) := \sum_{i=1}^{n'} \sigma_i |v_i|, \tag{6.4.9}$$

where $\sigma_i > 0$ for $i = 1, \ldots, n'$ and $n' \leq n$; $\Psi = \chi^\star_{\mathscr{E}}$ for the (interiorless, if $n' < n$) rectangular box

$$\mathscr{E} := \left\{ w \in (\mathbb{R}^n)^* \,\middle|\, \begin{array}{c} \text{for } 1 \leq i \leq n', |w_i| \leq \sigma_i, \\ \text{and for } n' < i \leq n, w_i = 0 \end{array} \right\}.$$

Obviously

$$\begin{aligned} \Psi(v) = 0 &\iff v \in \operatorname{span}\{e_{n'+1}, \ldots, e_n\} \\ &\iff v \perp \operatorname{span}\{e_1, \ldots, e_{n'}\}. \end{aligned}$$

Write $x \in \mathbb{R}^n$ as $(x', x'') \in \mathbb{R}^{n'} \oplus \mathbb{R}^{n-n'}$; similarly, write $\ell \in (\mathbb{R}^n)^*$ as $(\ell', \ell'') \in (\mathbb{R}^{n'})^* \oplus (\mathbb{R}^{n-n'})^*$. Write the affine operator $\mathrm{D}E(t,x) \equiv A_0 x - \ell(t)$ in block form as

$$A_0(x) - \ell(t) = \begin{pmatrix} A & B \\ B^* & C \end{pmatrix} \begin{pmatrix} x' \\ x'' \end{pmatrix} - \begin{pmatrix} \ell'(t) \\ \ell''(t) \end{pmatrix},$$

where $A\colon \mathbb{R}^{n'} \to (\mathbb{R}^{n'})^*$ is symmetric, $B\colon \mathbb{R}^{n-n'} \to (\mathbb{R}^{n'})^*$ with adjoint (transpose) $B^*\colon \mathbb{R}^{n'} \to (\mathbb{R}^{n-n'})^*$, and $C\colon \mathbb{R}^{n-n'} \to (\mathbb{R}^{n-n'})^*$ is symmetric and invertible. In this notation, the Schur complement of the affine operator $\mathrm{D}E(t,x) \equiv A_0 x - \ell(t)$ is given by $x' \mapsto \tilde{A}x' - \tilde{\ell}(t)$, where

$$\begin{aligned} \tilde{A} &:= A - BC^{-1}B^*\colon \mathbb{R}^{n'} \to (\mathbb{R}^{n'})^* \\ \tilde{\ell} &:= \ell' - BC^{-1}\ell'' \in (\mathbb{R}^{n'})^*. \end{aligned}$$

Hence, the x'-components (those which do experience friction) are, by virtue of the relationship $x'' = C^{-1}(\ell'' - B^* x')$, subject to the *reduced energetic potential* $E_{\mathrm{red}}\colon [0,T] \times \mathbb{R}^{n'} \to \mathbb{R}$ given in terms of the Schur complement

by

$$E_{\text{red}}(t, x') := \frac{1}{2}\langle \tilde{A}x', x'\rangle - \langle \tilde{l}(t), x'\rangle \tag{6.4.10}$$

Similarly, restriction to $\mathbb{R}^{n'}$ and its dual yields a reduced effective dual dissipation potential $\mathcal{F}^\star_{0,\text{red}} \colon (\mathbb{R}^{n'})^* \to [0, +\infty]$; $\mathcal{F}^\star_{0,\text{red}}$ is just the usual effective dual dissipation potential as generated by the weighted ℓ^1 norm on $\mathbb{R}^{n'}$, i.e., by proposition 6.3.4,

$$\mathcal{F}^\star_{0,\text{red}}(w_1, \ldots, w_{n'}) := -\sum_{i=1}^{n'} \log\left(\sigma_i^2 - |w_i|^2\right).$$

Note, however, that $\mathcal{F}^\star_{0,\text{red}}$ is not actually a "reduction" of some original $\mathcal{F}^\star_0 \colon (\mathbb{R}^n)^* \to [0, +\infty]$ — or, at least, the original $\mathcal{F}^\star_0$ might be identically $+\infty$. The trick that has been employed is one of reducing the energetic potential so that its derivative takes values in a subspace on which Ψ is non-degenerate. This trick offers a (formal) method of dimension reduction for certain infinite-dimensional problems.

Theorem 6.4.3 (Degenerate ℓ^1 dissipation)**.** *Let $X^{(P)}$ be the Markov chain in a quadratic energetic potential E and possibly degenerate ℓ^1 dissipation potential Ψ of the form* (6.4.9)*, and assume that the operator C in the the block decomposition of E is invertible. Then the projection $X'^{(P)}$ of $X^{(P)}$ onto the first n' components converges uniformly in probability to the solution of*

$$\frac{\mathrm{d}}{\mathrm{d}t}(x'(t)) = -\theta \mathrm{D}\mathcal{F}^\star_{0,\text{red}}(\mathrm{D}E_{\text{red}}(t, x'(t)))$$

and the other $n - n'$ components satisfy $x''(t) = C^{-1}(\ell''(t) - B^* x'(t))$.

In the next section, this dimension reduction trick will be formally applied to an evolution in an infinite-dimensional separable Hilbert space that experiences friction only on a finite-dimensional subspace.

6.5 Application: Andrade Creep

This section has more of a physical than a mathematical flavour, and mathematical rigour is sacrificed at a few points. Nonetheless, the overall argument is an important one since it provides additional justification for the validity of the interior-point regularization of the Moreau–Yosida approximation

as a model for the effect of a heat bath on a gradient descent. The non-rigorous steps lie solely in the reduction of an infinite-dimensional model to a one-dimensional one; what the theory developed above says about that one-dimensional model is entirely rigorous.

In 1910, Andrade [And10] [And14] reported that as a function of time, t, the creep deformation, y, of soft metals at constant temperature and applied stress can be described by a power law $y(t) \sim t^{1/3}$. Similar behavior has been observed in many classes of materials, including non-crystalline materials. In this section, Andrade's $t^{1/3}$ creep law emerges as the mean-field evolution of the Koslowski–Cuitiño–Ortiz model under the assumption of linear strain hardening.

The Koslowski–Cuitiño–Ortiz model is a phase-field model for dislocation dynamics in a single slip plane (thought of as the 2-dimensional torus $\mathbb{T}^2$), the state (i.e. phase field) is represented as an element u of the Sobolev space $H^{1/2}(\mathbb{T}^2; \mathbb{R})$: $u(x)$ represents the amount of slip at x in multiples of the Burgers vector. This model was introduced in [KCO02] and a mathematically rigorous analysis of the model in the line tension limit may be found in [GM05] [GM06]. The dissipation potential for this model is modeled as an ℓ^1 seminorm of the type considered earlier that takes account only of the values of the phase field at certain "obstacle sites"; this (formally) allows for the reduction from an infinite-dimensional setting to a finite-dimensional one. A mean-field approximation then reduces this (large) finite-dimensional evolution to a one-dimensional model, and it is this that exhibits Andrade creep under the additional assumption of linear strain hardening. It is possible to verify this procedure numerically; for the methodology and results, see [SKTO09].

In this setting, the energetic potential E will be regarded as a function of an externally applied load and the state, i.e. $E\colon (\mathbb{R}^n)^* \times \mathbb{R}^n \to \mathbb{R}$. It is easiest to write E partly in terms of the Fourier transform $\hat{u}\colon \mathbb{Z}^2 \to \mathbb{C}$ of u: define

$$\hat{u}(k) := \int_{\mathbb{T}^2} u(x) \exp(2\pi i x \cdot k) \, \mathrm{d}x$$

and let

$$E(s, u) := \sum_{k \in \mathbb{Z}^2 \setminus \{0\}} \frac{\mu b^2}{4} \frac{1}{1/K + d/2} |\hat{u}(k)|^2 - (s \cdot b) \int_{\mathbb{T}^2} u(x) \, \mathrm{d}x. \qquad (6.5.1)$$

In the above, $k \in \mathbb{Z}^2$ is the wavenumber and

$$K := \frac{1}{1-\nu}\sqrt{k_1^2 + k_2^2} - \frac{\nu}{1-\nu}\frac{k_2^2}{\sqrt{k_1^2+k_2^2}};$$

b is the Burgers vector, s is the applied shear stress, μ is the shear modulus, ν is Poisson's ratio and d is the interplanar distance. Since it does not include the 0^{th} Fourier coefficient, the first term in the energy E is invariant under addition of constants to the phase field: whenever $u_1 \colon \mathbb{T}^2 \to \mathbb{R}$ is a constant field,

$$E(s, u + u_1) = E(s, u) + (s \cdot b)u_1. \tag{6.5.2}$$

The obstacles are assumed to be disc-like subsets $\mathbb{B}_r(p_i) \subsetneq \mathbb{T}^2$ where $p_1, \ldots, p_n \in \mathbb{T}^2$ are the centers of the obstacles and $\mathbb{B}_r(p_i)$ is the disk with radius $0 < r \ll 1$ and center at p_i. Typically it will be assumed that the positions p_i are random and that the phase field u is constant within each obstacle, i.e. $u(x) = \xi_i$ if $x \in \mathbb{B}_r(p_i)$; attention will shortly be focused on this reduced field $\xi \colon \mathcal{O} \to \mathbb{R}$.

The dissipation functional $\Psi \colon H^{1/2}(\mathbb{T}^2; \mathbb{R}) \to [0, +\infty)$ depends on the obstacles and is defined by

$$\Psi(u) = \sum_{p \in \mathcal{O}} \mu \int_{\mathbb{B}_r(p)} |u(x)| \, \mathrm{d}x.$$

Note that in this situation the friction functional Ψ vanishes on an infinite-dimensional set; however, Ψ is a norm on the finite-dimensional space $\mathbb{R}^{\mathcal{O}}$ of phase field values at the obstacle sites. This is exactly the situation for which the Schur complementation procedure of subsection 6.4.3 was designed. Strictly speaking, theorem 6.4.3 requires a finite-dimensional ambient state space; the conjecture that it applies to this infinite-dimensional context will be assumed without proof.

Thus, for each $\xi \in \mathbb{R}^{\mathcal{O}}$, define the reduced elastic energy by Schur complementation (i.e. by minimizing out the the degrees of freedom that do not experience friction), i.e.

$$E_{\text{red}}(s, \xi) := \inf \left\{ E(s, u) \,\middle|\, u \in H^{1/2}(\mathbb{T}^2; \mathbb{R}) \text{ and } u(x) = \xi_i \text{ for } x \in \mathbb{B}_r(p_i) \right\},$$

and define the reduced dissipative potential by

$$\Psi_{\mathrm{red}}(\xi) := \mu\, \lambda^2(\mathbb{B}_r(0)) \sum_{p\in\mathcal{O}} |\xi_p|.$$

Note that E_{red} is a quadratic form: there exists a matrix $G \in \mathbb{R}^{\mathcal{O}\times\mathcal{O}}$, a vector $\tau \in \mathbb{R}^{\mathcal{O}}$ and a scalar $h \in \mathbb{R}$ such that

$$E_{\mathrm{red}}(s,\xi) = \frac{1}{2}\xi\cdot G\xi - (s\cdot b)\tau\cdot\xi + \frac{h}{2}(s\cdot b)^2.$$

The coefficients of G and τ and the value of h are random and depend on the realization of the obstacle positions. The resulting dual dissipation potential is

$$\mathcal{F}^\star_{0,\mathrm{red}}(\mathrm{D}E_{\mathrm{red}}(\tau,\xi)) = -\sum_{p\in\mathcal{O}} \log\big((\mu\pi r^2)^2 - ((G\xi - (s\cdot b)\tau)\cdot e_p)^2\big),$$

where $e_p \in \mathbb{R}^{\mathcal{O}}$ denotes the unit dislocation over the obstacle site $p \in \mathcal{O}$. Since the number of obstacles is typically large the induced differential equation

$$\dot\xi(t) = -\mathrm{D}\mathcal{F}^\star_{0,\mathrm{red}}(\mathrm{D}E_{\mathrm{red}}(\tau,\xi(t)))$$

for the thermalized dynamics of ξ is difficult to analyze. The situation simplifies when one considers the asymptotic behavior of the solutions in the limit where the number of obstacles $|\mathcal{O}| \to \infty$ and $r \to 0$ such that $\mu\pi r^2|\mathcal{O}| = \sigma > 0$.

For each realization of the obstacle positions, (6.5.2) implies that, for the constant vector $\underline{1} = \sum_{p\in\mathcal{O}} e_p \in \mathbb{R}^{\mathcal{O}}$,

$$G\,\underline{1} = 0 \text{ and } \tau\cdot\underline{1} = \bar\tau = E(\tfrac{1}{|b|^2}b,\underline{0}) - E(\tfrac{1}{|b|^2}b,\underline{1}). \tag{6.5.3}$$

Numerical simulation using Monte Carlo methods strongly suggests that $\bar\tau$ converges to a deterministic value when the number of obstacles tends to infinity [SKTO09]. In view of this and (6.5.3), it makes sense to split ξ into its mean value, $\bar\xi \in \mathbb{R}$, and the fluctuation around the mean value, $\hat\xi \in \mathbb{R}^{\mathcal{O}}$:

$$\bar\xi := \tfrac{1}{|\mathcal{O}|}\xi\cdot\underline{1},$$
$$\hat\xi := \xi - \bar\xi\,\underline{1}.$$

Substituting these definitions into those of E_{red} and Ψ_{red} yields that

$$E_{\mathrm{red}}(s,\xi) = \tfrac{1}{2}\hat{\xi}\cdot G\hat{\xi} - (s\cdot b)\bar{\tau}\bar{\xi} - (s\cdot b)\hat{\tau}\cdot\hat{\xi} + \tfrac{h}{2}(b\cdot s)^2,$$
$$\Psi_{\mathrm{red}}(z) = \Psi_{\mathrm{red}}(\hat{z} + \bar{z}\,\underline{1}),$$

where $\hat{\tau} := \tau - \frac{\bar{\tau}}{|\mathcal{O}|}\underline{1}$, $\bar{z} := \frac{1}{|\mathcal{O}|}z\cdot\underline{1}$, and $\hat{z} := z - \bar{z}\,\underline{1}$. Since $\hat{z}\cdot\underline{1} = 0$, the dissipation function Ψ_{red} satisfies the inequality

$$\Psi_{\mathrm{red}}(\bar{z}\,\underline{1}) \le \Psi_{\mathrm{red}}(\bar{z}\,\underline{1} + \hat{z}) \le \Psi_{\mathrm{red}}(\bar{z}\,\underline{1}) + \Psi_{\mathrm{red}}(\hat{z}).$$

Numerical simulations suggest that as the number of obstacles $|\mathcal{O}|$ becomes large, $\Psi_{\mathrm{red}}(\hat{z})$ becomes small: that is, the fluctuations of ξ around the mean value $\bar{\xi}$ experience very little friction.

This leads to the final reduction step, the reduction from the high-dimensional variable $\xi \in \mathbb{R}^{\mathcal{O}}$ to its one-dimensional mean field $\bar{\xi} \in \mathbb{R}$. The mean-field energetic potential $E_{\mathrm{MF}}\colon \mathbb{R}^* \times \mathbb{R} \to \mathbb{R}$ is obtained, essentially, by reading off the terms that include a $\bar{\tau}$ or $\bar{\xi}$ from the definition of E_{red}, yielding

$$E_{\mathrm{MF}}(\bar{\tau},\bar{\xi}) := -(s\cdot b)\bar{\tau}\bar{\xi}.$$

The corresponding mean-field dissipation potential $\Psi_{\mathrm{MF}}\colon \mathbb{R} \to [0,+\infty)$ is given by

$$\Psi_{\mathrm{MF}}(\bar{z}) := \left(\mu\lambda^2(\mathbb{B}_r(0))|\mathcal{O}|\right)|\bar{z}| = \sigma|\bar{z}|,$$

The effective dual dissipation potential associated to this one-dimensional Ψ_{MF} is just $\mathcal{F}^\star_{0,\mathrm{MF}}(w) := -\log(\sigma^2 - w^2)$ and so

$$\mathcal{F}^\star_{0,\mathrm{MF}}(\mathrm{D}E_{\mathrm{MF}}(\bar{\tau},\bar{\xi})) = -\log\left(\sigma^2 - (s\cdot b)^2\bar{\tau}^2\right).$$

Note that $\mathcal{F}^\star_{0,\mathrm{MF}}$ is always a lower bound for $\mathcal{F}^\star_{0,\mathrm{red}}$ even if the number of obstacles is finite; cf. the Bogolyubov inequality for the free energy in statistical mechanics [Bog66] [Bog74].

These reduction steps yield a scalar differential equation for the mean slip $\bar{\xi}\colon [0,T] \to \mathbb{R}$:

$$\frac{\mathrm{d}\bar{\xi}}{\mathrm{d}t} = -\theta\mathrm{D}\mathcal{F}^\star_{0,\mathrm{MF}}\left(\mathrm{D}E_{\mathrm{MF}}\left(\bar{\tau},\bar{\xi}\right)\right) = \frac{2\theta\bar{\tau}}{\sigma^2 - \bar{\tau}^2}$$

with appropriate initial conditions; in order to simplify the notation it is

assumed that $s \cdot b = 1$. Assume a constant applied load $\bar{\tau}$ and that the resistance is σ, with $0 \leq |\bar{\tau}| < \sigma$ — once again, this is the stability criterion that the net load should remain in the interior of the elastic region. At least formally, σ may vary provided that the inequality $\sigma > |\bar{\tau}| \geq 0$ still holds.[6.2] Under the additional assumption of linear strain hardening, i.e. $\sigma = \sigma_0 \bar{\xi}$, the previous equation becomes

$$\frac{\mathrm{d}\bar{\xi}}{\mathrm{d}t} = \frac{2\theta\bar{\tau}}{\sigma_0^2\bar{\xi}(t)^2 - \bar{\tau}^2}.$$

Clearly, the behaviour of $\bar{\xi}(t)$ for small t will depend upon the magnitude of $\bar{\tau}$ as compared to σ_0. However, under the assumption that the effective applied stress $\bar{\tau}$ is small in comparison to the frictional resistance $\sigma = \sigma_0\bar{\xi}$,

$$\frac{\mathrm{d}\bar{\xi}}{\mathrm{d}t} \approx \frac{2\theta\bar{\tau}}{\sigma_0^2\bar{\xi}(t)^2}.$$

This is the same differential equation, under the same assumption of strain hardening, as that proposed in [Cot96a] [Cot96b] [Cot96c], but obtained *via* a different line of reasoning: Cottrell's approach is to consider creep as an instance of critical avalanche formation analogous to the sandhill models introduced to explain self-organized criticality. Integration of the above differential equation yields

$$\bar{\xi}(t) \approx \left(\bar{\xi}(0) + \frac{6\theta\bar{\tau}t}{\sigma_0^2}\right)^{1/3} \sim t^{1/3} \text{ as } t \to +\infty,$$

by the general solution

$$\frac{\mathrm{d}y}{\mathrm{d}t} = \frac{k\theta}{(\sigma_0\theta^\alpha y(t)^\beta)^2} \implies y(t) = \left((1+2\beta)\left(C_1 + \frac{k\theta^{1-2\alpha}t}{\sigma_0^2}\right)\right)^{\frac{1}{1+2\beta}}.$$

Thus, Andrade's $t^{1/3}$ creep law follows as a straightforward consequence of the limiting dynamics when applied to the mean-field reduction of the Koslowski–Cuitiño–Ortiz phase-field model under the assumption of linear strain hardening. Furthermore, the theory also predicts more rapid creep

[6.2] This is an example of a situation in which the dissipation should be seen as a Finsler metric; linear strain hardening corresponds to the Finsler metric $\Psi(x, \dot{x}) = \sigma_0 x|\dot{x}|$ on the manifold $(0, +\infty)$ with tangent bundle $(0, +\infty) \times \mathbb{R}$. For an introduction to Finsler geometry, see [Che96] and the references therein.

for larger values of the temperature-like parameter θ, as intuition would suggest. It would be of interest to see if the above methods also provide a good estimate for the creep rate in the presence of thermal softening, i.e. $\alpha \neq 0$.

6.6 Directions for Further Research

This section outlines two natural extensions of the theory to more general state spaces: manifolds and infinite-dimensional spaces. The sought-for result is to prove, in as many settings as possible, that the càdlàg interpolants of the Markov chain $X^{(P)}$ converge as $[\![P]\!] \to 0$ to the solutions of the nonlinear deterministic gradient descent

$$0 \in \partial \mathcal{F}_0(x(t), -\dot{x}(t)/\theta) - \partial E(t, x(t));$$

this evolutionary problem can be seen as a doubly nonlinear problem in the style of [CV90].

In both cases, as mentioned earlier in this chapter, it is helpful to see $\mathcal{F}_0$ (resp. $\mathcal{F}_0^\star$) as a function on the tangent (resp. cotangent) bundle of the state space. The notion of a tangent measure in the sense of [Pre87] [O'N95] *et al.* enters the analysis quite naturally. The outlook for infinite-dimensional spaces is somewhat less clear than the outlook for finite-dimensional manifolds, simply because measures on infinite-dimensional spaces are somewhat ill-behaved objects in comparison to their finite-dimensional counterparts.

6.6.1 Manifolds as State Spaces

Let $(\mathcal{M}, g)$ be a connected, real, m-dimensional Riemannian manifold.[6.3] Suppose that a (smooth enough) energetic potential $E\colon [0,T] \times \mathcal{M} \to \mathbb{R}$ is specified, and that the dissipation potential is given by the geodesic distance

$$d(x_i, x_{i+1}) := \inf \left\{ \int_0^1 \|\dot{\gamma}(t)\|_g \, \mathrm{d}t \,\middle|\, \begin{array}{c} \gamma\colon [0,1] \to \mathcal{M} \text{ is a smooth} \\ \text{curve with } \gamma(0) = x_i, \gamma(1) = x_{i+1} \end{array} \right\}.$$

[6.3] The assumption that the manifold has a Riemannian structure can probably be weakened. As noted in [MT04], "a more general treatment should use Banach manifolds and the dissipation is then a Finsler metric". See, e.g., [Mie03] for applications of this idea.

(The connectedness of $\mathcal{M}$ ensures that the geodesic distance is, indeed, a metric on $\mathcal{M}$.) Let $\mathrm{d}V$ denote the standard Riemannian volume form on $\mathcal{M}$: that is, if $E = \{E_1, \ldots, E_m\}$ is any frame and $\Phi = \{\Phi^1, \ldots, \Phi^m\}$ is the dual coframe (i.e. $\langle \Phi^i, E_j \rangle = \delta^i_j$), then $\mathrm{d}V$ is given by

$$\mathrm{d}V = \sqrt{\det(g_{ij})}\, \Phi^1 \wedge \ldots \wedge \Phi^m, \text{ where } g_{ij} := g(E_i, E_j).$$

The thermalized gradient descent Markov chain scheme again makes sense: the transition kernels are again given by

$$\mathbb{P}[X_{i+1} \in A \mid X_i = x_i] \propto \int_A e^{-(E(t_{i+1}, x_{i+1}) + \psi(d(x_i, x_{i+1})))/\varepsilon}\, \mathrm{d}V(x_{i+1}).$$

The idea is to proceed as before in the computation of the expected state of the Markov chain at the next time step, and to turn an integral over the manifold into an integral over (a suitable portion of) the tangent space at x_i. To do this requires more information than just the topological structure of $\mathcal{M}$: what is needed is the exponential map generated by the metric g — see, for example, [dC92, chapter 3] [Lee97, chapter 5]. For $x \in \mathcal{M}$, $\exp_x$ denotes the exponential map from some neighbourhood of the origin in $\mathrm{T}_x\mathcal{M}$ to $\mathcal{M}$: that is, $\exp_x v := \gamma_v(1)$, where γ_v is the unique geodesic with $\gamma_v(0) = x$, $\gamma_v'(0) = v$. For real-valued differentiable functions f defined near x, the exponential map interacts with the derivative as follows:

$$f(\exp_x v) = f(x) + \langle \mathrm{D}f(x), v \rangle + o(\|v\|_g) \text{ as } \|v\|_g \to 0.$$

For each $x \in \mathcal{M}$, let U_x be a star-shaped neighbourhood of the origin in $\mathrm{T}_x\mathcal{M}$ such that the exponential map $\exp_x$ maps U_x diffeomorphically onto a subset of $\mathcal{M}$ that has full volume measure. (Take, for example, the region bounded by the preimage under $\exp_x$ of the cut locus for x.) Thus, any integral over $\mathcal{M}$ can be turned into an integral over U_x. Thus, a calculation in the style of subsection 6.3.1 working in normal coordinates (see lemma 5.10 and proposition 5.11 of [Lee97]) at $x_i \in \mathcal{M}$ yields that

$$\begin{aligned}
&\mathbb{E}[X_{i+1} \mid X_i = x_i] \\
&\quad \approx \exp_{x_i}\left(\frac{1}{Z}\int_{U_{x_i}} u e^{-(\langle \mathrm{D}E(t_{i+1},x_i),u\rangle + \|u\|_g)/\varepsilon}\,\mathrm{d}(\exp_{x_i}^* V)(u)\right) \\
&\quad = \exp_{x_i}\left(\frac{\varepsilon}{Z}\int_{U_{x_i}} u e^{-(\langle \mathrm{D}E(t_{i+1},x_i),u\rangle + \|u\|_g)/\varepsilon}\,\mathrm{d}u^1 \wedge \ldots \wedge \mathrm{d}u^m\right) \\
&\quad = \exp_{x_i}\left(\frac{\varepsilon}{Z}\int_{U_{x_i}/\varepsilon} v e^{-(\langle \mathrm{D}E(t_{i+1},x_i),v\rangle + \|v\|_g)}\,\mathrm{d}v^1 \wedge \ldots \wedge \mathrm{d}v^m\right) \\
&\quad \approx \exp_{x_i}\left(\frac{\varepsilon}{Z}\int_{\mathrm{T}_p\mathcal{M}} v e^{-(\langle \mathrm{D}E(t_{i+1},x_i),v\rangle + \|v\|_g)}\,\mathrm{d}v^1 \wedge \ldots \wedge \mathrm{d}v^m\right).
\end{aligned}$$

(The first $\approx$ sign arises because the second- and higher-order derivatives of E have been neglected; the second $\approx$ sign arises from the heuristic that the integrand should be small outside U_{x_i}/ε for small ε.) The measure associated to $\mathrm{d}v^1 \wedge \ldots \wedge \mathrm{d}v^m$ is a multiple of m-dimensional Hausdorff measure $\mathscr{H}^m$, and is a tangent measure for the volume measure on $\mathcal{M}$. This suggests the following:

Definition 6.6.1. For a Riemannian manifold $(\mathcal{M}, g)$, define the *effective dual dissipation potential* $\mathcal{F}_0^\star \colon \mathrm{T}^*\mathcal{M} \to \mathbb{R} \cup \{+\infty\}$ by

$$\mathcal{F}_0^\star(x,\ell) := \log \int_{\mathrm{T}_x\mathcal{M}} e^{-(\langle \ell, v\rangle + \|v\|_g)}\,\mathrm{d}v, \tag{6.6.1}$$

and the *effective dissipation potential* $\mathcal{F}_0 \colon \mathrm{T}\mathcal{M} \to \mathbb{R} \cup \{+\infty\}$ by fibre-wise convex conjugation of $\mathcal{F}_0^\star$:

$$\mathcal{F}_0(x,v) := \sup\left\{\langle \ell, v\rangle - \mathcal{F}_0^\star(x,\ell) \mid \ell \in \mathrm{T}_x\mathcal{M}\right\}. \tag{6.6.2}$$

Physical examples of evolutions in non-Euclidean spaces that would require this manifold treatment are easy to come by. For example, the orientation of a magnetic domain or grain in a polycrystal is naturally modeled as an element of a sphere or special orthogonal group. There is also cause to consider mean-field problems as before: recent experimental observations [RSPR09] of the behaviour of polycrystalline bcc tantalum suggest that polycrystalline response is dominated by one single instantaneous grain orientation, i.e. a polycrystal behaves very similarly to a single crystal. This

suggests that a mean-field approximation is highly accurate in this case, and this calls for further investigation.

6.6.2 Infinite-Dimensional State Spaces

A second natural direction in which to extend the analysis is to infinite-dimensional state spaces. For example, in the Koslowski–Cuitiño–Ortiz phase-field model for dislocation dynamics [KCO02], the state space is the Sobolev space $H^{1/2}(\mathbb{T}^2; \mathbb{R})$. Since the dissipation potential used in this model an ℓ^1 seminorm that is non-trivial on a finite-dimensional subspace of $H^{1/2}(\mathbb{T}^2; \mathbb{R})$, it is possible to reduce the dynamics to a finite-dimensional problem. Clearly, however, any general theory has to be capable of tackling infinite-dimensional evolutions directly.

One of the first difficulties encountered in an infinite-dimensional setting is the lack of "decent" reference measures. The following theorems form part of the standard literature on the topic: see, for example, the discussion in [HSY92, introduction] or [Yam85, part B, chapter 1, section 5].

Theorem 6.6.2. *Let $\mathcal{X}$ be an infinite-dimensional, separable Banach space. Then the only locally finite and translation-invariant Borel measure μ on $\mathcal{X}$ is the trivial measure.*

Theorem 6.6.3. *Let $\mathcal{X}$ be an infinite-dimensional, locally convex topological vector space. Then the only σ-finite and translation-quasi-invariant Borel measure μ on $\mathcal{X}$ is the trivial measure.*

A usual candidate for a reference measure on a separable Banach space $\mathcal{X}$ is Wiener measure, or, more generally, any non-degenerate Gaussian measure. Such a measure would be strictly positive and locally finite, but quasi-invariant only under translations by elements of the Cameron–Martin space $\mathcal{H}$ [CM44], a dense proper Hilbertian subspace of $\mathcal{X}$; indeed, for translations by elements of $\mathcal{X} \setminus \mathcal{H}$, the push-forward measure is singular with respect to the original measure [Mar50]. A Gaussian measure is also mutually singular with respect to its push-forward under any non-trivial dilation of the space. That said, none of these obstacles make it impossible to define a thermalized gradient descent.

Suppose that $\mathcal{X}$ is a reflexive and separable Banach space and that γ is a reference Gaussian measure on $\mathcal{X}$. Let the energetic potential E, the dissipation potential Ψ and the Markov chain X be as usual. Let f be a

function on $\mathcal{X}$ that is homogeneous of degree d. Then a calculation in the style of subsection 6.3.1, neglecting the second and higher derivatives of E, yields that

$$\mathbb{E}[f(\Delta X_{i+1}) \mid X_i] \approx \varepsilon^d \frac{\displaystyle\int_{\mathcal{X}} f(z) e^{-(\langle \mathrm{D}E(t_{i+1}, x_i), z\rangle + \Psi(z))} \,\mathrm{d}(T_*^{x_i,\varepsilon}\gamma)(z)}{\displaystyle\int_{\mathcal{X}} e^{-(\langle \mathrm{D}E(t_{i+1}, x_i), z\rangle + \Psi(z))} \,\mathrm{d}(T_*^{x_i,\varepsilon}\gamma)(z)},$$

where $T^{x_i,\varepsilon} \colon \mathcal{X} \to \mathcal{X}$ is the translation-and-dilation map $x_{i+1} \mapsto (x_{i+1} - x_i)/\varepsilon$.

In view of the finite-dimensional results of this chapter, the natural conjecture would be that the Markov chain $X^{(P)}$ converges as $[\![P]\!] \to 0$ and that the continuous-time limit is a deterministic process that satisfies the deterministic nonlinear gradient descent

$$\mathrm{D}\mathcal{F}_0(\dot{y}(t)) = -\mathrm{D}E(t, y(t))$$

with the same initial conditions as X. However, there what is the effective dual dissipation potential $\mathcal{F}_0^\star$ in this case? It helps to have considered the manifolds case before, since the measure used to define $\mathcal{F}_0^\star$ depends on the current state, so $\mathcal{F}_0^\star$ is a function on the cotangent bundle $\mathrm{T}^*\mathcal{X} \cong \mathcal{X} \times \mathcal{X}^*$. $\mathcal{F}_0^\star$ "ought" to be something like

$$\mathcal{F}_0^\star(x, w) := \log \int_{\mathcal{X}} \exp(-(\langle w, z\rangle + \Psi(z))) \,\mathrm{d}\tilde{\gamma}_x(z), \tag{6.6.3}$$

where $\tilde{\gamma}_x$ is (a suitably normalized version of) the weak limit of the measures $T_*^{x,\varepsilon}\gamma$ as $\varepsilon \to 0$. That is, $\tilde{\gamma}$ is a tangent measure for γ at x.

The definition (6.6.3) and the attendant normalization require some care since the measures $T_*^{x_i,\varepsilon}\gamma$ themselves behave very badly as $\varepsilon \to 0$. For example, suppose that $T_*^{0,\varepsilon}\gamma \rightharpoonup \nu$ and let $U \subsetneq \mathcal{X}$ be any bounded open set. Then

$$\big(T_*^{0,\varepsilon}\gamma\big)(U) = \gamma\big((T^{0,\varepsilon})^{-1}(U)\big) = \gamma(\varepsilon U) \to 0 \text{ as } \varepsilon \to 0.$$

Yet, by the portmanteau theorem (theorem D.1),

$$\nu(U) \leq \limsup_{\varepsilon \to 0} \big(T_*^{0,\varepsilon}\gamma\big)(U) = 0.$$

Therefore, no bounded set can have positive measure under ν. In the $\mathbb{R}^n$

and manifolds cases, this issue was taken care of by the cancellation of ε^n's in the numerator and denominator of expressions for $\mathbb{E}[f(\Delta X_{i+1}) \mid X_i]$. Multiplicative renormalization of the measures $T_*^{x,\varepsilon}\gamma$ by a factor dependent on ε but not on w or x would only change $\mathcal{F}_0^\star(x,w)$ up to an additive constant, and such a change is irrelevant to the claimed limiting continuous-time gradient descent. A suitable renormalization might be $\gamma(\mathbb{B}_\varepsilon(x))$.

As a final note, observe that the choice of reference measure γ makes a difference to the dynamics and the equilibrium state of the system. For example, if E is trivial, and Ψ is (a multiple of) the norm $\|\cdot\|$ on $\mathcal{X}$, then the system does not rest at state $x \in \mathcal{X}$ unless the measure $\tilde{\gamma}_x$ is a symmetric measure about 0. For more general dissipations, equilibrium states are those $(t,x) \in [0,T] \times \mathcal{X}$ for which

$$\int_{\mathcal{X}} z \exp(-(\langle \mathrm{D}E(t,x), z\rangle + \Psi(z)))\, \mathrm{d}\tilde{\gamma}_x(z) = 0.$$

6.7 Proofs and Supporting Results

6.7.1 Effective (Dual) Dissipation Potential

Lemma 6.7.1. *Let $\mathscr{E}$ and Ψ satisfy the usual hypotheses. Let $m\colon (\mathbb{R}^n)^* \to \mathbb{R}$ be given by*

$$m(v) := \inf\left\{ \langle v, u\rangle + \Psi(u) \,\middle|\, u \in \mathbb{S}^{n-1} \right\},$$

where $\mathbb{S}^{n-1}$ denotes the Euclidean unit sphere in $\mathbb{R}^n$. Then m is continuous and satisfies

$$m(v) \begin{cases} > 0, & \text{if } -v \in \mathring{\mathscr{E}}, \\ = 0, & \text{if } -v \in \partial\mathscr{E}, \\ < 0, & \text{if } -v \notin \mathscr{E}. \end{cases}$$

Proof. To save space, write $f(v,x) := \langle v, x\rangle + \Psi(x)$. Since $\chi_{\mathscr{E}}$ is convex and lower semicontinuous,

$$\inf_{x\in\mathbb{R}^n} f(v,x) = -\Psi^\star(-v) = -\chi_{\mathscr{E}}(-v). \tag{6.7.1}$$

Since $f(v,x)$ is 1-homogeneous in x, it follows that $m(v) < 0$ for $-v \notin \mathscr{E}$ and that $m(v) \geq 0$ for $-v \in \mathscr{E}$.

Note that f is continuous. Since m is a pointwise infimum of a family of continuous functions, it is upper semicontinuous. Since $\mathbb{S}^{n-1}$ is compact,

m is a pointwise infimum of a compactly-parametrized family of continuous functions, and so is also lower semicontinuous [PT82]. Thus, m is continuous.

Suppose that there exists $-v \in \mathring{\mathscr{E}}$ with $m(v) = 0$. By the compactness of $\mathbb{S}^{n-1}$, this implies that there exists a unit vector $u_0 \in \mathbb{S}^{n-1}$ with

$$f(v, u_0) = 0. \tag{6.7.2}$$

But, since $-v \in \mathring{\mathscr{E}}$, there exists $\alpha > 1$ such that $-\alpha v \in \mathring{\mathscr{E}}$. Then (6.7.2) implies that $\langle v, u_0 \rangle < 0$, so $f(\alpha v, u_0) < 0$, and so $m(\alpha v) < 0$, which contradicts (6.7.1). Hence, $m(v) > 0$ for $-v \in \mathring{\mathscr{E}}$.

It remains only to show that $m(v) = 0$ for $-v \in \partial\mathscr{E}$. Suppose not, i.e. that there exists $-v \in \partial\mathscr{E}$ with $m(v) > 0$. Since $-v \in \partial\mathscr{E}$ and $\mathscr{E}$ is convex (and hence star-convex with respect to the origin in $(\mathbb{R}^n)^*$), for every $\alpha > 1$, $-\alpha v \notin \mathscr{E}$, and so $m(\alpha v) < 0$. Hence, by the continuity of m,

$$0 \geq \lim_{\alpha \searrow 1} m(\alpha v) = m(v) > 0,$$

which is a contradiction. This completes the proof. ■

Proof of theorem 6.3.3. In this proof, to save space, ψ will denote the finite Borel measure on $\mathbb{R}^n$ defined by

$$\mathrm{d}\psi(z) := \exp(-\Psi(z))\,\mathrm{d}z.$$

1. By the continuity-and-coercivity inequality that, for all $x \in \mathbb{R}^n$,

   $$c_\Psi |x| \leq \Psi(x) \leq C_\Psi |x|,$$

   ψ is a strictly positive and finite measure. Hence, since the exponential function in the integrand of (6.3.1) is never zero, the claim follows.

2. Consider the spherical integral form (6.3.4) for $\mathcal{F}_0^\star$. By lemma 6.7.1, if $-w \in \mathring{\mathscr{E}}$, then the integral is that of a continuous and bounded function over a compact set, so the integral exists and is finite.

   If $-w \in \partial\mathscr{E}$, then, as in the proof of lemma 6.7.1, there exists $u_w \in \mathbb{S}^{n-1}$ with $\langle w, u_w \rangle + \Psi(u_w) = 0$, so the integrand has a pole. The

triangle inequality for Ψ implies that for $u_w + u \in \mathbb{R}^n$,

$$\begin{aligned}\langle w, u_w + u\rangle + \Psi(u_w + u) &\leq \langle w, u_w\rangle + \langle w, u\rangle + \Psi(u_w) + \Psi(u)\\ &\leq \langle w, u\rangle + \Psi(u)\\ &\leq |u|\big(|w| + C_\Psi\big).\end{aligned}$$

Hence, the integrand in (6.3.4) lies in $\Omega(|u|^{-n})$; hence, by the standard result that $x \mapsto |x|^{-\alpha}$ lies in L^1 for a d-dimensional domain about 0 if, and only if, $\alpha < d$, it follows that $\mathcal{F}_0^\star(w) = +\infty$.

If $-w \notin \mathscr{E}$, then it is clear from lemma 6.7.1 that the integral in (6.3.1) does not converge, and so $\mathcal{F}_0^\star(w) = +\infty$.

3. Let $v, w \in -\mathring{\mathscr{E}}$ and $0 < t < 1$. Then

$$\begin{aligned}&\mathcal{Z}_0^\star(tv + (1-t)w)\\ &\quad= \int_{\mathbb{R}^n} \exp\big(-\langle tv + (1-t)w, z\rangle\big)\,\mathrm{d}\psi(z)\\ &\quad= \int_{\mathbb{R}^n} \exp\big(-t\langle v, z\rangle\big)\exp\big(-(1-t)\langle w, z\rangle\big)\,\mathrm{d}\psi(z)\end{aligned}$$

which, by the Hölder–Rogers inequality,

$$\begin{aligned}&\leq \left(\int_{\mathbb{R}^n} \exp\big(-\langle v, z\rangle\big)\,\mathrm{d}\psi(z)\right)^t \left(\int_{\mathbb{R}^n} \exp\big(-\langle w, z\rangle\big)\,\mathrm{d}\psi(z)\right)^{1-t}\\ &= \mathcal{Z}_0^\star(v^*)^t\,\mathcal{Z}_0^\star(w)^{1-t}.\end{aligned}$$

Hence, since the logarithm is a monotonically increasing function, for all $v, w \in -\mathring{\mathscr{E}}$,

$$\mathcal{F}_0^\star(tv + (1-t)w) \leq t\mathcal{F}_0^\star(v^*) + (1-t)\mathcal{F}_0^\star(w).$$

Moreover, since $-\mathscr{E}$ is convex and $\mathcal{F}_0^\star$ is identically $+\infty$ outside the interior of $-\mathscr{E}$, $\mathcal{F}_0^\star$ is convex on all of $(\mathbb{R}^n)^*$.

4. The derivative $\mathrm{D}\mathcal{Z}_0^\star\colon -\mathring{\mathscr{E}} \to (\mathbb{R}^n)^{**} \cong \mathbb{R}^n$ can be computed using the standard theorem on differentiation under the integral sign, yielding

$$\mathrm{D}\mathcal{Z}_0^\star(v) = \int_{\mathbb{R}^n} -z\exp\big(-\langle v, z\rangle\big)\,\mathrm{d}\psi(z),$$

and so on for higher-order derivatives:

$$\mathrm{D}^k \mathcal{Z}_0^\star(v) = \int_{\mathbb{R}^n} (-z)^{\otimes k} \exp\big(-\langle v, z\rangle\big)\, \mathrm{d}\psi(z).$$

The integrals involved are all finite for $-v \in \mathring{\mathscr{E}}$ because of the exponentially small tails of the measure ψ.

5. As in the proof of the second part of the claim, let $-w \in \mathring{\mathscr{E}}$ and let $u_w \in \mathbb{S}^{n-1}$ be such that $\langle w, u_w\rangle + \Psi(u_w)$ is minimal (i.e. equals $m(w)$). Then
$$\langle w, u_0 + u\rangle + \Psi(u_0 + u) \le m(w) + |u|\big(|w| + C_\Psi\big).$$
Since $m(w) \to 0$ as $-w \to \partial\mathscr{E}$, the same argument as in part 2 applies, and so $\mathcal{F}_0^\star(w) \to +\infty$ as $-w \to \partial\mathscr{E}$. Now suppose that $\mathrm{D}\mathcal{F}_0^\star$ does not blow up. Then, since $\mathcal{F}_0^\star$ is smooth and $\mathscr{E}$ is compact, $\mathcal{F}_0^\star$ would be bounded on $-\mathscr{E}$, which is a contradiction. ■

6.7.2 Convergence Theorems in $\mathbb{R}^n$

Proof of theorem 6.4.1. Without loss of generality, take $\theta = 1$, i.e. $\varepsilon_i = \Delta t_i$. Fix $T > 0$ and a partition P of $[0, T]$. Define a deterministic sequence $y^{(P)}$, approximating the solution y to (6.4.2), by the implicit Euler scheme:

$$\begin{aligned} y_0^{(P)} &:= x_0; \\ y_i^{(P)} &:= y_{i-1}^{(P)} - \Delta t_i \mathrm{D}\mathcal{F}_0^\star\big(\mathrm{D}E(t_i, y_{i-1}^{(P)})\big); \end{aligned}$$

Note, however, that for the energy E used in this theorem, $\Delta y_i^{(P)}$ does not depend on $y_{i-1}^{(P)}$. Let $Z^{(P)} := X^{(P)} - y^{(P)}$. Standard results on the global error associated to an implicit Euler scheme [KP92, section 8.1] imply that $\|y^{(P)} - y\|_\infty \le C[\![P]\!]$; thus, to prove (6.4.3), it is enough to show that, for all $\lambda > 0$,

$$\mathbb{P}\left[\max_{1 \le i \le |P|} \big|Z_i^{(P)}\big| \ge \lambda\right] \le C[\![P]\!]. \tag{6.7.3}$$

By lemma 6.3.1, for all $i = 1, \ldots, |P|$,

$$\begin{aligned} \mathbb{E}\big[\Delta X_i^{(P)} \big| X_{i-1}^{(P)}\big] &= -\Delta t_i \mathrm{D}\mathcal{F}_0^\star(\mathrm{D}E(t_i, X_i^{(P)})); \\ \mathbb{E}\left[\big|\Delta X_i^{(P)} - \mathbb{E}\big[\Delta X_i^{(P)} \big| X_{i-1}^{(P)}\big]\big|^2 \Big| X_{i-1}^{(P)}\right] &\le C_\delta (\Delta t_i)^2, \end{aligned}$$

and the constant C_δ depends only on $\delta := \inf_{t\in[0,T]} \operatorname{dist}(\ell(t), \partial\mathscr{E}) > 0$. In fact, in this case, the conditional expectation and variance do not depend on the prior state $X_{i-1}^{(P)}$, and so are equal to the unconditional expectation and variance respectively. Thus, for all $i = 1, \dots, |P|$,

$$\begin{aligned}\mathbb{E}\big[\Delta Z_i^{(P)}\big] &= 0;\\ \mathbb{E}\big[|\Delta Z_i^{(P)}|^2\big] &\leq C_\delta(\Delta t_i)^2.\end{aligned}$$

Applying Kolmogorov's inequality (D.2) to each of the n components of the $Z_i^{(P)}$ yields that, for some constant C_n depending on n,

$$\begin{aligned}\mathbb{P}\left[\max_{1\leq i\leq |P|} \big|Z_i^{(P)}\big| \geq \lambda\right] &\leq \frac{C_n}{\lambda^2}\sum_{i=1}^{|P|}\mathbb{E}\big[|\Delta Z_i^{(P)}|^2\big]\\ &\leq \frac{[\![P]\!]C_n}{\lambda^2}\sum_{i=1}^{|P|}\Delta t_i C_\delta\\ &= \frac{[\![P]\!]C_n C_\delta T}{\lambda^2},\end{aligned}$$

which establishes (6.7.3) and completes the proof. ■

The next lemma (lemma 6.7.2) concerns the closeness of the effective dual dissipation potential $\mathcal{F}_0^\star$ and the corresponding quantity $\mathcal{F}_\varepsilon^\star$ that controls the increments of the Markov chain. The next lemma after that (lemma 6.7.4) gives the resulting bound for the classical gradient descents in $\mathcal{F}_0^\star \circ \mathrm{D}E$ and $\mathcal{F}_\varepsilon^\star \circ \mathrm{D}E$. The proofs are not hard (and also not much fun). It is worth noting that these two lemmata apply to the prototypical case of a quadratic energetic potential $E(t,x) := \frac{1}{2}\langle Ax, x\rangle - \langle \ell(t), x\rangle$, where $A\colon \mathbb{R}^n \to (\mathbb{R}^n)^*$ is symmetric and positive-definite and $\ell \in W^{1,\infty}([0,T]; (\mathbb{R}^n)^*)$ (i.e. ℓ is uniformly Lipschitz).

Lemma 6.7.2. *Suppose that the energetic potential E is smooth enough that*

$$M := \sup_{t\in[0,T]}\sup_{k\geq 2}\big\|\mathrm{D}^k E(t,\cdot)\big\|_{\mathrm{op}} < +\infty.$$

Then, for every $K \subset\subset -\mathscr{E}$ and every $k \in \mathbb{N}_0$, $\mathrm{D}^k\mathcal{F}_\varepsilon^\star \to \mathrm{D}^k\mathcal{F}_0^\star$ uniformly on K as $\varepsilon \to 0$. More precisely, for every such K and k, there exists a constant

C such that

$$\sup_{w\in K}\left|\mathrm{D}^k\mathcal{F}_\varepsilon^\star(w)-\mathrm{D}^k\mathcal{F}_0^\star(w)\right|\le C\varepsilon^{1/2}\ \textit{for all small enough}\ \varepsilon>0.$$

Proof. The essential quantity to estimate is

$$\mathcal{E}_k^\varepsilon(w):=\int_{\mathbb{R}^n}\left|1-\exp\left(-\sum_{k=2}^\infty\frac{\varepsilon^{k-1}}{k!}\langle\mathrm{D}^kE(t,x),z^{\otimes k}\rangle\right)\right||z|^ke^{-(\langle w,z\rangle+\Psi(z))}\,\mathrm{d}z,$$

since, by the elementary inequality

$$\left|\frac{a}{b}-\frac{c}{d}\right|\le\frac{|a-c|}{|d|}+\frac{|c||b-d|}{|bd|},$$

it holds true that

$$|\mathrm{D}^k\mathcal{F}_\varepsilon^\star(w)-\mathrm{D}^k\mathcal{F}_0^\star(w)|\le\frac{1}{\mathcal{Z}_0^\star(w)}\mathcal{E}_k^\varepsilon(w)+\frac{|\mathrm{D}^k\mathcal{Z}_0^\star(w)|}{\mathcal{Z}_0^\star(w)\mathcal{Z}_\varepsilon(w)}\mathcal{E}_0^\varepsilon(w).\tag{6.7.4}$$

Let $m(w):=\inf\{\langle w,z\rangle+\Psi(z)\mid|z|=1\}$. By lemma 6.7.1, m is continuous and bounded away from 0 on K. Similarly, since $\mathcal{Z}_0^\star$ and $\mathcal{Z}_\varepsilon$ are continuous and positive, they are bounded away from 0 on K, and $|\mathrm{D}^k\mathcal{Z}_0^\star|$ is bounded on K. (Note that all these bounds fail on $-\partial\mathscr{E}$, so the assumption that $K\subset\subset-\mathscr{E}$ is essential.) Thus, the emphasis is on estimating $\mathcal{E}_k^\varepsilon(w)$ in terms of ε and uniformly over K.

It is important to note that the naïve bound

$$\mathcal{E}_k^\varepsilon(w)\le\int_{\mathbb{R}^n}|z|^ke^{-(\langle w,z\rangle+\Psi(z))}\,\mathrm{d}z,$$

is no help. Instead, $\mathcal{E}_k^\varepsilon(w)$ will be estimated by splitting the integral into two parts: an integral over a ball around the origin in $\mathbb{R}^n$ (where a better bound than the naïve one will apply) and the rest (where only the naïve one will be used). More precisely, for any $a\in(0,1)$, let $R=R(a,\varepsilon,x,t)>0$ be such that

$$|z|\le R\implies1-\exp\left(-\sum_{k=2}^\infty\frac{\varepsilon^{k-1}}{k!}\langle\mathrm{D}^kE(t,x),z^{\otimes k}\rangle\right)\le a.\tag{6.7.5}$$

Converting to spherical polar coordinates yields that, for some constant c_n

depending only on n,

$$\mathcal{E}_k^\varepsilon(w) \leq c_n a \int_0^R r^{k+n-1} e^{-m(w)r}\,\mathrm{d}r + c_n \int_R^{+\infty} r^{k+n-1} e^{-m(w)r}\,\mathrm{d}r$$

In view of the uselessness of the naïve bound, the trick lies in choosing a and R such that $a \to 0$ and $R \to \infty$ at the right rates. This estimate is valid for any $a \in (0,1)$ and corresponding R. The above integrals can be evaluated exactly using the recurrence relation

$$\int x^n e^{cx}\,\mathrm{d}x = \frac{x^n e^{cx}}{c} - \frac{n}{c}\int x^{n-1} e^{cx}\,\mathrm{d}x;$$

the resulting polynomial-exponential expressions are a bit cumbersome to deal with, but only the leading-order contributions as $\varepsilon \to 0$ are of interest here.

Now take a specific choice of a and R: let $a := \varepsilon^{1/2}$. Note that the right-hand side of (6.7.5) holds if

$$\frac{1}{\varepsilon}\sum_{k=2}^{\infty} \frac{\varepsilon^k}{k!}\big\langle \mathrm{D}^k E(t,x), z^{\otimes k}\big\rangle \leq -\log(1-\varepsilon^{1/2}),$$

and clearly

$$\frac{1}{\varepsilon}\sum_{k=2}^{\infty} \frac{\varepsilon^k}{k!}\big\langle \mathrm{D}^k E(t,x), z^{\otimes k}\big\rangle \leq \frac{1}{\varepsilon}\sum_{k=2}^{\infty} \frac{\varepsilon^k}{k!}\big\|\mathrm{D}^k E(t,\cdot)\big\|_{\mathrm{op}}|z|^k \leq \frac{M}{\varepsilon}\exp(\varepsilon|z|).$$

Thus, it is enough to have $|z| \leq R$ for

$$R := \frac{1}{\varepsilon}\log\left(-\frac{\varepsilon}{M}\log(1-\varepsilon^{1/2})\right).$$

(Note that, if $M = 0$, then this is the case of a flat energetic potential, in which case $\mathcal{F}_\varepsilon^\star \equiv \mathcal{F}_0^\star$ and there is nothing to prove.) By l'Hôpital's rule, for this choice of a and R, $a \to 0$ and $R \to +\infty$ as $\varepsilon \to 0$. Hence, for this choice of a and R, there exist constants c_1, c_2 such that

$$\mathcal{E}_k^\varepsilon(w) \leq c_1\varepsilon^{1/2}\left(\frac{R}{m(w)}\right)^{k+n-1} e^{-m(w)R} + c_2\left(\frac{R}{m(w)}\right)^{k+n-1} e^{-m(w)R}.$$

The dominant term here is the $\varepsilon^{1/2}$ term, since $R^{k+n-1}e^{-m(w)R}$ not only tends to 0, but does so with all derivatives tending to zero as well; $m(w)$ is

bounded away from zero for $w \in K$. Thus, there is a constant C_k (dependent on k and the other geometric parameters, but not on ε) such that

$$\sup_{w\in K} \mathcal{E}_k^\varepsilon(w) \le C_k \varepsilon^{1/2} \text{ for all small enough } \varepsilon > 0.$$

Thus, by (6.7.4), as claimed

$$\sup_{w\in K} |\mathrm{D}^k \mathcal{F}_\varepsilon^\star(w) - \mathrm{D}^k \mathcal{F}_0^\star(w)| \le C_k' \varepsilon^{1/2} + C_0' \varepsilon^{1/2} \in O(\varepsilon^{1/2}) \text{ as } \varepsilon \to 0. \quad \blacksquare$$

Lemma 6.7.3. *Suppose that the (spatial) Hessian $\mathrm{D}^2 E$ of E is uniformly elliptic, i.e. there exists $c > 0$ such that*

$$\mathrm{D}^2 E(t,x)(v,v) \ge c|v|^2 \text{ for all } t \in [0,T), \text{ and all } x, v \in \mathbb{R}^n,$$

and that $\|\partial_t \mathrm{D}E\|_{L^\infty} < +\infty$. Then (6.4.6) holds.

Proof. The energy evolution equation for $\mathcal{F}_0^\star$ along x^0 can be calculated using the chain rule, yielding

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}t} \mathcal{F}_0^\star\big(\mathrm{D}E(t,x^0(t))\big) \\
&\quad = -\left\langle \mathrm{D}^2 E(t,x^0(t)), \mathrm{D}\mathcal{F}_0^\star(\mathrm{D}E(t,x^0(t)))^{\otimes 2} \right\rangle \\
&\qquad + \left\langle \partial_t \mathrm{D}E(t,x^0(t)), \mathrm{D}\mathcal{F}_0^\star(\mathrm{D}E(t,x^0(t))) \right\rangle \\
&\quad \le -c \left|\mathrm{D}\mathcal{F}_0^\star\big(\mathrm{D}E(t,x^0(t))\big)\right|^2 + \|\partial_t \mathrm{D}E\|_{L^\infty} \left|\mathrm{D}\mathcal{F}_0^\star\big(\mathrm{D}E(t,x^0(t))\big)\right|
\end{aligned}$$

Theorem 6.3.3(5) implies that if $\mathcal{F}_0^\star$ blows up along any curve (i.e. one that approaches $-\partial\mathscr{E}$ in the dual space), then so does $|\mathrm{D}\mathcal{F}_0^\star|$. However, the mean value theorem and the above calculation imply that $\mathcal{F}_0^\star$ must be decreasing when $|\mathrm{D}\mathcal{F}_0^\star|$ is large. This yields the desired contradiction. $\blacksquare$

Lemma 6.7.4. *Suppose that x^ε, $\varepsilon > 0$, and x^0 solve*

$$\begin{aligned}
\dot{x}^\varepsilon &= -\mathrm{D}\mathcal{F}_\varepsilon^\star(\mathrm{D}E(t,x^\varepsilon)), \\
\dot{x}^0 &= -\mathrm{D}\mathcal{F}_0^\star(\mathrm{D}E(t,x^0)),
\end{aligned}$$

with initial conditions $x^\varepsilon(0) = x^0(0) = x_0$ such that $-\mathrm{D}E(0,x_0) \in \mathring{\mathscr{E}}$, and that (6.4.6) holds. Then there exists a constant C such that

$$\sup_{t\in[0,T]} \left|x^\varepsilon(t) - x^0(t)\right| \le C\varepsilon^{1/2} \text{ for all small enough } \varepsilon > 0.$$

Proof. The strategy, of course, is to appeal to lemma 6.7.2 and Grönwall's inequality in the form of theorem B.1. Without loss of generality, assume that $\varepsilon > 0$ is small enough that the conclusion of lemma 6.7.2 holds.

First, it is claimed that there exists a $K \subset\subset \mathscr{E}$ such that $-\mathrm{D}E(t, x^0(t)) \in K$ for all $t \geq 0$, i.e. that

$$\inf_{t\in[0,T]} \operatorname{dist}\big(-\mathrm{D}E(t, x^0(t)), \partial\mathscr{E}\big) > 0.$$

For a contradiction, suppose not. Then, since $\mathcal{F}_0^\star$ blows up to $+\infty$ on $\partial\mathscr{E}$, this would imply that $t \mapsto \mathcal{F}_0^\star(\mathrm{D}E(t, x^0(t)))$ blows up to $+\infty$. This, however, is ruled out by (6.4.6).

Thus, by lemma 6.7.2, there exists $C > 0$ such that

$$\sup_{t\in[0,T]} \big|\mathrm{D}\mathcal{F}_\varepsilon^\star(t, x^0(t)) - \mathrm{D}\mathcal{F}_0^\star(t, x^0(t))\big| \leq C\varepsilon^{1/2}.$$

Hence, by Grönwall's inequality, for all $t \in [0, T]$,

$$|x^\varepsilon(t) - x^0(t)| \leq \frac{C}{L}\varepsilon^{1/2},$$

where L is at most the product of the (finite) Lipschitz constants for $\mathrm{D}E$ and $\mathrm{D}\mathcal{F}_0^\star|_K$. ■

Lemma 6.7.5. *Consider a uniform partition P of $[0, T]$ with $[\![P]\!] = h > 0$. Let X be the Markov chain generated by E (convex) and Ψ (as usual) with $\varepsilon = h$, and assume the convexity/monotonicity property* (6.4.5)*. Let y be the Euler approximation to*

$$\dot{x} = -\mathrm{D}\mathcal{F}_h^\star(\mathrm{D}E(t, x(t)))$$

given by

$$\Delta y_i := -h\mathrm{D}\mathcal{F}_h^\star(\mathrm{D}E(t_{i+1}, y_i)),$$

with $X_0 = y_0$ such that $-\mathrm{D}E(0, y_0) \in \mathring{\mathscr{E}}$. Then, for every $\lambda > 0$,

$$\mathbb{P}\left[\max_{0\leq i\leq T/h} \big|X_i - y_i\big| \geq \lambda\right] \in O(h) \text{ as } h \to 0. \tag{6.7.6}$$

Proof. In order to simplify the notation, assume that the partition P is a uniform partition with $[\![P]\!] = h > 0$, and define a time-dependent vector

field f_h by $f_h(t,x) := -\mathrm{D}\mathcal{F}_h^\star(\mathrm{D}E(t,x))$. Let

$$K(t) := \{x \in \mathcal{S}(t) \mid \mathrm{dist}(x, \partial\mathcal{S}(t)) > \delta\}$$

with $\delta > 0$ smaller than the modulus of continuity for ℓ with step h and also small enough that $y_t \in K(t)$ for all $t \in [0,T]$. Write (dropping the superscript that indicates the partition P or its mesh size h)

$$\begin{aligned} X_{i+1} &= X_i + hf_h(t_{i+1}, X_i) + \Xi_{i+1}(X_i), \\ y_{i+1} &= y_i + hf_h(t_{i+1}, y_i). \end{aligned}$$

By lemma 6.3.1, for each x, $\Xi_{i+1}(x)$ is a random variable with mean 0 and k^{th} central moment at most $C_k(x)h^k$. The "errors" $Z := X - y$ satisfy

$$Z_{i+1} = Z_i + h\big(f_h(t_{i+1}, X_i) - f_h(t_{i+1}, y_i)\big) + \Xi_{i+1}(X_i). \tag{6.7.7}$$

Recall that the vector field $f_h(t_{i+1}, \cdot)$

(M) is a monotonically decreasing vector field on $\mathcal{S}(t_{i+1})$;

(B) is bounded on compactly-embedded subsets of $\mathcal{S}(t_{i+1})$;

and also that

(Z) for every x, $\mathbb{E}[\Xi_{i+1}(x)] = 0$.

Let $\mathcal{K}_i$ be (the σ-algebra generated by) the event that $X_j \in K(t_{j+1})$ for $0 \le j \le i$. Applying the conditional expectation operator $\mathbb{E}[-|\mathcal{K}_i]$ (which is never conditioning on an event of zero probability) to the square of (6.7.7) yields:

$$\begin{aligned} &\mathbb{E}\big[|Z_{i+1}|^2\big|\mathcal{K}_i\big] - \mathbb{E}\big[|Z_i|^2\big|\mathcal{K}_i\big] \\ &\quad = 2h\mathbb{E}\big[(f_h(t_{i+1}, X_i) - f_h(t_{i+1}, y_i)) \cdot Z_i\big|\mathcal{K}_i\big] \quad \le 0 \text{ by (M)} \\ &\qquad + 2\mathbb{E}\big[Z_i \cdot \Xi_{i+1}(X_i)\big|\mathcal{K}_i\big] \quad = 0 \text{ by (Z)} \\ &\qquad + 2h\mathbb{E}\big[(f_h(t_{i+1}, X_i) - f_h(t_{i+1}, y_i)) \cdot \Xi_{i+1}(X_i)\big|\mathcal{K}_i\big] \quad = 0 \text{ by (Z)} \\ &\qquad + h^2\mathbb{E}\big[|f_h(t_{i+1}, X_i) - f_h(t_{i+1}, y_i)|^2\big|\mathcal{K}_i\big] \quad \le Ch^2 \text{ by (B)} \\ &\qquad + \mathbb{E}\big[|\Xi_{i+1}(X_i)|^2\big|\mathcal{K}_i\big] \quad \le Ch^2 \text{ by lemma 6.3.1} \\ &\quad \le Ch^2. \end{aligned}$$

Applying the unconditional expectation operator to both sides yields the following uniform bound for the second moment of the deviations:

$$\max_{0\le i\le T/h} \mathbb{E}\big[|Z_i|^2\big] \le CTh. \tag{6.7.8}$$

The inequality (6.7.8) is insufficient to establish (6.7.6), but can be used to "bootstrap" a similar inequality for the fourth moments. Define a tetralinear form $\langle\cdot,\cdot,\cdot,\cdot\rangle\colon \mathbb{R}^n \to \mathbb{R}$ by

$$\langle w, x, y, z\rangle := (w\cdot x)(y\cdot z), \tag{6.7.9}$$

so that $|x|^4 = \langle x, x, x, x\rangle$. This tetralinear form is invariant under arbitrary compositions of the following interchanges of entries: $(1,2)$, $(3,4)$ and $(1,3)(2,4)$. The Cauchy–Bunyakovskiĭ–Schwarz inequality for the Euclidean inner product implies a corresponding inequality for this tetralinear form: for all $w, x, y, z \in \mathbb{R}^n$,

$$|\langle w, x, y, z\rangle| \le |w||x||y||z|. \tag{6.7.10}$$

Hence, $\mathbb{E}\big[|Z_{i+1}|^4\big] \equiv \mathbb{E}\big[\mathbb{E}\big[|Z_{i+1}|^4\big|\mathcal{K}_i\big]\big]$ can be expanded using the tetralinear form (6.7.9) and (6.7.7) and each term estimated as in the derivation of (6.7.8). By (Z), those terms containing precisely one $\Xi_{i+1}(X_i)$ have zero expectation; the terms of the form

$$\mathbb{E}\left[\,\langle Z_i, Z_i, Z_i, h(f_h(t_{i+1}, X_i) - f_h(t_{i+1}, y_i))\rangle\big|\,\mathcal{K}_i\right]$$

are non-positive by (M); the remaining terms can all be estimated using (B), (6.7.8), (6.7.10) and lemma 6.3.1, with the worst bound being $O(h^3)$. Thus, the following uniform bound for the fourth moment of the deviations holds:

$$\max_{0\le i\le T/h} \mathbb{E}\big[|Z_i|^4\big] \le CTh^2. \tag{6.7.11}$$

Hence, for $\lambda > 0$,

$$\begin{aligned}
&\mathbb{P}\big[|Z_i| \geq \lambda \text{ for some } 0 \leq i \leq T/h\big] \\
&\quad \leq \sum_{i=0}^{T/h} \mathbb{P}\big[|Z_i| \geq \lambda\big] \\
&\quad \leq \sum_{i=0}^{T/h} \lambda^{-4}\mathbb{E}\big[|Z_i|^4\big] \text{ by the Bienayme–Chebyshëv inequality} \\
&\quad \leq \lambda^{-4} C T^2 h \text{ by (6.7.11)},
\end{aligned}$$

which establishes (6.7.6) and completes the proof. ■

Chapter 7

Conclusions and Outlook

7.1 Summary and Conclusions

As noted in the abstract and introduction, this thesis has focused on two extremal scaling regimes in the problem of a gradient descent in a random energy landscape subject to a heat bath. The first regime was the regime in which the microstructural variations were dominant; the second was the regime in which the thermal effects were dominant and superimposed on the effective macroscopic gradient descent for the microstructure. The intermediate regime has been left for further work — see the next section.

In chapter 3, it was shown that a one-dimensional rate-independent process arises as the limit process of a classical gradient descent a smooth, convex energetic potential subject to suitable random perturbation. Inertial effects were neglected and the limit was taken as the scale on which the microstructure varied and the relaxation time both converged to zero. The one-dimensional result was essentially deterministic, with the case of random microstructure following as a corollary. In higher dimension (chapter 4), though, the scaling result required a more specific model of the random perturbation and some truly probabilistic analysis. In both cases, though, one of the key points is that the microstructure prevents the fast relaxation time from causing the total variation of the process to blow up. How to extend these results to infinite-dimensional settings, the natural setting for many problems in elastoplasticity, is not immediately clear since the model class in the $\mathbb{R}^n$ case was quite restricted, and does not generalize easily to infinite-dimensional spaces.

Chapter 5 presented a heuristic derivation for a model of the effect of a

heat bath on any gradient descent, justified by appeal to the fact that it generates well-known stochastic differential equations in the classical cases. The case of one-homogeneous dissipation (i.e. an underlying rate-independent system) in $\mathbb{R}^n$ was the topic of chapter 6. The results of chapter 6 showed that the coupling to the heat path destroys the original rate-independence in a controlled way, and that the quantitative "softening" of the system's response can be captured by a simple non-linear transformation of the original dissipation potential. This approach appears to be robust with respect to dimension and geometry, and may be extended to non-Euclidean and/or infinite-dimensional state spaces.

7.2 Directions for Further Research: Chapter 4.5?

For the most part, directions for further research have been outlined in the relevant chapters. However, one other obvious topic for future work lies in the white space between chapters 4 and 5. The work before this point established that a differential inclusion of the form

$$\partial\Psi(\dot{z}(t)) = -\mathrm{D}E(t, z(t))$$

is the macroscopic limiting behaviour of a classical gradient descent

$$\dot{z}_\varepsilon(t) = -\frac{1}{\varepsilon}\nabla E_\varepsilon(t, z_\varepsilon(t))$$

in a suitable perturbation E_ε of E. After this point, the thermalized gradient descent model for the effect of a heat bath was applied to the limiting differential inclusion directly. This step was justified by the assumption that the thermal effects were dominant, the agreement of results like theorem 5.4.1 with other theoretical elements, and the ability of the results of chapter 6 to make exact physical predictions consistent with observed physical phenomena (Andrade creep).

A more sensitive analysis would provide a mathematically complete picture. It should be possible to show that the solutions to the random stochastic differential equation

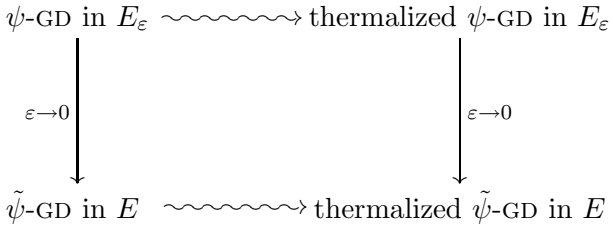
$$\ddot{z}_\varepsilon(t) = -\eta\dot{z}_\varepsilon(t) - \rho\nabla E_\varepsilon(t, z_\varepsilon(t)) + \sigma\dot{W}(t) \tag{7.2.1}$$

converge as $\varepsilon \to 0$ to the limiting gradient descent

$$\mathrm{D}\mathcal{F}_0(-\dot{x}(t)/\theta) = \mathrm{D}E(t, x(t)). \tag{7.2.2}$$

Here η is a viscosity parameter, ρ is a rate, and σ is a diffusivity; these may need to be scaled appropriately with respect to ε. It is reasonable to expect that this objective can be accomplished using the tools of averaging and homogenization theory [BLP78] [PS08] [ZKO94]. If the law of $\lim_{\varepsilon\to 0} z_\varepsilon$ is non-singular, but has the solution of (7.2.2) as its mean, then so much the better, since this would provide a truly positive-temperature description of the limiting system, unlike the diagonal limit employed in chapter 6.

Alternatively, since the thermalized gradient descent for 2-homogeneous dissipation coincides with the usual interpretation of noise in the sense of stochastic calculus, the objective of "chapter 4.5" can be seen as a study of the stability of the thermalized gradient descent method: that is, under what circumstances does the following diagram "commute"?



Appendices

A Symbols and Notation

Symbol	Denotes
$\mathbb{1}$	Identity operator.
$\mathbb{1}_A$	Indicator function of a measurable set A; *cf.* χ_A.
$\arg\min \mathcal{J}$	The set of global minimizers of a functional $\mathcal{J}$.
$\mathbb{B}_r(p)$	Open ball of radius r about a point p.
c_n	n-dimensional Lebesgue measure of the Euclidean unit ball in $\mathbb{R}^n$; $c_n = \pi^{n/2}/\Gamma(1+\frac{n}{2})$.
D	Fréchet/Gâteaux derivative.
D_{KL}	Kullback–Leibler divergence (relative entropy); see (5.3.4).
$\partial\Psi$	The set-valued subdifferential of a convex function Ψ.
E	Energetic potential.
$\mathbb{E}[X]$	Expectation of a random variable X.
$\mathscr{E}$	Elastic region for a given dissipative potential; see (2.4.1).
$f^\star$	Convex conjugate of f; see (C.2).
$\mathcal{F}_0$	Effective dissipation potential; $\mathcal{F}_0 := (\mathcal{F}_0^\star)^\star$.
$\mathcal{F}_\varepsilon^\star, \mathcal{F}_0^\star$	Effective dual dissipation potential; see (6.3.1).
$\mathscr{H}^s_{\mathcal{X}}$	s-dimensional Hausdorff measure on a space $\mathcal{X}$, $s \geq 0$.
K-lim	Kuratowski limit; see definition 3.5.1.
$\mathscr{M}(\mathcal{X})$	The space of Radon measures on a given space $\mathcal{X}$.
$\mathbb{N}, \mathbb{N}_0$	The natural numbers, with $\mathbb{N}_0 = \mathbb{N} \uplus \{0\}$.
$\mathscr{P}(\mathcal{X})$	Space of Borel probability measures on a given space $\mathcal{X}$.
$\mathscr{P}_{\preceq\pi}(\mathcal{X})$	Those measures in $\mathscr{P}(\mathcal{X})$ that admit a probability density function with respect to π.

$\mathcal{P}([a,b])$	The set of finite ordered partitions of an interval $[a,b]$.
$\mathcal{S}, \mathcal{S}_{\text{loc}}$	The set of (globally/locally) stable states for a given energy and dissipative potential; see definition 2.4.6.
sgn	Signum/direction function on a normed vector space: $\operatorname{sgn}(x) := x/\|x\|$ for $x \neq 0$, and $\operatorname{sgn}(0) := 0$.
$\operatorname{Var}_{[a,b]}(u)$	Variation seminorm of $u\colon [a,b] \to \mathcal{X}$.
$\operatorname{Var}[X]$	Variance (second central moment) of a random variable X.
W	Standard Wiener process (Brownian motion).
$\mathscr{W}$	Incremental work function.
$z(t\pm)$	Right/left limit of z at t: $z(t\pm) := \lim_{s \searrow 0} z(t \pm s)$.
$\mathcal{Z}_\varepsilon, \mathcal{Z}_0$	Effective "partition function"; $\mathcal{Z}_\varepsilon(w) \equiv \exp \mathcal{F}_\varepsilon(w)$.
γ, γ^n	Standard Gaussian measure on $\mathbb{R}^n$, the same as $\mathcal{N}(0, \mathbb{1})$.
Δ	Backward-difference operator acting on sequences in space or time: $\Delta x_{i+1} := x_{i+1} - x_i$.
λ^n	Lebesgue measure on $\mathbb{R}^n$.
ϖ_f	Modulus of continuity for a function $f\colon X \to Y$: for $\delta > 0$, $\varpi_f(\delta) := \sup\{d_Y(f(x), f(y)) \mid d_X(x,y) \leq \delta\}$.
χ_K	Convex characteristic function of a convex set K, see (C.6); *cf.* $\mathbb{1}_A$.
Ψ	Dissipation potential, $\Psi = \chi^\star_{\mathscr{E}}$.
$\preceq$	Absolute continuity of measures; see (5.3.1).
$\overset{O(f)}{=}$	Equality up to an error in $O(f)$; similarly for $\overset{O(f)}{\leq}$.
$\lvert\cdot\rvert$	Absolute value on $\mathbb{R}$; the Euclidean norm on $\mathbb{R}^n$.
$\lvert\cdot\rvert_p$	p-norm on $\mathbb{R}^n$; subscript usually omitted for $p = 2$.
$[\![P]\!]$	The mesh of a partition $P \in \mathcal{P}([0,T])$; see (2.3.2).
$\langle \mathcal{X}^*, \mathcal{X} \rangle$	Dual pairing between a topological vector space $\mathcal{X}$ and its continuous dual space $\mathcal{X}^*$.

B Unfamiliar Forms of Two Familiar Results

B.1 Grönwall's Inequality

The original and usual formulation of Grönwall's inequality [Grö19] for ordinary differential equations is essentially an estimate of how two initial conditions evolve under a single vector field. However, with only a little more effort one can obtain a much more powerful result that allows the two flows to have different vector fields as well as different initial conditions. For the proof of this formulation of Grönwall's inequality, see Howard's excellent classroom note [How98].

Theorem B.1 (Grönwall's inequality)**.** *Let U be an open subset of a Banach space $(\mathcal{X}, \|\cdot\|)$. Let $f, g\colon [a,b] \times U \to \mathcal{X}$ be continuous and suppose that $y, z\colon [a,b] \to U$ solve the initial value problems*

$$\begin{cases} \dot{y}(t) = f(t, y(t)), \\ y(a) = y_0 \in U; \end{cases} \qquad \textit{and} \qquad \begin{cases} \dot{z}(t) = g(t, z(t)), \\ z(a) = z_0 \in U. \end{cases}$$

Suppose that there exists a continuous function $\phi\colon [a,b] \to \mathbb{R}$ and a constant $L \geq 0$ such that, for all $t \in [a,b]$,

$$\|f(t, y(t)) - g(t, y(t))\| \leq \phi(t).$$

$$\|g(t, x_1) - g(t, x_2)\| \leq L\|x_1 - x_2\|,$$

Then, for all $t \in [t_1, t_2]$,

$$\|y(t) - z(t)\| \leq e^{L|t-a|}\|y_0 - z_0\| + e^{L|t-a|} \int_a^t e^{-L|s-a|}\phi(s)\,\mathrm{d}s. \tag{B.1}$$

B.2 Lebesgue's Dominated Convergence Theorem

Any respectable introductory course in measure theory covers Lebesgue's dominated convergence theorem, and quite rightly points out that it relies only upon pointwise almost everywhere convergence of the integrands. What is not always stressed is that the theorem is essentially a "continuity under the integral sign" result in the same spirit as differentiation under the integral sign.

Theorem B.2 (Lebesgue's dominated convergence theorem)**.** *Let $\mathcal{X}$ be a first-countable topological space and let $(\Omega, \mathscr{F}, \mu)$ be a measure space. Let $f\colon \mathcal{X} \times \Omega \to \mathbb{R}$ be such that*

1. *for μ-almost all ω, $f(\cdot, \omega)\colon \mathcal{X} \to \mathbb{R}$ is continuous;*
2. *for every $x \in \mathcal{X}$, $f(x, \cdot)\colon \Omega \to \mathbb{R}$ is measurable;*
3. *for all $x \in \mathcal{X}$, there is a function $\Theta_x \in L^1(\Omega, \mathscr{F}, \mu; \mathbb{R})$ and a neighbourhood $U[x]$ of x in $\mathcal{X}$ such that, for all $y \in U[x]$ and μ-almost all $\omega \in \Omega$, $|f(y, \omega)| \leq \Theta_x(\omega)$.*

Then the function $\mathcal{X} \to \mathbb{R}\colon x \mapsto \int_\Omega f(x, \omega)\,\mathrm{d}\mu(\omega)$ is continuous.

Proof. Since $\mathcal{X}$ is a first-countable space, continuity is equivalent to sequential continuity, and $V \subseteq \mathcal{X}$ is open if, and only if, for any sequence that converges to a point of V eventually lies in V. Fix any point $x_0 \in \mathcal{X}$ and any sequence $(x_n)_{n\in\mathbb{N}}$ converging to x_0. Note that, for large enough n, $x_n \in U[x_0]$, where $U[x_0]$ is as in the statement of the theorem. It is required to show that

$$\int_\Omega f(x, \omega)\,\mathrm{d}\mu(\omega) = \lim_{n\to\infty} \int_\Omega f(x_n, \omega)\,\mathrm{d}\mu(\omega),$$

but this is exactly the conclusion of the usual formulation of Lebesgue's dominated convergence theorem for the sequence of functions $f_n\colon \Omega \to \mathbb{R}$ given by $f_n(\omega) := f(x_n, \omega)$ for $n \in \mathbb{N}_0$, with dominating function Θ_{x_0}, since, by the continuity of $f(\cdot, \omega)$, for μ-almost all $\omega \in \Omega$,

$$f_n(\omega) = f(x_n, \omega) \xrightarrow[n\to\infty]{} f(x_0, \omega) = f_0(\omega). \qquad \blacksquare$$

In general, theorem B.2 does not hold if $\mathcal{X}$ is replaced by a general topological space. The reason for this is that Lebesgue's dominated convergence theorem requires Fatou's lemma, which in turn rests on that fact that the limit inferior of a *countable* family of measurable functions is a measurable function — and this last property can fail for uncountable families. In general topological spaces, continuity may be inequivalent to sequential continuity but *is* equivalent to continuity for nets (Moore–Smith sequences). However, a net may use an uncountable index set, in which case Fatou's lemma and its consequents are no longer guaranteed to hold.

C Convex Analysis

Standard references on convex analysis include [BV04], [ET99] and [Roc70]. The elements of convex analysis that are needed in this thesis are the basic notions of convexity and semicontinuity, and the duality relationships.

C.1 Convexity

Definition C.1. Let $\mathcal{X}$ be a real vector space. $K \subseteq \mathcal{X}$ is said to be *convex* if, whenever $x_0, x_1 \in K$ and $t \in [0,1]$,

$$x_t := (1-t)x_0 + tx_1 \in K.$$

If K is convex, then a function $f\colon K \to \mathbb{R} \cup \{\pm\infty\}$ is said to be *convex* if

$$f(x_t) \leq (1-t)f(x_0) + tf(x_1) \text{ for all } x_0, x_1 \in K, t \in [0,1],$$

and *strictly convex* if

$$f(x_t) < (1-t)f(x_0) + tf(x_1) \text{ for all } x_0, x_1 \in K, t \in (0,1);$$

if, in addition, $\mathcal{X}$ is normed, then f is said to be α-*uniformly convex* if

$$f(x_t) \leq (1-t)f(x_0) + tf(x_1) - \frac{\alpha t(1-t)}{2}\|x_0 - x_1\|^2$$

for all $x_0, x_1 \in K$, $t \in [0,1]$.

Example C.2. Any (positive multiple of the) norm induced by an inner product is a uniformly convex function, as the following shows. Let $E(x) := \frac{\kappa}{2}\langle x, x\rangle$ with $\kappa > 0$ for some inner product $\langle\cdot,\cdot\rangle$. Then

$$\begin{aligned}
&E(x_t)\\
&\quad= \frac{\kappa}{2}\left((1-t)^2\|x_0\|^2 + t^2\|x_1\|^2 + 2t(1-t)\langle x_0, x_1\rangle\right)\\
&\quad= \frac{\kappa}{2}\left((1-t)^2\|x_0\|^2 + t^2\|x_1\|^2 + t(1-t)\left(\|x_0\|^2 + \|x_1\|^2 - \|x_0 - x_1\|^2\right)\right)\\
&\quad\leq (1-t)E(x_0) + tE(x_1) - \frac{\kappa t(1-t)}{2}\|x_0 - x_1\|^2.
\end{aligned}$$

From the point of view of gradient descent theory, the advantage of convex functions is the well-known property that their gradients are monotone

vector fields. Indeed, this property is so well-known that the following result, due to Kachurovskiĭ [Sho97], is often not even referred to by name.

Theorem C.3 (Kachurovskiĭ)**.** *Let K be a convex subset of a Banach space $\mathcal{X}$ and let $f\colon K \to \mathbb{R} \cup \{+\infty\}$ be an extended real-valued function that is Fréchet differentiable with derivative $\mathrm{D}f\colon K \to \mathcal{X}^*$. Then the following are equivalent:*

1. *f is a convex function;*
2. *for all $x, y \in K$, $\langle \mathrm{D}f(x), y - x\rangle \leq f(y) - f(x)$;*
3. *$\mathrm{D}f$ is an (increasing) monotone operator, i.e., for all $x, y \in K$,*

$$\langle \mathrm{D}f(x) - \mathrm{D}f(y), x - y\rangle \geq 0.$$

C.2 Semicontinuity

Definition C.4. Let $\mathcal{X}$ be any topological space. A function $f\colon \mathcal{X} \to \mathbb{R} \cup \{\pm\infty\}$ is said to be *lower semicontinuous at $x \in \mathcal{X}$* if, for every $\varepsilon > 0$, there exists a neighbourhood U of x such that

$$y \in U \implies f(y) \geq f(x) - \varepsilon;$$

equivalently, f is lower semicontinuous at x if

$$\liminf_{y \to x} f(y) \geq f(x).$$

f is said to be *lower semicontinuous* if it is lower semicontinuous at every point of its domain; equivalently, f is lower semicontinuous if all its sublevel sets are closed, i.e.

$$\alpha \in \mathbb{R} \implies \{x \in \mathcal{X} \mid f(x) \leq \alpha\} \text{ is closed in } \mathcal{X}.$$

Lemma C.5. 1. *Pointwise (resp. global) continuity implies pointwise (resp. global) lower semicontinuity.*

2. *If $\mathcal{X}$ is a compact space and $f\colon \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ is lower semicontinuous, then f attains a minimum somewhere in $\mathcal{X}$*

3. *$f\colon \mathcal{X} \to \mathbb{R} \cup \{\pm\infty\}$ is lower semicontinuous if, and only if, its epigraph is closed in $\mathcal{X} \times (\mathbb{R} \cup \{\pm\infty\})$ with the product topology, where*

$$\operatorname{epi} f := \{(x, v) \in \mathcal{X} \times (\mathbb{R} \cup \{\pm\infty\}) \,|\, v \geq f(x)\}\,. \tag{C.1}$$

4. *Let $I \neq \emptyset$ be an index set and suppose that $f_i\colon \mathcal{X} \to \mathbb{R} \cup \{\pm\infty\}$ is a lower semicontinuous function for every $i \in I$. Then the pointwise supremum*

$$x \mapsto \sup_{i \in I} f_i(x)$$

is lower semicontinuous.

5. *If $\mathcal{X}$ is a uniform space (e.g. a metric space), then $f\colon \mathcal{X} \to \mathbb{R} \cup \{\pm\infty\}$ is lower semicontinuous if, and only if, it can be written as a pointwise supremum of a family of continuous functions.*

C.3 Duality

Now let $\mathcal{X}$ be a topological vector space and let $\mathcal{X}^*$ denote the continuous dual space; for simplicity, assume that $\mathcal{X}$ is reflexive, i.e. $\mathcal{X} \cong \mathcal{X}^{**}$.

Definition C.6. Let $f\colon \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ be any function. The *convex conjugate* (also known as the *Legendre–Fenchel transform* or *Fenchel–Moreau conjugate*) of f is $f^\star\colon \mathcal{X}^* \to \mathbb{R} \cup \{+\infty\}$ defined by

$$f^\star(x^*) := \sup_{x \in \mathcal{X}} \big(\langle x^*, x\rangle - f(x)\big). \tag{C.2}$$

For any f, $f^\star$ is a convex function. If f is convex and lower semicontinuous, then $f^{\star\star} = f$; for any f, $f^{\star\star}$ is the greatest convex lower semicontinuous function less than or equal to f. A convex, lower semicontinuous function f and its convex conjugate $f^\star$ always satisfy *Fenchel's inequality*:

$$\langle x^*, x\rangle \leq f^\star(x^*) + f(x) \text{ for all } x \in \mathcal{X}, x^* \in \mathcal{X}^*. \tag{C.3}$$

Definition C.7. Given a function $f\colon \mathcal{X} \to \mathbb{R}_{+\infty}$, the *subdifferential* $\partial f(x)$ of f at x is defined to be

$$\partial f(x) := \big\{\ell \in \mathcal{X}^* \big| f(x) + \langle \ell, y - x\rangle \leq f(y) \text{ for all } y \in \mathcal{X}\big\}.$$

Theorem C.8. *Let $f\colon \mathcal{X} \to \mathbb{R}\cup\{+\infty\}$ be convex and lower semicontinuous, $x \in \mathcal{X}$ and $x^* \in \mathcal{X}^*$. Then the following are equivalent:*

1. $x^* \in \partial f(x)$;
2. *for all* $w \in \mathcal{X}$, $f(x+w) \geq f(x) + \langle x^*, w\rangle$;
3. $x \in \arg\max\{\langle x^*, w\rangle - f(w) \mid w \in \mathcal{X}\}$;
4. $\langle x^*, x\rangle = f^\star(x^*) + f(x)$ *(cf. Fenchel's inequality* (C.3)*)*;
5. $x^* \in \arg\max\{\langle w^*, x\rangle - f^\star(w^*) \mid w^* \in \mathcal{X}^*\}$;
6. *for all* $w^* \in \mathcal{X}^*$, $f^\star(x^*+w^*) \geq f^\star(x^*) + \langle w^*, x\rangle$;
7. $x \in \partial f^\star(x^*)$.

Suppose that $K \subseteq \mathcal{X}$ is a closed and convex set. Let $\mathrm{T}_x^* K$ denote the *inward tangent cone* to K at x:

$$\mathrm{T}_x^* K := \overline{\{w \in \mathcal{X} \mid \text{for some } r > 0, x + rw \in K\}}. \tag{C.4}$$

Let $\mathrm{N}_x^* K$ denote the *outward normal cone* to K at x:

$$\mathrm{N}_x^* K := \{y^* \in \mathcal{X}^* \mid \langle \text{for all } w \in \mathrm{T}_x K, y^*, w\rangle \leq 0\}. \tag{C.5}$$

Let $\chi_K\colon \mathcal{X} \to \mathbb{R}\cup\{+\infty\}$ denote the convex characteristic function of K:

$$\chi_K(x) := \begin{cases} 0, & \text{if } x \in K, \\ +\infty, & \text{if } x \in \mathcal{X}\setminus K. \end{cases} \tag{C.6}$$

The subdifferential of χ_K may be succinctly expressed using normal cones:

$$\partial\chi_K(x) = \mathrm{N}_x^* K.$$

Similarly, if $\mathscr{E} \subseteq \mathcal{X}^*$ is closed, convex and bounded, with $0 \in \mathring{\mathscr{E}}$, and $\Psi := \chi_{\mathscr{E}}^\star$, then

$$\partial\Psi(x) = \operatorname*{arg\,max}_{y^*\in\mathscr{E}} \langle y^*, x\rangle = \begin{cases} \mathscr{E}, & \text{if } x = 0, \\ \{z^* \in \mathscr{E} \mid \Psi(x) = \langle z^*, x\rangle\}, & \text{if } x \neq 0. \end{cases} \tag{C.7}$$

D Probability Measures

This section summarizes some of the central elements of the convergence theory of probability measures and random variables in metric spaces. Thorough treatments in the literature include the monographs of Billingsley [Bil99] and Parthasarathy [Par05]. In the category of Banach spaces, [LT91] is another comprehensive reference, and [Pol84] focuses on the convergence of stochastic processes.

For a topological space $\mathcal{X}$, let $\mathscr{P}(\mathcal{X})$ denote the set of all Borel probability measures on $\mathcal{X}$.

D.1 Convergence of Measures

Theorem D.1 (Portmanteau theorem for weak convergence)**.** *Let $\mathcal{X}$ be a separable topological space; let $(\mu_\alpha)_{\alpha\in A}$ be a net in $\mathscr{P}(\mathcal{X})$ and let $\mu \in \mathscr{P}(\mathcal{X})$. Then the following are all equivalent:*

1. *for all bounded and continuous functions $\phi\colon \mathcal{X} \to \mathbb{R}$,*

$$\lim_\alpha \int_{\mathcal{X}} \phi \,\mathrm{d}\mu_\alpha = \int_{\mathcal{X}} \phi \,\mathrm{d}\mu;$$

2. *for all closed subsets $F \subseteq \mathcal{X}$, $\limsup_\alpha \mu_\alpha(F) \leq \mu(F)$;*

3. *for all open subsets $G \subseteq \mathcal{X}$, $\liminf_\alpha \mu_\alpha(G) \geq \mu(G)$;*

4. *for all Borel subsets $S \subseteq \mathcal{X}$ with $\mu(\partial S) = 0$, $\lim_\alpha \mu_\alpha(S) = \mu(S)$.*

If one (and hence all) of the conditions of theorem D.1 holds, then the net $(\mu_\alpha)_{\alpha\in A}$ is said to *converge weakly* to μ, and this will be denoted $\mu_\alpha \rightharpoonup \mu$. Henceforth, $\mathscr{P}(\mathcal{X})$ will be assumed to be equipped with the topology of weak convergence.

Theorem D.2. *1. $\mathscr{P}(\mathcal{X})$ can be metrized as a separable metric space if, and only if, $\mathcal{X}$ is itself a separable metric space.*

2. *If $\mathcal{X}$ is a separable metric space and $E \subseteq \mathcal{X}$ is dense in $\mathcal{X}$, then*

$$\{\mu \in \mathscr{P}(\mathcal{X}) | \operatorname{supp}\mu \subseteq E \text{ and } \operatorname{supp}\mu \text{ is finite}\}$$

is dense in $\mathscr{P}(\mathcal{X})$.

3. *$\mathscr{P}(\mathcal{X})$ is a compact metric space if, and only if, $\mathcal{X}$ is itself a compact metric space.*

4. *If $\mathcal{X}$ is a separable metric space, then $\mathscr{P}(\mathcal{X})$ is a topologically complete space if, and only if, $\mathcal{X}$ is itself a topologically complete space.*

The compact subsets of of $\mathscr{P}(\mathcal{X})$ when $\mathcal{X}$ is a complete, separable metric space (a so-called *Polish space*) are characterized by a celebrated theorem of Prokhorov [Pro56]:

Definition D.3. A collection of probability measures $\mathcal{K} \subseteq \mathscr{P}(\mathcal{X})$ is said to be *uniformly tight* if, for all $\varepsilon > 0$, there exists a compact set $K_\varepsilon \subseteq \mathcal{X}$ such that, for all $\mu \in \mathcal{K}$, $\mu(\mathcal{X} \setminus K_\varepsilon) < \varepsilon$.

Theorem D.4 (Prokhorov)**.** *Let $\mathcal{X}$ be a Polish space. Then $\mathcal{K} \subseteq \mathscr{P}(\mathcal{X})$ is relatively compact if, and only if, it is uniformly tight.*

In the case of the classical Wiener space $\mathcal{C}^0([0,T];\mathbb{R}^n)$ of continuous functions from $[0,T]$ into $\mathbb{R}^n$, equipped with the uniform norm topology, Prokhorov's theorem and the Arzelà–Ascoli theorem together give a criterion for tightness.

Theorem D.5 (Arzelà–Ascoli)**.** *A subset K of $\mathcal{C}^0([0,T];\mathbb{R}^n)$ is relatively compact if, and only if, it is uniformly bounded and equicontinuous, i.e. there exists $R > 0$ such that $|f(t)| \leq R$ for all $f \in K$ and $t \in [0,T]$, and*

$$\text{for all } \varepsilon > 0, \text{ there exists } \delta > 0 \text{ such that } \sup_{f \in K} \varpi_f(\delta) \leq \varepsilon,$$

where ϖ_f denotes the modulus of continuity for f.

Theorem D.6 (Arzelà–Ascoli–Prokhorov)**.** *Let $(\mu_\alpha)_{\alpha \in A}$ be a net of probability measures on $\mathcal{C}^0([0,T];\mathbb{R}^n)$. Then $(\mu_\alpha)_{\alpha \in A}$ is uniformly tight if, and only if, both*

$$\lim_{R \to \infty} \limsup_\alpha \mu_\alpha\{f \mid |f(0)| \geq R\} = 0$$

and, for all $\theta > 0$,

$$\lim_{\delta \to 0} \limsup_\alpha \mu_\alpha\{f \mid \varpi_f(\delta) \geq \theta\} = 0.$$

D.2 Convergence of Random Variables

Definition D.7. A *random variable* X taking values in a measurable space $(\mathcal{X}, \mathscr{A})$ is simply a measurable function $X\colon \Omega \to \mathcal{X}$ for some probability space $(\Omega, \mathscr{F}, \mathbb{P})$. The *law* or *distribution* of X is the push-forward measure $X_*\mathbb{P}$ on $\mathcal{X}$:

$$(X_*\mathbb{P})(S) = \mathbb{P}[X \in S] := \mathbb{P}(X^{-1}(S)) \text{ for } S \in \mathscr{A}.$$

A net of $\mathcal{X}$-valued random variables $(X_\alpha)_{\alpha \in A}$, not necessarily all defined on the same probability space, is said to *converge weakly* to another random variable X if the laws of the X_α converge weakly to that of X.

Stronger notions of convergence of random variables require a notion of distance in the target space so that correlations between the random variables can be quantified. Furthermore, it is necessary to assume that the metric space $(\mathcal{X}, d)$ is separable, since this ensures that $d(X_\alpha, X)\colon \Omega \to \mathbb{R}$ is a measurable function, i.e. that, for every $\lambda > 0$,

$$[d(X_\alpha, X) \geq \lambda] := \{\omega \in \Omega \mid d(X_\alpha(\omega), X(\omega)) \geq \lambda\}$$

is a measurable subset of Ω. More advanced treatments drop the separability requirement on $\mathcal{X}$ itself and focus on random variables with almost-surely-separable range, i.e. *Radon random variables*. This criterion is closely related to questions of tightness/compactness: X is Radon if, and only if, its law is a tight measure on $\mathcal{X}$.

Definition D.8. Let $(X_\alpha)_{\alpha \in A}$, X be random variables defined on a common probability space and taking values in a separable metric space $(\mathcal{X}, d)$. Then X_α is said to *converge in probability* to X if, for all $\lambda > 0$,

$$\lim_\alpha \mathbb{P}\big[d(X_\alpha, X) \geq \lambda\big] = 0;$$

this will be denoted by $X_\alpha \xrightarrow[\alpha]{\mathbb{P}} X$ or $\mathbb{P}\text{-}\lim_\alpha X_\alpha = X$. For $p > 0$, X_α is said to *converge in* p^{th} *mean* to X if

$$\lim_\alpha \mathbb{E}\big[d(X_\alpha, X)^p\big] = 0;$$

this will be denoted by $X_\alpha \xrightarrow[\alpha]{L^p} X$ or $L^p\text{-}\lim_\alpha X_\alpha = X$. Finally, X_α is said

to *converge* $\mathbb{P}$*-almost surely* to X if

$$\mathbb{P}\left[\lim_{\alpha} X_\alpha = X\right] = 1;$$

this will be denoted by $X_\alpha \xrightarrow[\alpha]{\mathbb{P}\text{-a.s.}} X$ or $\lim_\alpha X_\alpha = X$ $\mathbb{P}$-a.s..

Convergence in probability defines a topology on the space of $\mathcal{X}$-valued random variables: in fact, this topology is metrizable and a metric ρ that induces the topology of convergence in probability is given by

$$\rho(X, Y) := \mathbb{E}[\min\{1, d(X, Y)\}].$$

In [LT91], this metric is referred to as the L^0 metric; it is, in some sense, the L^p space in which *every* random variable has been made integrable by taking the minimum with 1. However, it is a curious (even pathological) fact that there is, in general, no such thing as a topology of almost sure convergence. An example of a probability space that admits no topology of almost sure convergence is given by $[0, 1]$ with the uniform (Lebesgue) measure, as illustrated in [Ord66]; indeed, Ordman's counterexample is applicable to any probability space that is not completely atomic.

The various modes of convergence are related as listed in the following theorem. In general, the converse implications do not hold: point 6 is one of the few converse implications that do hold.

Theorem D.9. *Let $(X_\alpha)_{\alpha\in A}$, X be random variables defined on a common probability space and taking values in a metric space $(\mathcal{X}, d)$. Then*

1. $X_\alpha \xrightarrow{\mathbb{P}\text{-a.s.}} X \implies X_\alpha \xrightarrow{\mathbb{P}} X$;
2. $X_\alpha \xrightarrow{L^p} X$ *for some* $p > 0 \implies X_\alpha \xrightarrow{\mathbb{P}} X$;
3. *for* $r > s \geq 1$, $X_\alpha \xrightarrow{L^r} X \implies X_\alpha \xrightarrow{L^s} X$;
4. $X_\alpha \xrightarrow{\mathbb{P}} X \implies X_\alpha \rightharpoonup X$;
5. $X_\alpha \xrightarrow{\mathbb{P}} X \implies$ *every subsequence of* (X_α) *has a further subsequence that converges to* X $\mathbb{P}$*-almost surely;*
6. $X_\alpha \rightharpoonup X$ *and* $X = x$ $\mathbb{P}$-a.s. $\implies X_\alpha \xrightarrow{\mathbb{P}} X$.

Another converse implication is an application of the Borel–Cantelli lemma, which shows that sufficiently fast convergence in probability implies almost sure convergence:

Theorem D.10. *Let $(X_k)_{k\in\mathbb{N}}$, X be random variables defined on a common probability space and taking values in a metric space $(\mathcal{X}, d)$, and suppose that, for every $\delta > 0$,*

$$\sum_{k\in\mathbb{N}} \mathbb{P}\big[d(X_k, X) > \delta\big] < +\infty.$$

Then $\lim_{k\in\infty} X_k = X$ $\mathbb{P}$-almost surely.

The following maximal inequality, known as Doob's submartingale inequality, bounds the probability that a sequence of random variables exceeds some value λ in terms of the expected value of the last term of the sequence. It includes as special cases the well-known Bienayme–Chebyshëv and Kolmogorov inequalities.

Theorem D.11 (Doob's submartingale inequality)**.** *Let $(X_k)_{k\in\mathbb{N}}$ be a sequence of non-negative $\mathbb{R}$-valued random variables satisfying the submartingale inequality that, for each $k \in \mathbb{N}$,*

$$\mathbb{E}\big[X_{k+1}\big|X_1, \ldots, X_k\big] \geq X_k.$$

Then, for any $n \in \mathbb{N}$, $p \geq 1$ and $\lambda > 0$,

$$\mathbb{P}\left[\max_{1\leq k\leq n} X_k \geq \lambda\right] \leq \frac{1}{\lambda^p}\mathbb{E}[X_n^p]. \tag{D.1}$$

In particular, taking $X_k = |Y_1 + \ldots + Y_k|^2$ for a sequence of mean-zero, independent $\mathbb{R}$-valued random variables $(Y_k)_{k\in\mathbb{N}}$, (D.1) *implies Kolmogorov's inequality:*

$$\mathbb{P}\left[\max_{1\leq k\leq n} \big|Y_1 + \ldots + Y_k\big| \geq \lambda\right] \leq \frac{1}{\lambda^2}\mathrm{Var}[Y_1 + \ldots + Y_n] \equiv \frac{1}{\lambda^2}\sum_{k=1}^{n}\mathrm{Var}[Y_k]; \tag{D.2}$$

and also implies the Bienayme–Chebyshëv inequality for a single random variable Y:

$$\mathbb{P}[|Y| \geq \lambda] \leq \frac{1}{\lambda^p}\mathbb{E}[|Y|^p]. \tag{D.3}$$

Bibliography

[ABBM90] Bruno Alessandro, Cinzia Beatrice, Giorgio Bertotti, and Arianna Montorsi. Domain-wall dynamics and Barkhausen effect in metallic ferromagnetic materials. I. Theory. *J. Appl. Phys.*, 68(6):2901–2907, 1990.

[ACJ96] Rohan Abeyaratne, Chun-Hwa Chu, and Richard D. James. Kinetics of materials with wiggly energies: theory and application to the evolution of twinning microstructures in a Cu-Al-Ni shape memory alloy. *Phil. Mag. A*, 73:457–497, 1996.

[AGS08] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient Flows in Metric Spaces and in the Space of Probability Measures.* Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, second edition, 2008.

[And10] Edward N. da C. Andrade. On the Viscous Flow in Metals, and Allied Phenomena. *Proc. Roy. Soc. London*, A(84):1–12, 1910.

[And14] Edward N. da C. Andrade. The Flow in Metals under Large Constant Stresses. *Proc. Roy. Soc. London*, A(90):329–342, 1914.

[AS66] S.M. Ali and Samuel D. Silvey. A general class of coefficients of divergence of one distribution from another. *J. Roy. Statist. Soc. Ser. B*, 28:131–142, 1966.

[BE76a] Haïm Brézis and Ivar Ekeland. Un principe variationnel associé à certaines équations paraboliques. Le cas dépendant du temps. *C. R. Acad. Sci. Paris Sér. A–B*, 282(20):Ai, A1197–A1198, 1976.

[BE76b] Haïm Brézis and Ivar Ekeland. Un principe variationnel associé à certaines équations paraboliques. Le cas indépendant du

temps. *C. R. Acad. Sci. Paris Sér. A–B*, 282(17):Aii, A971–A974, 1976.

[Ber98] Giorgio Bertotti. *Hysteresis in Magnetism for Physicists, Materials Scientists and Engineers*. Electromagnetism. Academic Press, San Diego, 1998.

[Bil99] Patrick Billingsley. *Convergence of Probability Measures*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons Inc., New York, 1999.

[BK52] Evgenij A. Barbašin and Nikolaĭ N. Krasovskiĭ. On stability of motion in the large. *Doklady Akad. Nauk. SSSR (N.S.)*, 86:453–456, 1952.

[BLP78] Alain Bensoussan, Jacques-Louis Lions, and George Papanicolaou. *Asymptotic Analysis for Periodic Structures*, volume 5 of *Studies in Mathematics and its Applications*. North-Holland Publishing Co., Amsterdam, 1978.

[Bog66] Nikolaĭ N. Bogolyubov. On model dynamical systems in statistical mechanics. *Physica*, 32:933–944, 1966.

[Bog74] Nikolaĭ N. Bogolyubov. *Metod issledovaniya model′nykh gamil′tonianov [A method for studying model Hamiltonians]*. Izdat. "Nauka", Moscow, 1974.

[BV04] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, Cambridge, 2004.

[CFL28] Richard Courant, Kurt Friedrichs, and Hans Lewy. Über die partiellen Differenzengleichungen der mathematischen Physik. *Math. Ann.*, 100(1):32–74, 1928.

[Che96] Shiing-Shen Chern. Finsler geometry is just Riemannian geometry without the quadratic restriction. *Notices Amer. Math. Soc.*, 43(9):959–963, 1996.

[CM44] Robert H. Cameron and W. T. Martin. Transformations of Wiener integrals under translations. *Ann. of Math. (2)*, 45:386–396, 1944.

[CO08] Sergio Conti and Michael Ortiz. Minimum principles for the trajectories of systems governed by rate problems. *J. Mech. Phys. Solids*, 56(5):1885–1904, 2008.

[Cot96a] Alan H. Cottrell. Andrade creep. *Phil. Mag. Letters*, 73(1):35–37, 1996.

[Cot96b] Alan H. Cottrell. Criticality in Andrade creep. *Phil. Mag. A*, 74(4):1041–1046, 1996.

[Cot96c] Alan H. Cottrell. Strain hardening in Andrade creep. *Phil. Mag. Letters*, 74(5):375–379, 1996.

[Csi67] Imre Csiszár. Information-type measures of difference of probability distributions and indirect observations. *Studia Sci. Math. Hungar.*, 2:299–318, 1967.

[CV90] Pierluigi Colli and Augusto Visintin. On a class of doubly nonlinear evolution equations. *Comm. Partial Differential Equations*, 15(5):737–756, 1990.

[dC92] Manfredo P. do Carmo. *Riemannian Geometry*. Mathematics: Theory & Applications. Birkhäuser Boston Inc., Boston, MA, 1992. Translated from the second Portuguese edition by Francis Flaherty.

[DDM06] Gianni Dal Maso, Antonio DeSimone, and Maria Giovanna Mora. Quasistatic evolution problems for linearly elastic-perfectly plastic materials. *Arch. Ration. Mech. Anal.*, 180(2):237–291, 2006.

[DKMO00] Antonio DeSimone, Robert V. Kohn, Stefan Müller, and Felix Otto. Magnetic microstructures – a paradigm of multiscale problems. In *ICIAM 99 (Edinburgh)*, pages 175–190. Oxford Univ. Press, Oxford, 2000.

[Don51] Monroe D. Donsker. An invariance principle for certain probability limit theorems. *Mem. Amer. Math. Soc.*, 1951(6):12, 1951.

[DT02] Gianni Dal Maso and Rodica Toader. A model for the quasistatic growth of brittle fractures: existence and approximation results. *Arch. Ration. Mech. Anal.*, 162(2):101–135, 2002.

[DV08] Daryl J. Daley and David Vere-Jones. *An Introduction to the Theory of Point Processes. Vol. II.* Probability and its Applications (New York). Springer, New York, second edition, 2008. General theory and structure.

[DZ98] Amir Dembo and Ofer Zeitouni. *Large Deviations Techniques and Applications*, volume 38 of *Applications of Mathematics (New York)*. Springer-Verlag, New York, second edition, 1998.

[ET99] Ivar Ekeland and Roger Témam. *Convex Analysis and Variational Problems*, volume 28 of *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, english edition, 1999. Translated from the French.

[Fid06] Alexander Ya. Fidlin. *Nonlinear Oscillations in Mechanical Engineering*. Birkhäuser, 2006.

[FM98] Gilles A. Francfort and Jean-Jacques Marigo. Revisiting brittle fracture as an energy minimization problem. *J. Mech. Phys. Solids*, 46(8):1319–1342, 1998.

[FW98] Mark I. Freĭdlin and Alexander D. Wentzell. *Random Perturbations of Dynamical Systems*, volume 260 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, New York, second edition, 1998. Translated from the 1979 Russian original by Joseph Szücs.

[GM05] Adriana Garroni and Stefan Müller. Γ-limit of a phase-field model of dislocations. *SIAM J. Math. Anal.*, 36(6):1943–1964 (electronic), 2005.

[GM06] Adriana Garroni and Stefan Müller. A variational model for dislocations in the line tension limit. *Arch. Ration. Mech. Anal.*, 181(3):535–578, 2006.

[Gor94] Russell A. Gordon. *The Integrals of Lebesgue, Denjoy, Perron, and Henstock*, volume 4 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 1994.

[Gre54] Melville S. Green. Markoff Random Processes and the Statistical Mechanics of Time-Dependent Phenomena. II. Irreversible Processes in Fluids. *J. Chem. Phys.*, 22(3):398–413, 1954.

[Gri20] Alan A. Griffith. The phenomena of rupture and flow in solids. *Phil. Trans. Roy. Soc. London Ser. A*, 221:163–198, 1920.

[Grö19] Thomas H. Grönwall. Note on the derivatives with respect to a parameter of the solutions of a system of differential equations. *Ann. of Math. (2)*, 20(4):292–296, 1919.

[Gru04] Natalie Grunewald. *Mathematische Analyse einer Stick-Slip-Bewegung in zufälligem Medium*. PhD thesis, Mathematisches Institut, Rheinische Friedrich-Wilhelms-Universität, Bonn, Germany, 2004.

[Gru05] Natalie Grunewald. Barkhausen effect: a stick-slip motion in a random medium. *Methods Appl. Anal.*, 12(1):29–41, 2005.

[GS01] Geoffrey R. Grimmett and David R. Stirzaker. *Probability and Random Processes*. Oxford University Press, New York, third edition, 2001.

[HMS02] Desmond J. Higham, Xuerong Mao, and Andrew M. Stuart. Strong convergence of Euler-type methods for nonlinear stochastic differential equations. *SIAM J. Numer. Anal.*, 40(3):1041–1063 (electronic), 2002.

[How98] Ralph E. Howard. The Gronwall Inequality, November 1998. http://www.math.sc.edu/∼howard/Notes/gronwall.pdf.

[HSY92] Brian R. Hunt, Tim Sauer, and James A. Yorke. Prevalence: a translation-invariant "almost every" on infinite-dimensional spaces. *Bull. Amer. Math. Soc. (N.S.)*, 27(2):217–238, 1992.

[Jay57a] Edwin T. Jaynes. Information theory and statistical mechanics. *Phys. Rev. (2)*, 106:620–630, 1957.

[Jay57b] Edwin T. Jaynes. Information theory and statistical mechanics. II. *Phys. Rev. (2)*, 108:171–190, 1957.

[JKO98] Richard Jordan, David Kinderlehrer, and Felix Otto. The variational formulation of the Fokker-Planck equation. *SIAM J. Math. Anal.*, 29(1):1–17 (electronic), 1998.

[Kan42] Leonid V. Kantorovich. On the translocation of masses. *C. R. (Doklady) Acad. Sci. USSR (N.S.)*, 37:199–201, 1942.

[KCO02] Marisol Koslowski, Alberto M. Cuitiño, and Michael Ortiz. A phase-field theory of dislocation dynamics, strain hardening and hysteresis in ductile single crystals. *J. Mech. Phys. Solids*, 50(12):2597–2635, 2002.

[Ken05] Wilfrid Kendall. Notes on perfect simulation. In *Markov chain Monte Carlo*, volume 7 of *Lect. Notes Ser. Inst. Math. Sci. Natl. Univ. Singap.*, pages 93–146. World Sci. Publ., Hackensack, NJ, 2005.

[Khi24] Aleksandr Ya. Khinchin. Über einen Satz der Wahrscheinlichkeitsrechnung. *Fund. Math.*, 6:9–20, 1924.

[Kin93] John F.C. Kingman. *Poisson Processes*, volume 3 of *Oxford Studies in Probability*. The Clarendon Press Oxford University Press, New York, 1993. Oxford Science Publications.

[KK73] Stanley Kogelman and Joseph B. Keller. Asymptotic theory of nonlinear wave propagation. *SIAM J. Appl. Math.*, 24:352–361, 1973.

[KM97] Markus Kunze and Manuel D.P. Monteiro Marques. Existence of solutions for degenerate sweeping processes. *J. Convex Anal.*, 4(1):165–176, 1997.

[KM98] Markus Kunze and Manuel D.P. Monteiro Marques. On the discretization of degenerate sweeping processes. *Portugal. Math.*, 55(2):219–232, 1998.

[Kol29] Andreĭ N. Kolmogorov. Über das Gesetz des iterierten Logarithmus. *Math. Ann.*, 101(1):126–135, 1929.

[Kos03] Marisol Koslowski. *A phase-field model of dislocations in ductile single crystals*. PhD thesis, California Institute of Technology, Pasadena, California, USA, 2003.

[KP92] Peter E. Kloeden and Eckhard Platen. *Numerical Solution of Stochastic Differential Equations*, volume 23 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, 1992.

[KS91] Ioannis Karatzas and Steven E. Shreve. *Brownian Motion and Stochastic Calculus*, volume 113 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1991.

[Kub57] Ryogo Kubo. Statistical-mechanical theory of irreversible processes I. general theory and simple applications to magnetic and conduction problems. *J. Phys. Soc. Jpn.*, 12:570, 1957.

[Kur66] Kazimierz Kuratowski. *Topology. Vol. I.* Academic Press, New York, 1966. New edition, revised and augmented. Translated from the French by J. Jaworowski.

[Lam77] John Lamperti. *Stochastic Processes.* Springer-Verlag, New York, 1977. A survey of the mathematical theory, Applied Mathematical Sciences, Vol. 23.

[Lar75] Edward W. Larsen. Neutron transport and diffusion in inhomogeneous media. I. *J. Mathematical Phys.*, 16:1421–1427, 1975.

[LaS76] Joseph P. LaSalle. Stability theory and invariance principles. In *Dynamical systems (Proc. Internat. Sympos., Brown Univ., Providence, R.I., 1974), Vol. I*, pages 211–222. Academic Press, New York, 1976.

[Lee97] John M. Lee. *Riemannian Manifolds: An Introduction to Curvature*, volume 176 of *Graduate Texts in Mathematics.* Springer-Verlag, New York, 1997.

[LT91] Michel Ledoux and Michel Talagrand. *Probability in Banach Spaces*, volume 23 of *Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)].* Springer-Verlag, Berlin, 1991. Isoperimetry and processes.

[Mar50] Gisiro Maruyama. Notes on Wiener integrals. *Kōdai Math. Sem. Rep.*, 2:41–44, 1950.

[Mat60] Bertil Matérn. *Spatial variation: Stochastic models and their application to some problems in forest surveys and other sampling investigations.* Meddelanden Fran Statens Skogsforskningsinstitut, Band 49, Nr. 5, Stockholm, 1960.

[Men01] Govind Menon. *Geometric methods for the Maxwell–Bloch equations and the kinetics of martensitic phase transitions.* PhD thesis, Brown University, Providence, Rhode Island, USA, 2001.

[Men02] Govind Menon. Gradient systems with wiggly energies and related averaging problems. *Arch. Ration. Mech. Anal.*, 162(3):193–246, 2002.

[Mie03] Alexander Mielke. Energetic formulation of multiplicative elasto-plasticity using dissipation distances. *Contin. Mech. Thermodyn.*, 15(4):351–382, 2003.

[Mie05] Alexander Mielke. Evolution of rate-independent systems. In *Evolutionary equations. Vol. II*, Handb. Differ. Equ., pages 461–559. Elsevier/North-Holland, Amsterdam, 2005.

[Mie07] Alexander Mielke. Modeling and analysis of rate-independent processes, January 2007. Lipschitz Lecture held in Bonn: http://www.wias-berlin.de/people/mielke/papers/Lipschitz07Mielke.pdf.

[MO08] Alexander Mielke and Michael Ortiz. A class of minimum principles for characterizing the trajectories and the relaxation of dissipative systems. *ESAIM: Control Optim. Calc. Var.*, 14:494–516, 2008.

[Mon81] Gaspard Monge. Mémoire sur la théorie des déblais et de remblais. In *Histoire de l'Académie Royale des Sciences de Paris, avec les Mémoires de Mathématique et de Physique pour la même année*, pages 666–704. 1781.

[Mon93] Manuel D.P. Monteiro Marques. *Differential Inclusions in Nonsmooth Mechanical Problems*. Progress in Nonlinear Differential Equations and their Applications, 9. Birkhäuser Verlag, Basel, 1993. Shocks and dry friction.

[Mor65] Jean-Jacques Moreau. Proximité et dualité dans un espace hilbertien. *Bull. Soc. Math. France*, 93:273–299, 1965.

[Mor70] Jean-Jacques Moreau. Sur les lois de frottement, de plasticité et de viscosité. *C. R. Acad. Sci. Paris Sér. II Méc. Phys. Chim. Sci. Univers Sci. Terre*, 271:608–611, 1970.

[Mor71] Jean-Jacques Moreau. Sur l'évolution d'un système élasto-visco-plastique. *C. R. Acad. Sci. Paris Sér. A-B*, 273:A118–A121, 1971.

[Mor76] Jean-Jacques Moreau. Application of convex analysis to the treatment of elastoplastic systems. In Paul Germain and Bernard Nayroles, editors, *Applications of Methods of Functional Analysis to Problems in Mechanics*, volume 503 of *Lecture Notes in Mathematics*, pages xix+531, Berlin, 1976. Springer-Verlag. Joint Symposium, IUTAM/IMU, held in Marseille, September 1–6, 1975.

[MR09] Alexander Mielke and Filip Rindler. Reverse approximation of energetic solutions to rate-independent processes. *NoDEA Nonlinear Differential Equations Appl.*, 16(1):17–40, 2009.

[MT04] Alexander Mielke and Florian Theil. On rate-independent hysteresis models. *NoDEA Nonlinear Differential Equations Appl.*, 11(2):151–189, 2004.

[MTL02] Alexander Mielke, Florian Theil, and Valery I. Levitas. A variational formulation of rate-independent phase transformations using an extremum principle. *Arch. Ration. Mech. Anal.*, 162(2):137–177, 2002.

[Mul87] Patrick Muldowney. *A General Theory of Integration in Function Spaces, Including Wiener and Feynman Integration*, volume 153 of *Pitman Research Notes in Mathematics Series*. Longman Scientific & Technical, Harlow, 1987.

[Nay76a] Bernard Nayroles. Deux théorèmes de minimum pour certains systèmes dissipatifs. *C. R. Acad. Sci. Paris Sér. A-B*, 282(17):Aiv, A1035–A1038, 1976.

[Nay76b] Bernard Nayroles. Un théorème de minimum pour certains systèmes dissipatifs. Variante hilbertienne. *Travaux Sém. Anal. Convexe*, 6(Exp. 2):22, 1976.

[O'N95] Toby O'Neil. A measure with a large set of tangent measures. *Proc. Amer. Math. Soc.*, 123(7):2217–2220, 1995.

[OR99] Michael Ortiz and Eduardo A. Repetto. Nonconvex energy minimization and dislocation structures in ductile single crystals. *J. Mech. Phys. Solids*, 47(2):397–462, 1999.

[Ord66] Edward T. Ordman. Classroom Notes: Convergence Almost Everywhere is Not Topological. *Amer. Math. Monthly*, 73(2):182–183, 1966.

[Ort81] Michael Ortiz. *Topics in constitutive theory for inelastic solids*. PhD thesis, University of California at Berkeley, Berkeley, California, USA, 1981.

[Ött96] Hans C. Öttinger. *Stochastic Processes in Polymeric Fluids*. Springer-Verlag, Berlin, 1996. Tools and examples for developing simulation algorithms.

[Par05] Kalyanapuram R. Parthasarathy. *Probability Measures on Metric Spaces*. AMS Chelsea Publishing, Providence, RI, 2005. Reprint of the 1967 original.

[Pol84] David Pollard. *Convergence of Stochastic Processes*. Springer Series in Statistics. Springer-Verlag, New York, 1984.

[Pre87] David Preiss. Geometry of measures in $\mathbb{R}^n$: distribution, rectifiability, and densities. *Ann. of Math. (2)*, 125(3):537–643, 1987.

[Pro54] Yuriĭ V. Prokhorov. Methods of functional analysis in limit theorems of probability theory. *Vestnik Leningrad Univ.*, 11, 1954.

[Pro56] Yuriĭ V. Prokhorov. Convergence of random processes and limit theorems in probability theory. *Teor. Veroyatnost. i Primenen.*, 1:177–238, 1956.

[PS08] Grigorios A. Pavliotis and Andrew M. Stuart. *Multiscale Methods: Averaging and Homogenization*, volume 53 of *Texts in Applied Mathematics*. Springer, New York, 2008.

[PT82] Jean-Paul Penot and Michel Théra. Semicontinuous mappings in general topology. *Arch. Math. (Basel)*, 38(2):158–166, 1982.

[Roc70] R. Tyrrell Rockafellar. *Convex Analysis*. Princeton Mathematical Series, No. 28. Princeton University Press, Princeton, N.J., 1970.

[RSPR09] Daniel Rittel, Mike L. Silva, P.C. Benny Poon, and Guruswami Ravichandran. Thermomechanical behaviour of single crystalline tantalum in the static and dynamic regime. 2009. In preparation.

[Sch66] Michael Schilder. Some asymptotic formulas for Wiener integrals. *Trans. Amer. Math. Soc.*, 125:63–85, 1966.

[Sho97] Ralph E. Showalter. *Monotone Operators in Banach Space and Nonlinear Partial Differential Equations*, volume 49 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 1997.

[SKM87] Dietrich Stoyan, Wilfrid S. Kendall, and Joseph Mecke. *Stochastic Geometry and its Applications*. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. John Wiley & Sons Ltd., Chichester, 1987. With a foreword by David G. Kendall.

[SKTO09] Timothy J. Sullivan, Marisol Koslowski, Florian Theil, and Michael Ortiz. On the behaviour of dissipative systems in contact with a heat bath: Application to Andrade creep. *J. Mech. Phys. Solids*, 2009. To appear.

[Slu25] Evgeniĭ E. Slutskiĭ. Über stochastische Asymptoten und Grenzwerte. *Metron*, 5(3):3–89, 1925.

[Smi02] Georgi V. Smirnov. *Introduction to the Theory of Differential Inclusions*, volume 41 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2002.

[ST07] Timothy J. Sullivan and Florian Theil. Deterministic stick-slip dynamics in a one-dimensional random potential. In Martin Kružík, editor, *Analysis and Numerics for Rate-Independent Processes, 25 February–3 March 2007*, volume 11/2007 of *Oberwolfach Reports*, pages 62–65, Zürich, 2007. Mathematisches Forschungsinstitut Oberwolfach, European Mathematical Society Publishing House.

[Ste08] Ulisse Stefanelli. The Brezis–Ekeland principle for doubly nonlinear equations. *SIAM J. Control Optim.*, 47(3):1615–1642, 2008.

[Ste09] Ulisse Stefanelli. The discrete Brezis–Ekeland principle. *J. Convex Anal.*, 2009. To appear.

[Vil09] Cédric Villani. *Optimal Transport, Old and New*, volume 338 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2009.

[Was69] Leonid N. Wasserstein. Markov processes over denumerable products of spaces describing large system of automata. *Problemy Peredači Informacii*, 5(3):64–72, 1969.

[Wie86] Frederik W. Wiegel. *Introduction to Path-Integral Methods in Physics and Polymer Science*. World Scientific Publishing Co., Singapore, 1986.

[Yam85] Yasuo Yamasaki. *Measures on Infinite-Dimensional Spaces*, volume 5 of *Series in Pure Mathematics*. World Scientific Publishing Co., Singapore, 1985.

[YM08] Chenggui Yuan and Xuerong Mao. A note on the rate of convergence of the Euler-Maruyama method for stochastic differential equations. *Stoch. Anal. Appl.*, 26(2):325–333, 2008.

[Yos65] Kōsaku Yosida. *Functional Analysis*. Die Grundlehren der Mathematischen Wissenschaften, Band 123. Academic Press Inc., New York, 1965.

[Zei95] Eberhard Zeidler. *Applied Functional Analysis*, volume 109 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 1995. Main principles and their applications.

[Zha05] Fuzhen Zhang, editor. *The Schur Complement and its Applications*, volume 4 of *Numerical Methods and Algorithms*. Springer-Verlag, New York, 2005.

[ZKO94] Vasiliĭ V. Zhikov, Sergeĭ M. Kozlov, and Olga A. Oleĭnik. *Homogenization of Differential Operators and Integral Functionals*. Springer-Verlag, Berlin, 1994. Translated from the Russian by G.A. Iosif′yan.