European Journal of Applied Mathematics (2026), **37**, pp. 123–179
doi:10.1017/S0956792525100120

CAMBRIDGE UNIVERSITY PRESS

PAPER

Adversarial flows: A gradient flow characterization of adversarial attacks

Lukas Weigand[1], Tim Roith[1] and Martin Burger[1,2]

[1]Helmholtz Imaging, Deutsches Elektronen-Synchrotron DESY, Notkestr. 85, 22607 Hamburg, Germany
[2]Department of Mathematics, Bundesstr. 55, University of Hamburg, 20146 Hamburg, Germany
Corresponding author: Tim Roith; Email: tim.roith@desy.de

Received: 08 June 2024; **Revised:** 04 August 2025; **Accepted:** 04 August 2025; **First published online:** 12 September 2025

Keywords: adversarial attacks; adversarial training; metric gradient flows; Wasserstein gradient flows

2020 Mathematics Subject Classification: 49Q20, 34A60, 68Q32, 65K15

Abstract

A popular method to perform adversarial attacks on neural networks is the so-called fast gradient sign method and its iterative variant. In this paper, we interpret this method as an explicit Euler discretization of a differential inclusion, where we also show convergence of the discretization to the associated gradient flow. To do so, we consider the concept of p-curves of maximal slope in the case $p = \infty$. We prove existence of ∞-curves of maximum slope and derive an alternative characterization via differential inclusions. Furthermore, we also consider Wasserstein gradient flows for potential energies, where we show that curves in the Wasserstein space can be characterized by a representing measure on the space of curves in the underlying Banach space, which fulfil the differential inclusion. The application of our theory to the finite-dimensional setting is twofold: On the one hand, we show that a whole class of normalized gradient descent methods (in particular, signed gradient descent) converge, up to subsequences, to the flow when sending the step size to zero. On the other hand, in the distributional setting, we show that the inner optimization task of adversarial training objective can be characterized via ∞-curves of maximum slope on an appropriate optimal transport space.

1. Introduction

This paper considers gradient flows in metric spaces, following the seminal work by [2]. There, the authors introduce the concept of p-curves of maximal slope, with origins dating back to [31]. This concept is further generalized in [87]. As for our main contribution, we study the less-known limit case $p = \infty$ and adapt current theory to this setting. The main incentive for our work is the adversarial attack problem as introduced in [46, 101]. Here one considers a classification task, where a classifier $h : \mathcal{X} \to \mathcal{Y}$ – typically parametrized as a neural network – is given an input $x \in \mathcal{X}$, which it *correctly* classifies as $y \in \mathcal{Y}$, where $\mathcal{Y}$ is assumed to be a subset of a finite dimensional vector space. The goal is to obtain a perturbed input $\tilde{x} \in \mathcal{X}$, the *adversarial example*, which is misclassified, while its difference to x is "imperceptible". In practice, the latter condition is enforced by requiring that $\tilde{x}$ has at most distance ε to x in an ℓ^p distance, where $\varepsilon > 0$ is called the *adversarial budget*. Given some loss function $\ell : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}$, one then formulates the *adversarial attack* problem [46, 101],

$$\sup_{\tilde{x} \in \overline{B}_\varepsilon(x)} \ell(h(\tilde{x}), y). \tag{AdvAtt}$$

The above problem is also called an *untargeted* attack, since we are solely interested in the misclassification. This is opposed to *targeted* attacks, where one prescribes $y_{\text{target}} \in \mathcal{Y}$ and wants to obtain an adversarial example, s.t. $h(\tilde{x}) = y_{\text{target}}$. This basically amounts to changing the loss function in (AdvAtt),

© The Author(s), 2025. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/S0956792525100120 Published online by Cambridge University Press

namely to $-\ell(\cdot, y_{\text{target}})$, without changing the inherent structure of the problem, which is why we do not consider it separately in the following. Methods for generating adversarial examples include first-order attacks [12, 71, 80], momentum-variants [35], second-order attacks [55] or even zero-order attacks, not employing the gradient of the classifier [11, 53]. Especially for classifiers induced by neural networks, it was noticed in [101] that approximate maximizers of (AdvAtt) completely corrupt the classification performance, even for a very small budget ε. This observation created severe concerns about the robustness and reliability of neural networks (see e.g. [59]) and has sparked a general interest in both the adversarial attack and the defence problem. The connection between the attack and defence task was already introduced in [46], where the authors propose *adversarial training* (similarly derived in [58, 64]). Here, the standard empirical risk minimization is modified to

$$\inf_{h\in\mathcal{H}} \sum_{(x,y)\in\mathcal{T}} \sup_{\tilde{x}\in\overline{B_\varepsilon}(x)} \ell(h(\tilde{x}), y) \tag{AdvTrain}$$

for a training set $\mathcal{T} \subset \mathcal{X} \times \mathcal{Y}$ and a hypothesis class $\mathcal{H} \subset \{h | h : \mathcal{X} \to \mathcal{Y}\}$. Since this requires solving (AdvAtt) for every data point x, the authors then propose an efficient one-step method, called fast gradient sign method (FGSM),

$$x_{\text{FGS}} = x + \varepsilon \ \operatorname{sign}(\nabla_x \ell(h(x), y)). \tag{FGSM}$$

The motivation, as provided in [46], was to consider a linear model $x \mapsto \langle w, x\rangle$, with weights w. The maximum over the input x constrained to the budget ball $\overline{B_\varepsilon^\infty}(x)$ is then attained in a corner of the hypercube, which validates the use of the sign. From a practical perspective, also for more complicated models, the sign operation ensures that $x_{\text{FGS}} \in \partial B_\varepsilon^\infty(x)$, i.e., x_{FGS} uses all the given budget in the ℓ^∞ distance after just one update step. This adversarial training setup was similarly employed in [64, 88, 94, 105] and analyzed as regularization of the empirical risk in [18, 20]. For other strategies to obtain robust classifiers, we refer, e.g., to [21, 47, 57, 77]. In situations where only the attack problem is of interest, multistep methods are feasible, which led to the iterative FGS method [58, 59]

$$x_{\text{IFGS}}^{k+1} = \Pi_{\overline{B_\varepsilon^p}(x)}(x_{\text{IFGS}}^k + \tau \ \operatorname{sign}(\nabla_x \ell(h(x_{\text{IFGS}}^k), y))), \tag{IFGSM}$$

where $\tau > 0$ now defines a step size and $\Pi_{\overline{B_\varepsilon^p}(x)}$ denotes the orthogonal projection to the ε-ball in the ℓ^p-norm around the original image. Originally, the case $p = \infty$ was employed, where the projection is then a simple clipping operation. Other choices of p are usually limited to $\{0, 1, 2\}$, which is also due to the computational effort of computing the projection (see [80] for $p = 0$ and [36] for $p = 1$). Signed gradient descent can also be interpreted as a form of normalized gradient descent in the ℓ^∞ topology as in [27], where our framework allows considering a general ℓ^q norm. Apart from the adversarial setting, signed gradient descent, without the projection step, is an established optimization algorithm itself, see e.g., [70, 106] for other applications. The idea of using signed gradients can also be found in the RPROP algorithm [83]. The convergence to minimizers of signed gradient descent and its variants was analyzed in [5, 26, 61, 74]. A slightly different kind of projected version, using linear constraints, was considered in [25], where the authors also considered a continuous time version; however, the results therein and the considered flow are not directly connected to our work here. We consider the limit $\tau \to 0$ of signed gradient descent and the projected variant (IFGSM), for which we derive a gradient flow characterization. This is visualized in Figure 1. In the Euclidean setting with a differentiable energy $\mathcal{E} : \mathbb{R}^d \to \mathbb{R}$ and $p \in (1, \infty)$, a differentiable curve $u : [0, T] \to \mathbb{R}^d$ is a p-curve of maximum slope, if it solves the p-gradient flow equation

$$\left(\left|u'\right|(t)\right)^{p-2} u'(t) = -\nabla\mathcal{E}(u(t)).$$

Here, we also refer to [14, 15] for a study of gradient-flow type equations in Hilbert spaces, for non-differentiable functionals. Following the approach in [2, 31, 32, 65], the above equation is equivalent to

$$\frac{\mathrm{d}}{\mathrm{d}t}(\mathcal{E} \circ u) \le -\frac{1}{p}\left|u'\right|^p - \frac{1}{q}\left|\nabla\mathcal{E}(u)\right|^q,$$

https://doi.org/10.1017/S0956792525100120 Published online by Cambridge University Press

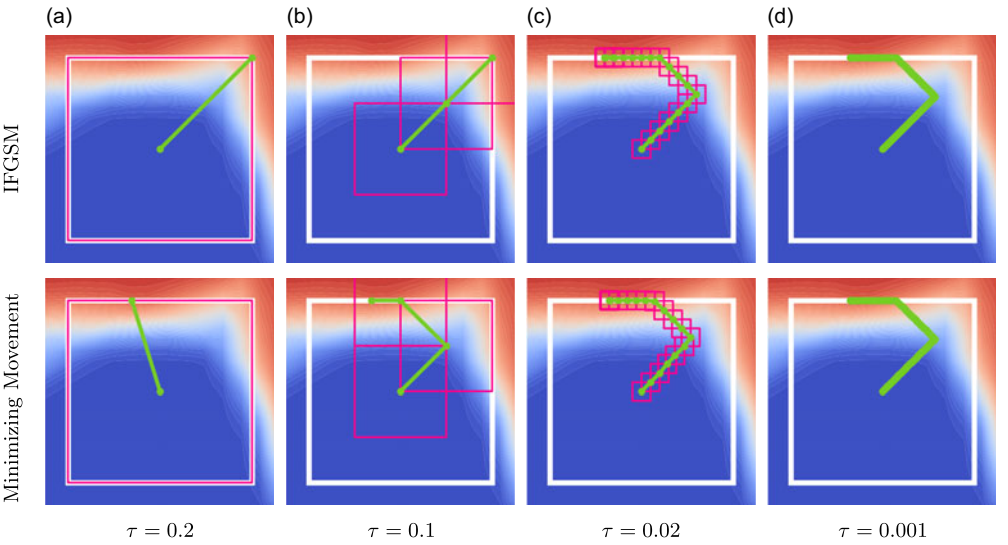


Figure 1. *Behavior of (IFGSM) (top) and the minimizing movement scheme (MinMove) (bottom), for a binary classifier – parametrized as a neural network – on $\mathbb{R}^2$, a budget of $\varepsilon = 0.2$ and $\tau \in \{0.2, 0.1, 0.02, 0.001\}$. The white box indicates the maximal distance to the initial value, and the pink boxes indicate the step size τ of the scheme. Details on this experiment can be found in Appendix H.*

where $1/p + 1/q = 1$. The strength of this approach is that all derivatives in the above inequality have meaningful generalizations to the metric space setting, which we repeat in the next section. Motivated by signed gradient descent, in this paper, we draw the connection to the case $p = \infty$. In the Euclidean setting, with a differentiable functional $\mathcal{E}$, the energy dissipation inequality we derive for $p = \infty$ reads

$$\begin{aligned} &|u'| \leq 1, \\ &\frac{\mathrm{d}}{\mathrm{d}t}(\mathcal{E} \circ u) \leq -|\nabla \mathcal{E}(u)|. \end{aligned}$$

Intuitively, a ∞-curve of maximal slope minimizes the energy $\mathcal{E}$ as fast as possible under the restriction that its velocity $|u'|$ is bounded by 1. Like in [2], our results consider general metric spaces, Banach spaces and Wasserstein spaces, which are further detailed in the following sections. Typically, curves of maximum slope can be approximated via a minimizing movement scheme, which in our case translates to

$$x_\tau^{k+1} \in \operatorname*{arg\,min}_{x \in \mathcal{X}} \{\mathcal{E}(x) : \|x - x_\tau^k\| \leq \tau\},$$

where $x_\tau^0 = x^0$ is a given initial value. A main insight, explored in section 5, is that under certain assumptions, (FGSM) and (IFGSM) fulfil this scheme, if we replace the energy by a semi-implicit version.

A further aspect is the characterization of adversarial attacks in the distributional setting, where the sum is replaced by an integral over the data distribution μ. Interchanging the integral and the supremum (see Corollary 5.7) yields the characterization of adversarial training (AdvTrain) as a distributionally robust optimization (DRO) problem,

$$\inf_{h \in \mathcal{H}} \sup_{\tilde{\mu}: D(\mu, \tilde{\mu}) \leq \varepsilon} \int \ell(h(x), y)\, \mathrm{d}\tilde{\mu}(x, y), \tag{DRO}$$

where D denotes a distance on the space of distributions. This formulation of adversarial training was the subject of many studies in recent years, see, e.g., [18, 20, 22, 23, 107]. Typically, the distance D is

https://doi.org/10.1017/S0956792525100120 Published online by Cambridge University Press

chosen as an optimal transport distance,

$$D(\mu, \tilde{\mu}) := \inf_{\gamma \in \Gamma(\mu, \tilde{\mu})} \int_{\gamma} c((x, y), (\tilde{x}, \tilde{y}))^2 d\gamma,$$

with $\Gamma(\mu, \tilde{\mu})$ denoting the set of all couplings and the cost

$$c((x, y), (\tilde{x}, \tilde{y})) := \begin{cases} \|x - \tilde{x}\| & \text{if } y = \tilde{y}, \\ +\infty & \text{if } y \neq \tilde{y}. \end{cases} \tag{1.1}$$

The goal here is then to derive a characterization of curves $\mu : [0, T] \to \mathcal{W}_p$, where $\mathcal{W}_p$ denotes the p-Wasserstein space. In this regard, we mention the related work [107], where the authors proposed to solve the inner optimization problem

$$\sup_{\tilde{\mu}: D(\mu, \tilde{\mu}) \leq \varepsilon} \int \ell(h(x), y) \, \mathrm{d}\tilde{\mu}(x, y)$$

by disintegrating the data distribution $\mathrm{d}\mu(x, y) = \mathrm{d}\mu_y(x)\mathrm{d}\nu(y)$ (see Appendix E), and calculating for ν-a.e. $y \in \mathcal{Y}$ the corresponding 2-gradient flow in $\mathcal{W}_2$ with initial condition μ_y^0. As shown in [2], solving this gradient flow is equivalent to solving the partial differential equation

$$\begin{aligned} \partial_t (\mu_y)_t &= \nabla \cdot ((\mu_y)_t \nabla_x \ell(h(x), y)) \quad \text{on } (0, T) \\ (\mu_y)_0 &= \mu_y^0, \end{aligned} \tag{1.2}$$

which is to be understood in the distributional sense. The authors in [107] then approximate a maximizer by $\mathrm{d}\tilde{\mu}(x, y) \approx \mathrm{d}(\mu_y)_T(x)\mathrm{d}\nu(y)$, where T has to be chosen small enough such that the approximation is still within the ε ball around μ.

In the following, we first provide the necessary notions for gradient flows in metric spaces and then proceed to discuss the main contributions and the outline of this paper.

1.1. Setup

We give a brief recap on classical notation and preliminaries on evolution in metric spaces. More details can be found in [2, 68]. In the following, we denote by $(\mathcal{S}, d)$ a complete metric space, while $\mathcal{X}$ denotes a Banach space. We consider a proper functional $\mathcal{E} : \mathcal{S} \to (-\infty, +\infty]$, i.e., the effective domain $\mathrm{dom}\,(\mathcal{E}) := \{x \in \mathcal{S} : \mathcal{E}(x) < \infty\}$ is assumed to be nonempty. Throughout this paper, we denote by

$$B_\tau(x) := \{\tilde{x} \in \mathcal{S} : d(x, \tilde{x}) < \tau\}, \qquad \overline{B_\tau}(x) := \{\tilde{x} \in \mathcal{S} : d(x, \tilde{x}) \leq \tau\}$$

the ball and its closed variant, induced by the given metric d, where we employ the abbreviation $B_\tau(0) = B_\tau$. In the finite-dimensional case, we write B_τ^p to denote the ball induced by the ℓ^p norm on $\mathbb{R}^d$. Note that there is a notation conflict with d denoting both the distance and the dimension of the finite-dimensional space $\mathbb{R}^d$. However, the concrete meaning is always clear from the context.

Metric derivative. We consider curves $u : [0, T] \to \mathcal{S}$ with $T > 0$ for which we want to have a notion of velocity. For this purpose, we need a generalization of the absolute value of the derivatives, which is provided by the *metric derivative* as introduced by [1]. Here, one usually considers p-absolutely continuous curves [2], i.e., for $p \in [1, \infty]$, there exists $m \in L^p(0, T)$ such that

$$d(u(t), u(s)) \leq \int_s^t m(r) \, \mathrm{d}r \tag{1.3}$$

for all $0 \leq s < t \leq T$. The set of all p-absolutely continuous curves is denoted by $AC^p(0, T; \mathcal{S})$. We are especially interested in the case $p = \infty$, where the condition in Equation (1.3) is equivalent to the Lipschitzness of the curve, i.e., the existence of a constant $L \geq 0$ such that

$$d(u(t), u(s)) \leq L \, (t - s)$$

https://doi.org/10.1017/S0956792525100120 Published online by Cambridge University Press

for all $0 \le s < t \le T$. For the special case $p = \infty$, we have the following result as a special case of [2, Theorem 1.1.2].

Lemma 1.1 (Metric derivative). *Let $u : [0, T] \to \mathcal{S}$ be a Lipschitz curve with Lipschitz constant L, then the limit*

$$|u'|(t) := \lim_{s \to t} \frac{d(u(s), u(t))}{|s - t|}$$

exists for a.e. $t \in [0, T]$ and is referred to as the metric derivative. Moreover, the function $t \mapsto |u'|(t)$ belongs to $L^\infty(0, T)$ with $\||u'|\|_{L^\infty(0,T)} \le L$, and

$$d(u(s), u(t)) \le \int_s^t |u'|(r)\, dr \quad \text{for all } 0 \le s \le t \le T.$$

Remark 1.2. The metric derivative $|u'|$ is actually minimal in the sense that for every m satisfying (1.3),

$$|u'|(t) \le m(t) \quad \text{for a.e. } t \in (0, T).$$

Remark 1.3. If $\mathcal{S} = \mathcal{X}$ is a Banach space and satisfies the Radon–Nikodým property (c.f. [89, p. 106]), e.g., if it is reflexive, then $u \in AC^p(0, T; \mathcal{X})$ if and only if

- u is differentiable a.e. on $(0, T)$
- $u'(t) \in L^p(0, T; \mathcal{X})$
- $u(t) - u(s) = \int_s^t u'(t)\, \mathrm{d}r$ for $0 \le s \le t \le T$.

Upper gradients We consider *upper gradients* as a generalization of the absolute value of the gradient in the metric setting. Namely, we employ the following definitions from [2, Definition 1.2.1] and [2, Definition 1.2.2].

Definition 1.4. *A function $g : \mathcal{S} \to [0, +\infty]$ is called a strong upper gradient for $\mathcal{E}$ if, for every absolutely continuous curve $u : [0, T] \to \mathcal{S}$, the function $g \circ u$ is Borel and*

$$|\mathcal{E}(u(t)) - \mathcal{E}(u(s))| \le \int_s^t g(u(r))|u'|(r)\, dr \quad \forall\, 0 \le s \le t \le T \tag{1.4}$$

If $(g \circ u)\, |u'| \in L^1(0, T)$, then $\mathcal{E} \circ u$ is absolutely continuous and

$$|(\mathcal{E} \circ u)'(t)| \le g(u(t))|u'|(t) \quad \text{for a.e. } t \in (0, T). \tag{1.5}$$

Definition 1.5. *A function $g : \mathcal{S} \to [0, +\infty]$ is called a weak upper gradient for $\mathcal{E}$, if for every absolutely continuous curve $u : [0, T] \to \mathcal{S}$ that fulfils*

(i) $(g \circ u)\ |u'| \in L^1(0, T)$,

(ii) $\mathcal{E} \circ u$ *is a.e. in* $(0, T)$ *equal to a function* $\psi : (0, T) \to \mathbb{R}$ *with bounded variation,*

it follows that

$$\left|\psi'\right| \le (g \circ u)\ \left|u'\right| \ \text{a.e. in } (0, T).$$

Remark 1.6. We note that for a function ψ with bounded variation, i.e.,

$$\sup \left\{ \sum_{i=0}^{N-1} |\psi(t_{i+1}) - \psi(t_i)| \ : 0 = t_0 < \ldots < t_N = T \right\} < \infty,$$

we have that the derivative ψ' exists a.e. in the interval $(0, T)$, see [90, Theorem 9.6, Chapter IV].

Remark 1.7. Admissible curves u in the above definition fulfil that $u^{-1}(\mathcal{S} \setminus \mathrm{dom}\,(\mathcal{E}))$ is a null set, because of (ii). Therefore, the behaviour of g outside of $\mathrm{dom}\,(\mathcal{E})$ is negligible.

https://doi.org/10.1017/S0956792525100120 Published online by Cambridge University Press

Metric slope We now consider the *metric slope*, as defined in [31], as a special realization of a weak upper gradient. Intuitively, the slope gives the value of the maximal descent at a point u at an infinitesimal small distance.

Definition 1.8. *For a proper functional $\mathcal{E} : \mathcal{S} \to (-\infty, +\infty]$, the local slope of $\mathcal{E}$ at $x \in \operatorname{dom}(\mathcal{E})$ is defined as*

$$|\partial \mathcal{E}|(x) := \limsup_{z \to x} \frac{(\mathcal{E}(x) - \mathcal{E}(z))^+}{d(x, z)}.$$

The definition of the slope does, in fact, yield an upper gradient, which is provided by the following statement from [2].

Theorem 1.9 [2, Theorem 1.2.5]. *Let $\mathcal{E}$ be a proper functional, then the function $|\partial \mathcal{E}|$ is a weak upper gradient.*

Curves of maximal slope Curves of maximal slope were introduced in [31] and are a possible generalization of a gradient evolution in metric spaces. They are usually formulated for the case $p \in (1, \infty)$ as follows, see, e.g., [2].

Definition 1.10 (p-Curves of maximal slope). *For $p \in (1, \infty)$, we say that an absolutely continuous curve $u : [0, T] \to \mathcal{S}$ is a p-curve of maximal slope, for the functional $\mathcal{E}$ with respect to an upper gradient g, if $\mathcal{E} \circ u$ is a.e. equal to a non-increasing map ψ and*

$$\psi'(t) \leq -\frac{1}{p} \left|u'\right|^p (t) - \frac{1}{q} g^q(u(t)) \tag{1.6}$$

for almost every $t \in (0, T)$ and $1 = \frac{1}{p} + \frac{1}{q}$.

For $p \in (1, \infty)$, the existence of such curves is provided, see for example [2].

1.2. Main results

Here, we summarize the main contributions of this paper. The most important one is the development and application of a gradient flow framework that allows for a theoretical study of adversarial attacks. Concerning the theory of metric gradient flows, we introduce notions tailored to this application and also provide adapted proofs, as detailed below. Here, it should be noted however that many of our results in metric and Banach spaces can be obtained from the theory of doubly nonlinear equations [69, 87]. Therefore, the main contribution from this side is to draw the connection between the previously mentioned works and the field of adversarial attacks. On top of that, the proofs that are adapted to our scenario allow for additional insights into the concrete application we consider. Beyond single adversarial examples, we also treat distributional adversaries, which we link to curves of maximal slope in the ∞ -Wasserstein space. For potential energies, we derive a (to our knowledge novel) characterization of curves of maximal slope via the superposition principle, which highlights the connection between single adversarial attacks and the distributional adversary. We give more details on the results below.

In section 2, we extend the notion of p-curves of maximal slope to the case $p = \infty$, for Lipschitz curves u. As hinted in the introduction, in the limit $p \to \infty$ of Definition 1.10, we replace (1.6) by the following condition,

$$\begin{aligned} \left|u'\right|(t) &\leq 1, \\ \psi'(t) &\leq -g(u(t)). \end{aligned}$$

Such curves are then called ∞-curves of maximal slope. We want to highlight that similar considerations already appeared in the early works of De Giorgi, see for example, [31, Definition 1.2] and [43, Example 1.3]. For our concrete setup here, we dedicate section 2 to an existence proof of such curves. We note that this can also be obtained as a corollary of a more general existence result in [87, Theorem 3.5].

https://doi.org/10.1017/S0956792525100120 Published online by Cambridge University Press

Therein, the authors prove existence of curves of maximal slope fulfilling

$$\psi'(t) \leq -f^*(g(u(t))) - f(\left|u'\right|(t))$$

for a convex and lower semicontinuous function $f : [0, \infty) \to [0, \infty]$. When choosing $f = \chi_{[0,1]}$, we recover our notion of ∞-curves of maximal slope. Although the existence proof in section 2 employs similar concepts, we choose to include it here. On the one hand, the treatment of this specific case allows for certain arguments that are not directly possible in the general case. On the other hand, this already introduces the main steps for the convergence proof in section 3, which can not directly be deduced from [87]. The existence result in Theorem 2.11 is summarized below.

Existence: *Under the assumptions specified in section 2, for every* $\mathcal{E} : \mathcal{S} \to (-\infty, +\infty]$ *and for every* $x^0 \in \operatorname{dom}(\mathcal{E})$*, there exists a 1-Lipschitz curve* $u : [0, T] \to \mathcal{S}$ *with* $u(0) = x^0$*, which is an* ∞*-curve of maximum slope for* $\mathcal{E}$ *with respect to its strong upper gradient* $|\partial\mathcal{E}|$.

In section 3, we consider the specific case of ∞-curves of maximal slope in a Banach space $\mathcal{X}$, and an energy E that is a C^1 perturbation of a convex function. Note that here and in the following, when the functional takes the role of a C^1-perturbation as in section 3, we use the symbol E instead of $\mathcal{E}$. We derive an equivalent characterization of ∞-curves of maximal slope via a differential inclusion. We note that this differential inclusion can be obtained from [87, Proposition 8.2], with the same choice of f as for the existence result above. The statement in our setting can be found in Theorem 3.8 and is summarized below.

Differential inclusion: *Let* $E : \mathcal{X} \to (-\infty, +\infty]$ *satisfy (3.7) and* $u : [0, 1] \to \mathcal{X}$ *be an a.e. differentiable Lipschitz curve. Let further* $E \circ u$ *be a.e. equal to a non-increasing function* ψ*, then the following are equivalent:*

(i) $|u'|(t) \leq 1$ *and* $\psi'(t) \leq -|\partial E|(u(t))$ *for a.e.* $t \in [0, 1]$,

(ii) $u'(t) \in \partial \|\cdot\|_*(-\xi) \quad \forall \xi \in \partial^\circ E(u(t)) \neq \emptyset$, *for a.e.* $t \in [0, 1]$,

where $\partial^\circ E(u(t))$ *denotes the elements of minimal norm of* $\partial E(u(t))$.

For an energy $E = E^{\mathrm{d}} + E^{\mathrm{c}}$ consisting of a differentiable part E^{d} and a convex part E^{c}, we consider the linearization in the differentiable part around a point z,

$$E^{\mathrm{sl}}(x; z) := E^{\mathrm{d}}(z) + \langle DE^{\mathrm{d}}(z), x - z\rangle + E^{\mathrm{c}}(x).$$

This then leads us to the semi-implicit minimizing movement scheme in Definition 3.10

$$x^{k+1}_{\mathrm{si},\tau} \in \operatorname*{arg\,min}_{x \in \overline{B_\tau}(x^k_{\mathrm{si},\tau})} E^{\mathrm{sl}}(x; x^k_{\mathrm{si},\tau}),$$

which we also employ to approximate curves of maximal slope. In the case of $p = 2$, we refer to [40, 98] for other works that also consider approximate minimizing movement schemes. This semi-implicit scheme is useful, since in the finite dimensional adversarial setting, it allows us to choose $-\ell(h(\cdot), y)$ as the differentiable part, and additionally to incorporate the budget constraint via the indicator function $\chi_{\overline{B_\varepsilon}(x)}$. We denote by $\bar{x}_{\mathrm{si},\tau}$ the step function associated to the iterates $x^k_{\mathrm{si},\tau}$, see Definition 3.10. We can show that up to a subsequence, this scheme also converges to an ∞-curve of maximum slope in the topology σ as specified in Assumption 1.a. The result can be found in Theorem 3.16, which we hint at below.

Convergence to curves of maximal slope: *Under the assumptions specified in section 3, there exists a* ∞*-curve of maximal slope* u *and a subsequence of* $\tau_n = T/n$ *such that*

$$\bar{x}_{\mathrm{si},\tau_n}(t) \xrightarrow{\sigma} u(t) \text{ as } n \to \infty \quad \forall t \in [0, T].$$

In order to better understand the connection between the differential inclusion and (IFGSM), we want to highlight that ∞-curves of maximum slope yield a general concept, which is not directly tied to

https://doi.org/10.1017/S0956792525100120 Published online by Cambridge University Press

signed gradient descent and the choice of the projection. The intuition behind ∞-curves is rather connected to employing normalized gradient descent (NGD) [27]. Choosing $(\mathbb{R}^d, \|\cdot\|_p)$ as the underlying Banach space, in section 5 we see that for $1/p + 1/q = 1$ the following iteration fulfils the semi-implicit minimizing movement scheme,

$$x^{k+1} = x^k + \tau \ \operatorname{sign}(\nabla_x E(x^k)) \cdot \left(\frac{|\nabla_x E(x^k)|}{\|\nabla_x E(x^k)\|_q} \right)^{q-1},$$

where the absolute value and multiplication are understood entrywise. Choosing $p = 2$ or $p = \infty$ recovers the notion of NGD as in [27]. Normalized gradient methods have gained significant attention outside the adversarial context. For example, in the context of saddle point evasion [50, 60, 75], subgradient corruption [103], machine learning [28] and even variational quantum algorithms [100]. In the setting of adversarial attacks, normalization means that we want to ensure that the iterates exploit the maximum allowed budget (locally on $\overline{B_\tau}(x^k)$ ball) in each step. This was similarly observed in [35]. As long as the iterates stay within the given budget ε, one can directly show that (IFGSM) is an explicit solution to the semi-implicit scheme and therefore converges to ∞-curves of maximum slope. In the more interesting case, where the projection has an effect, we need to ensure that minimizing on $\overline{B^p_\tau}(x)$ and then projecting to $\overline{B^p_\varepsilon}(x^0)$ is equivalent to directly minimizing on $\overline{B^p_\varepsilon}(x^0) \cap \overline{B^p_\tau}(x)$. We show this property for the case $p = \infty$ in Lemma 5.4. Employing the convergence result for the semi-implicit minimizing movement scheme, then yields the convergence up to subsequences of (IFGSM), employing the ℓ^∞ norm. Denoting by $x^k_{\mathrm{IFGS},\tau}$ the k-th iterate obtained in (IFGSM) with stepsize τ, Corollary 5.3 then presents the following result.

Convergence of IFGSM: *Under the assumptions specified in section 5, for $T > 0$, there exists a ∞-curve of maximal slope $u : [0, T] \to \mathbb{R}^d$, with respect to E, and a subsequence of $\tau_n := T/n$ such that*

$$\left\| x^{\lceil t/\tau_{n_i} \rceil}_{\mathrm{IFGS},\tau_{n_i}} - u(t) \right\| \xrightarrow{i\to\infty} 0 \qquad \text{for all } t \in [0, T].$$

In section 4, we consider potential energies

$$\mathcal{E} : W_\infty(\mathcal{X}) \ni \mu \mapsto \int E(x)\, \mathrm{d}\mu(x),$$

where in our context, the potential $E : \mathcal{X} \to (-\infty, +\infty]$ has the form $E(x) = -\ell(h(x), y)$. The basis for our main result in this section is given by [63, Theorem 3.1], which is repeated as Theorem 4.7 in this paper. Namely, we characterize absolutely continuous curves $\mu \in \mathrm{AC}^p(0, T; \mathcal{W}_p)$ by a measure η on the space of curves $u : [0, T] \to \mathcal{X}$, which is concentrated on $\mathrm{AC}^p(0, T; \mathcal{X})$. Using this representation, in Theorem 4.18, we show that being a ∞-curve of maximum slope in the Wasserstein space is equivalent to the differential inclusion on the underlying Banach space, for η-a.e. curve.

Characterization of curves in Wasserstein space: *Under the assumptions specified in Theorem 4.18, for a curve $\mu \in \mathrm{AC}^\infty(0, T; \mathcal{W}_\infty)$ with η from Theorem 4.7, the following statements are equivalent:*

(i) The curve μ is ∞-curve of maximal slope w.r.t. to the weak upper gradient $|\partial \mathcal{E}|$.

(ii) For η-a.e. curve $u \in C(0, T; \mathcal{X})$ it holds that $E \circ u$ is for a.e. $t \in (0, T)$, equal to a non-increasing map ψ_u and

$$u'(t) \ \in \ \partial \|\cdot\|_*(-\xi) \quad \forall \xi \in \partial^\circ E(u(t)) \neq \emptyset, \quad \text{for a.e. } t \in (0, T).$$

When applying this result to adversarial training, we slightly deviate from the Wasserstein setting by choosing the extended distance in (1.1) and the associated transport distance in order to prohibit mass transport into the label direction.

Here, we want to refer to other works considering distributional adversarial attacks, e.g., [18, 23, 66, 81, 82, 96, 97, 107]. We can adjust the arguments in section 4 to derive an analogous result for the energy $\mathcal{E}(\mu) := \int -\ell(h(x), y)\, \mathrm{d}\mu(x, y)$, which we state in Theorem 5.10. Here, we only enforce the budget constraint by setting the end time of the flow to $T = \varepsilon$.

https://doi.org/10.1017/S0956792525100120 Published online by Cambridge University Press

1.3. Outline

The paper is organized as follows: In section 2, we start by introducing ∞-curves of maximal slope, as the limit case of p-curves of maximal slope. Section 2.3 then provides an existence result for those curves in a general metric setting. The underlying assumptions for its proof are stated in Section 2.1.

In section 3, we consider ∞-curves of maximal slope when the underlying metric space is a Banach space. Section 3.1 introduces C^1-perturbations of convex functions as a convenient class of functionals that covers most of the energies we consider in this paper. In section 3.2, we derive equivalent characterizations of ∞-curves of maximal slope via a doubly nonlinear differential inclusion. This section is concluded by investigating first-order approximation techniques of those differential inclusions in section 3.3.

Section 4 is devoted to ∞-curves of maximal slope, when the underlying space is the ∞-Wasserstein space. For potential energies, we give an equivalent characterization of ∞-curves of maximal slope via a probability measures η on the space $C(0,T;\mathcal{X})$, which is concentrated on ∞-curves of maximal slope on the underlying Banach space $\mathcal{X}$. From η, we can then derive a corresponding continuity equation for those curves of maximal slope.

In section 5, we discuss the application of differential inclusions derived in section 3 to generate adversarial examples. We show that the popular FGSM and its iterative variant (IFGSM) are simple first-order approximations of ∞-curves of maximal slope. Insection 5.2, we rewrite adversarial training as a distributional robust optimization problem and discuss the usage of ∞-curves in the corresponding probability space to generate distributional adversaries.

2. Infinity flows in metric spaces

In this section, we generalize the notion of p-curves of maximal slope to the case $p=\infty$. We consider the convex function $f(x)=\frac{1}{p}|x|^p$, which allows us to express the energy dissipation inequality (1.6) in Definition 1.10 as follows,

$$\psi'(t) \le -f(|u'|(t)) - f^*(g(u(t))), \tag{2.1}$$

where $f^*(x^*)=\frac{1}{q}|x^*|^q$ denotes the convex conjugate of f. Considering the above inequality for arbitrary convex functions f leads to the general framework as introduced in [87]. For our setting, we consider the indicator function, which is obtained as the following pointwise limit,

$$\frac{1}{p}|x|^p \xrightarrow{p\to\infty} \chi_{[-1,1]}(x) = \begin{cases} 0 \text{ if } |x|\le 1, \\ +\infty \text{ else,} \end{cases}$$

where $\chi_{[-1,1]}$ is a convex function with conjugate $\chi^*_{[-1,1]}(x^*)=x^*$. Using $f=\chi_{[-1,1]}$ in (2.1) forces the curves of maximal slope to obey $|u'|\le 1$ almost everywhere and the energy dissipation inequality becomes

$$\psi'(t) \le -(g\circ u)(t),$$

which motivates the following definition.

Definition 2.1 (∞-Curve of maximal slope). *We say an absolutely continuous curve $u:[0,T]\to\mathcal{S}$ is an ∞-curve of maximal slope for the functional $\mathcal{E}$ with respect to an upper gradient g, if $\mathcal{E}\circ u$ is a.e. equal to a non-increasing map ψ and*

$$\begin{aligned} &|u'|(t) \le 1, \\ &\psi'(t) \le -(g\circ u)(t), \end{aligned} \tag{InfFlow}$$

holds for a.e. $t\in(0,T)$.

Remark 2.2. We note that the condition $|u'|\le 1$ a.e. implies that u is a Lipschitz curve with Lipschitz constant 1, see Lemma 1.1.

https://doi.org/10.1017/S0956792525100120 Published online by Cambridge University Press

Remark 2.3. (*Dissipation equality*). If g is a strong upper gradient of $\mathcal{E}$ and $\psi:[0,T]\to\mathbb{R}$ is finite then by Definition 1.4 and (InfFlow)

$$|\mathcal{E}(u(t))-\mathcal{E}(u(s))|\le\int_s^t g(u(r))|u'|(r)\,\mathrm{d}r\le\int_s^t g(u(r))\,\mathrm{d}r\le\int_s^t -\psi'(r)\,\mathrm{d}r\le\psi(s)-\psi(t)<+\infty,$$

where in the last inequality we use that non-increasing functions are differentiable a.e. and an upper bound on the second fundamental theorem of calculus holds [102, Proposition 1.6.37]. This in particular implies that $\mathcal{E}\circ u$ is absolutely continuous and $\psi(t)=(\mathcal{E}\circ u)(t)$ for all $t\in(0,T)$ (see Lemma E.1). Furthermore, Remark 1.3 implies

$$\mathcal{E}(u(t))-\mathcal{E}(u(s))=\int_s^t(\mathcal{E}\circ u)'(r)\,\mathrm{d}r\quad\text{for } 0\le s\le t\le T$$

and we can estimate

$$\mathcal{E}(u(t))-\mathcal{E}(u(s))=\int_s^t(\mathcal{E}\circ u)'(r)\,\mathrm{d}r\le\int_s^t -d(u(r))\,\mathrm{d}r\quad\text{for } 0\le s\le t\le T$$

and on the other hand, using (1.5), we obtain

$$\begin{aligned}\mathcal{E}(u(t))-\mathcal{E}(u(s))&=\int_s^t(\mathcal{E}\circ u)'(r)\,\mathrm{d}r\ge\int_s^t -|(\mathcal{E}\circ u)'(r)|\,\mathrm{d}r\\&\ge\int_s^t -g(u(r))|u'|(r)\,\mathrm{d}r\ge\int_s^t -g(u(r))\,\mathrm{d}r\end{aligned}$$

for $0\le s\le t\le T$. Therefore, the energy dissipation equality

$$\mathcal{E}(u(t))-\mathcal{E}(u(s))=\int_s^t -g(u(r))\,\mathrm{d}r\tag{EnDisEq}$$

holds for every $0\le s\le t\le T$.

Example 1. *As an easy example, let us look at the quadratic energy* $\mathcal{E}:x\mapsto\frac{1}{2}x^2$ *on the space* $(\mathcal{S},d)=(\mathbb{R},|\cdot-\cdot|)$. *Its metric slope and thus weak upper gradient is given by* $|\partial\mathcal{E}|(x)=\left|\frac{d}{dx}\mathcal{E}\right|(x)=|x|$. *We choose* $x^0=1$ *as the starting point, then the corresponding* ∞*-curve of maximal slope is*

$$u(t)=\begin{cases}1-t & \text{if } 0\le t\le 1,\\ 0 & \text{if } t>1\end{cases}.$$

We directly observe that $|u'|\le 1$ *and*

$$\mathcal{E}(u(t))=\begin{cases}\frac{1}{2}(1-t)^2 & \text{if } 0\le t\le 1,\\ 0 & \text{if } t>1,\end{cases}$$

is a non-increasing map with

$$\frac{d}{dt}\mathcal{E}(u(t))=\begin{cases}t-1 & \text{if } 0\le t<1,\\ 0 & \text{if } t>1,\end{cases}=-|u(t)|=-|\partial\mathcal{E}|(u(t)),$$

and therefore the conditions (InfFlow) are fulfilled. Here we can already observe a typical behaviour of ∞*-curves of maximal slope. They have a constant velocity of* 1 *until they hit a local minimum where they stop abruptly.*

The rest of this section is devoted to an existence proof for ∞-curves of maximal slope.

2.1. Assumptions for existence

Here, we state the assumptions needed for the proof of existence. Approximations of curves of maximal slope are constructed via a minimizing movement scheme. To guarantee convergence of those

https://doi.org/10.1017/S0956792525100120 Published online by Cambridge University Press

approximations, a form of relative compactness is essential. This is guaranteed by Assumption 1.b. Furthermore, relative compactness with respect to the topology induced by the metric $d(\cdot,\cdot)$ may not be given. However, relative compactness with respect to a weaker topology σ is sufficient, as long as it is compatible with the topology induced by the metric $d(\cdot,\cdot)$, Assumption 1.a. These assumptions were also employed in [2].

Assumption 1.a (Weak topology). *In addition to the metric topology, $(\mathcal{S}, d)$ is assumed to be endowed with a Hausdorff topology σ. We assume that σ is compatible with the metric d, in the sense that σ is weaker than the topology induced by d and d is sequentially σ-lower semicontinuous, i.e.,*

$$(x^n, z^n) \overset{\sigma}{\rightharpoonup} (x, z) \Longrightarrow \liminf_{n\to\infty} d(x^n, z^n) \geq d(x, z).$$

Assumption 1.b (Relative compactness). *Every d-bounded set contained in sublevels of $\mathcal{E}$ is relatively σ-sequentially compact, i.e.,*

$$\text{if} \quad \{x^n\}_{n\in\mathbb{N}} \subset \mathcal{S} \quad \text{with} \quad \sup_{n\in\mathbb{N}} \mathcal{E}(x^n) < +\infty, \quad \sup_{n,m} d(x^n, x^m) < +\infty,$$

$$\text{then } (x^n)_{n\in\mathbb{N}} \text{ admits a } \sigma\text{-convergent subsequence.}$$

Assumptions 2.a and 2.b ensure the lower semicontinuity of the energy functional and the lower semicontinuity of its metric slope. These regularity assumptions are required for the energy dissipation inequality during the limiting process in the proof of Theorem 2.11.

Assumption 2.a (Lower semicontinuity). *We assume sequential σ-lower semicontinuity of $\mathcal{E}$ for bounded sequences, namely,*

$$\left.\begin{aligned} \sup_{n,m\in\mathbb{N}} \{d(x^n, x^m)\} < +\infty, \\ x^n \overset{\sigma}{\rightharpoonup} x \end{aligned}\right\} \Longrightarrow \mathcal{E}(x) \leq \liminf_{n\to\infty} \mathcal{E}(x^n). \tag{2.2}$$

Assumption 2.b (Lower semicontinuity of slope). *In addition, we ask that $|\partial\mathcal{E}|$ is a strong upper gradient and it is sequentially σ-lower semicontinuous on d-bounded sublevels of $\mathcal{E}$.*

Remark 2.4. The proof of existence is possible with a wide variety of regularity assumptions on the energy $\mathcal{E}$, which can be tailored to a variety of different situations. For example, if the sequentially σ-lower semicontinuous envelope of $|\partial\mathcal{E}|$

$$|\partial^-\mathcal{E}| := \left\{ \liminf_{n\to\infty} |\partial\mathcal{E}|(x^n) : x^n \overset{\sigma}{\rightharpoonup} x, \sup_n d(x^n, x), \mathcal{E}(x^n) < +\infty \right\}$$

is a strong upper gradient, one can drop Assumption 2.b and instead prove existence of curves of maximal slope with respect to $|\partial^-\mathcal{E}|$. Further, if $|\partial\mathcal{E}|$ (or $|\partial^-\mathcal{E}|$ respectively) is only a weak upper gradient (compare [2, Theorem 2.3.3]), then Assumption 2.a has to be replaced by continuity of the energy.

2.2. Minimizing movement for $p = \infty$

The minimizing movement scheme is an implicit time discretization of curves of maximal slope. The existence of curves of maximal slope is proven by sending the discrete time step τ of the minimizing movement scheme to 0. For the time interval $[0, T]$ and some $n \in \mathbb{N}$, we use an equidistant time discretization $t^k = k \cdot \tau$ for $k \in \{0, \ldots, n\}$ with $\tau = T/n$. Starting with $x_\tau^0 = x^0$ the classical minimizing movement scheme to approximate p-curves of maximum slope reads

$$x_\tau^{k+1} \in \arg\min_{\tilde{x}\in\mathcal{S}} \left\{ \frac{1}{p\tau^{p-1}} d^p(\tilde{x}, x_\tau^k) + \mathcal{E}(\tilde{x}) \right\}.$$

Taking formally the limit $p \to \infty$ under the constraint $d(\tilde{x}, x_\tau^k) \leq \tau$, we arrive at the corresponding minimizing movement scheme for $p = \infty$, which we define in the following.

https://doi.org/10.1017/S0956792525100120 Published online by Cambridge University Press

Definition 2.5 (Minimizing movement scheme for $p = \infty$). *For $\tau = T/n$ and $x_\tau^0 = x^0$, we consider the iteration defined for $k \in \mathbb{N}_0$ as*

$$x_\tau^{k+1} \in \arg\min_{\tilde{x} \in \mathcal{S}} \{\mathcal{E}(\tilde{x}) : d(\tilde{x}, x_\tau^k) \leq \tau\}. \tag{MinMove}$$

We define the step function $\bar{x}_\tau$ by

$$\bar{x}_\tau(0) = x^0, \quad \bar{x}_\tau(t) = x_\tau^k \text{ if } t \in (t_\tau^{k-1}, t_\tau^k], k \geq 1.$$

Furthermore, we define

$$|x_\tau'|(t) := \frac{d\left(x_\tau^k, x_\tau^{k-1}\right)}{t_\tau^k - t_\tau^{k-1}} \text{ if } t \in \left(t_\tau^{k-1}, t_\tau^k\right),$$

as the metric derivative of the corresponding piecewise affine linear interpolation.

Assumptions 2.a and 1.b guarantee the existence of minimizers in (MinMove) via the direct method in the calculus of variations [30], which ensures that the minimizing movement scheme can be defined. Now for all $x \in \mathcal{S}$, we set

$$\mathcal{E}_\tau(x) := \min_{\tilde{x} \in \overline{B_\tau(x)}} \mathcal{E}(\tilde{x}). \tag{2.3}$$

Remark 2.6. The function defined in (2.3) is similarly employed in [3, 16, 17] and the proof strategy as displayed in Figure 2 resembles the max-ball arguments as in the previously mentioned works. The expression in (2.3) can also be seen as the infimal convolution [39, 49] of $\mathcal{E}$ and $\chi_{\overline{B_\tau(0)}}$, i.e., $\mathcal{E}_\tau = \chi_{\overline{B_\tau}} \square \mathcal{E}$ and can also be considered as the limit $p \to \infty$ of the Moreau envelope [72],

$$\inf_{\tilde{x}} \left\{ \mathcal{E}(\tilde{x}) + \frac{1}{p} \|x - \tilde{x}\|^p \right\}$$

which is typically defined for $p = 2$.

Remark 2.7. More recently, similar schemes to the one defined in (MinMove) have been introduced in an optimization context in [48]. Here, the operation on the right-hand side of (MinMove) was labelled the "ball-proximal" or "brox" operator.

The next lemma gives an equivalent characterization of the metric slope and provides its relation to the minimizing movement scheme. In fact, it is a special case of [2, Lemma 3.1.5, Remark 3.1.7]. For completeness, we provide an adapted proof in Appendix E.

Lemma 2.8. *For all $x \in \operatorname{dom}(\mathcal{E})$, we have that*

$$|\partial \mathcal{E}|(x) = \limsup_{\tau \to 0^+} \frac{\mathcal{E}(x) - \mathcal{E}_\tau(x)}{\tau}. \tag{2.4}$$

Further, we are interested in the behaviour of the mapping $\tau \mapsto \mathcal{E}_\tau(x)$ when varying τ. By definition, it is monotone decreasing in τ and thus differentiable a.e. This allows us to derive an integral inequality that gives an upper bound to $\mathcal{E}_\tau(x)$ as τ increases.

Lemma 2.9 (Differentiability of $\mathcal{E}_\tau(x)$). *For $x \in \operatorname{dom}(\mathcal{E})$, the derivative $\frac{d}{d\tau}\mathcal{E}_\tau(x)$ exists for a.e. $\tau \in (0, +\infty)$ and*

$$\mathcal{E}_{\tau_1}(x) + \int_{\tau_1}^{\tau_2} \frac{d}{d\tilde{\tau}} \mathcal{E}_{\tilde{\tau}}(x) d\tilde{\tau} \geq \mathcal{E}_{\tau_2}(x) \quad \text{for } 0 \leq \tau_1 \leq \tau_2 < +\infty. \tag{2.5}$$

Furthermore,

$$\frac{d}{d\tau} \mathcal{E}_\tau(x) \leq -|\partial \mathcal{E}|(x_{\min,\tau}) \quad \text{for a.e. } \tau \in (0, +\infty), \tag{2.6}$$

where

$$x_{\min,\tau} \in \arg\min_{\tilde{x}} \{\mathcal{E}(\tilde{x}) : d(x, \tilde{x}) \leq \tau\}. \tag{2.7}$$

https://doi.org/10.1017/S0956792525100120 Published online by Cambridge University Press

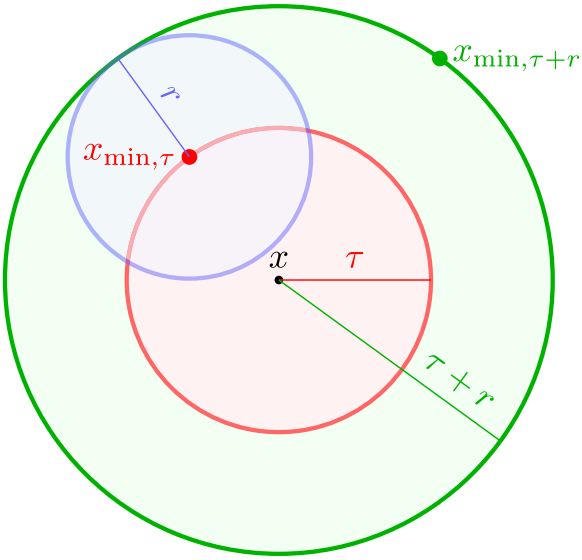


Figure 2. *Visualization of the ball inclusion used for the proof of (2.6).*

Proof. Let $x \in \operatorname{dom}(\mathcal{E})$, for any $\tau^* < \infty$ we know that the mapping $\tau \mapsto \mathcal{E}_\tau(x)$ is monotone decreasing on $[0, \tau^*]$ and thus its variation can be bounded,

$$\mathcal{E}_0(x) - \mathcal{E}_{\tau^*}(x) = \mathcal{E}(x) - \mathcal{E}(x_{\min,\tau^*}) < \infty.$$

Employing [90, Theorem 9.6, Chapter IV], this yields that the derivative exists for almost every $t \in (0, \tau^*)$ and that (2.5) holds. To show (2.6), we observe that

$$B_r(x_{\min,\tau}) \subset B_{\tau+r}(x) \text{ and thus } \mathcal{E}_{\tau+r}(x) \leq \mathcal{E}_r(x_{\min,\tau}),$$

see Figure 2, which yields

$$-\left(\frac{\mathcal{E}(x_{\min,\tau}) - \mathcal{E}_{\tau+r}(x)}{r}\right) \leq -\left(\frac{\mathcal{E}(x_{\min,\tau}) - \mathcal{E}_r(x_{\min,\tau})}{r}\right).$$

It follows that

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\tau}\mathcal{E}_\tau(x) &= \lim_{r\to 0} \frac{\mathcal{E}_{\tau+r}(x) - \mathcal{E}_\tau(x)}{r} \\ &= -\lim_{r\to 0} \frac{\mathcal{E}(x_{\min,\tau}) - \mathcal{E}_{\tau+r}(x)}{r} \\ &\leq -\limsup_{r\to 0} \frac{\mathcal{E}(x_{\min,\tau}) - \mathcal{E}_r(x_{\min,\tau})}{r} = -|\partial\mathcal{E}|(x_{\min,\tau}),\end{aligned}$$

where we used the characterization of the slope from Lemma 2.8. □

2.3. Proof of existence

Together with the previous lemmas, we are now able to prove the existence of ∞-curves of maximal slope. Besides the piecewise constant interpolation $\bar{x}$, we use a variational interpolation. This interpolation, combined with estimate in (2.9), later yields the differential inequality (InfFlow).

https://doi.org/10.1017/S0956792525100120 Published online by Cambridge University Press

Definition 2.10 (De Giorgi variational interpolation). *We denote by* $\tilde{x}_\tau : [0,T] \to \mathcal{S}$ *any interpolation of the discrete values satisfying*

$$\tilde{x}_\tau(t) \in \arg\min_{\tilde{x}} \left\{ \mathcal{E}(x) : d\left(\tilde{x}, x_\tau^{k-1}\right) \le t - t_\tau^{k-1} \right\}$$

if $t \in (t_\tau^{k-1}, t_\tau^k]$ *and* $k \ge 1$. *Furthermore, we define*

$$D_\tau(t) := \frac{d}{dt} \mathcal{E}_{(t-t_\tau^{k-1})}\left(x_\tau^{k-1}\right) \qquad \text{if } t \in \left(t_\tau^{k-1}, t_\tau^k\right]. \tag{2.8}$$

Employing Lemma 2.9, the above definition directly yields

$$\mathcal{E}(\tilde{x}_\tau(s)) + \int_s^t D_\tau(r)\, \mathrm{d}r \ge \mathcal{E}(\tilde{x}_\tau(t)) \quad \forall\, 0 \le s \le t \le T, \tag{2.9}$$

which is used in the following existence proof, Theorem 2.11. We employ the arguments of [2, Ch. 3] and transfer them to our setting, where a crucial statement is the refined version of Ascoli–Arzelà in [2, Proposition 3.3.1], which is repeated for convenience, in the appendix, see Proposition B.1.

As detailed in section 1, this can also be obtained via the results in [87]. Nevertheless, we include a proof here, since this introduces the main arguments for the proof of Theorem 3.16.

Theorem 2.11 (Existence of ∞-curves of maximal slope). *Under the Assumptions 1.a to 2.b for every* $x^0 \in \mathrm{dom}\,(\mathcal{E})$, *there exists a 1-Lipschitz curve* $u : [0,T] \to \mathcal{S}$ *with* $u(0) = x^0$, *which is an* ∞*-curve of maximum slope for* $\mathcal{E}$ *with respect to its strong upper gradient* $|\partial\mathcal{E}|$ *and* u *satisfies the energy dissipation equality*

$$\mathcal{E}(u(0)) = \mathcal{E}(u(t)) + \int_0^t |\partial\mathcal{E}|(u(r))\, dr \quad \text{for all } t \in [0,T]. \tag{2.10}$$

Proof. We consider the set of all possible iterates in the minimizing movement scheme $K = \{x_{\tau_n}^i : 0 \le i \le n, n \in \mathbb{N}\} \subset \mathcal{S}$. Recalling Definition 2.5, for every $n \in \mathbb{N}$ and $i, j \in \{0, \dots, n\}$, we have the estimate

$$d(x_{\tau_n}^i, x_{\tau_n}^j) \le \sum_{k=i}^{j-1} d\left(x_{\tau_n}^k, x_{\tau_n}^{k+1}\right) \le (j-i)\, \tau_n \le T,$$

and therefore for every $n, m \in \mathbb{N}$ and $0 \le i \le n,\ 0 \le j \le m$, we have

$$d(x_{\tau_n}^i, x_{\tau_m}^j) \le d(x_{\tau_n}^i, x^0) + d(x^0, x_{\tau_m}^j) \le 2T.$$

Furthermore, since $x^0 \in \mathrm{dom}\,(\mathcal{E})$, we also know that $\mathcal{E}(x_{\tau_n}^i) \le \mathcal{E}(x^0) < \infty$ and thus K is a d-bounded set, contained in sublevels of $\mathcal{E}$. Using relative compactness, i.e., Assumption 1.b, this ensures that $\overline{K}$ is a σ-sequentially compact set and therefore fulfils 1 of Proposition B.1. In order to apply the latter, it remains to choose a function ω that fulfils 2. For this, we consider the sequence of curves $\left|x'_{\tau_n}\right| : [0,T] \to \mathbb{R}$, which is by definition bounded in $L^\infty(0,T)$, i.e.,

$$\left\| \left|x'_{\tau_n}\right| \right\|_{L^\infty(0,T)} \le 1, \qquad \text{for every } n \in \mathbb{N}.$$

For fixed $0 \le s \le t \le T$, let us define

$$s(n) := \min_{k \in \{0,\dots,n\}} \{k \cdot \tau_n : s \le k \cdot \tau_n\}, \qquad t(n) := \min_{k \in \{0,\dots,n\}} \{k \cdot \tau_n : t \le k \cdot \tau_n\}. \tag{2.11}$$

Using the triangle inequality and the fact that the distance between two consecutive iterates is bounded by τ, we obtain

$$\limsup_{n \to +\infty} d(\bar{x}_{\tau_n}(s), \bar{x}_{\tau_n}(t)) \le \limsup_{n \to +\infty} \sum_{i=1}^{\frac{t(n)-s(n)}{\tau_n}} d(\bar{x}_{\tau_n}(s(n) + (i-1)\tau), \bar{x}_{\tau_n}(s(n) + i\tau_n)) \tag{2.12}$$

$$\le \lim_{n \to +\infty} (t(n) - s(n)) = |t - s| =: \omega(s,t). \tag{2.13}$$

https://doi.org/10.1017/S0956792525100120 Published online by Cambridge University Press

Therefore, 2 in Proposition B.1 is fulfilled, allowing us to apply [2, Proposition 3.3.1] to extract another subsequence such that

$$\bar{x}_{\tau_n}(t) \overset{\sigma}{\rightharpoonup} u(t) \text{ as } n \to \infty \quad \forall t \in [0,T], \quad u \text{ is } d\text{-continuous in } [0,T].$$

This in particular ensures $u(0) = x^0$ and (2.12) together with Assumption 1.a yields 1-Lipschitzness of u, since for $s \le t$, we have

$$d(u(s), u(t)) \le \liminf_{n\to\infty} d(\bar{x}_{\tau_n}(s), \bar{x}_{\tau_n}(t)) \le t - s.$$

By construction, it holds that $d(\bar{x}_{\tau_n}, \tilde{x}_{\tau_n}) \le \tau$, which also yields

$$\tilde{x}_{\tau_n}(t) \overset{\sigma}{\rightharpoonup} u(t) \text{ as } n \to \infty \quad \forall t \in [0,T].$$

Observing that $\tilde{x}_{\tau_n}(0) = x^0 = u(0)$ independent of n, we take the limes inferior for (2.9) and use Assumption 2.a and Fatou's lemma to obtain for all $t \in [0,T]$

$$\begin{aligned}\mathcal{E}(u(0)) &\ge \liminf_{n\to\infty} \left\{ \mathcal{E}(\tilde{x}_{\tau_n}(t)) - \int_0^t D_{\tau_n}(r)\,\mathrm{d}r \right\} \ge \mathcal{E}(u(t)) + \int_0^t \liminf_{n\to\infty} -D_{\tau_n}(r)\,\mathrm{d}r \\ &\ge \mathcal{E}(u(t)) + \int_0^t |\partial\mathcal{E}(u(r))|\,\mathrm{d}r.\end{aligned}$$

The last inequality follows by the estimate

$$|\partial\mathcal{E}|(u(t)) \le \liminf_{n\to\infty} |\partial\mathcal{E}|(\tilde{x}_{\tau_n}(t)) \le \liminf_{n\to\infty} -D_{\tau_n}(t) \quad \text{for a.e.} \quad t \in (0,T),$$

which is a consequence of (2.8) and (2.6) and the σ-lower semicontinuity of the slope. On the other hand, we know that $|\partial\mathcal{E}|$ is a strong upper gradient and $|u'|(r) \le 1$ for a.e. $r \in [0,T]$, such that

$$\mathcal{E}(u(0)) \le \mathcal{E}(u(t)) + \int_0^t |\partial\mathcal{E}|(u(r))|u'|(r)\,\mathrm{d}r \le \mathcal{E}(u(t)) + \int_0^t |\partial\mathcal{E}|(u(r))\,\mathrm{d}r.$$

In particular, the equality

$$\mathcal{E}(u(0)) = \mathcal{E}(u(t)) + \int_0^t |\partial\mathcal{E}|(u(r))|u'|(r)\,\mathrm{d}r = \mathcal{E}(u(t)) + \int_0^t |\partial\mathcal{E}|(u(r))\,\mathrm{d}r$$

must hold. It follows that $t \mapsto \mathcal{E}(u(t))$ is locally absolutely continuous and

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{E}(u(t)) = -|\partial\mathcal{E}|(u(t))|u'|(t) = -|\partial\mathcal{E}|(u(t)) \text{ for a.e.} t \in (0,T).$$

□

3. Banach space setting

In this section, we consider the Banach space setting, i.e., we assume that $\mathcal{S} = \mathcal{X}$, where $\mathcal{X}$ is a Banach space with norm $\|\cdot\|$ and $(\mathcal{X}^*, \|\cdot\|_*)$ denoting its dual. In this section, we assume the functional to be a C^1-perturbation (see section 3.1) and use the symbol E to distinguish it from general functionals $\mathcal{E}$ in the previous section. We want to give an equivalent characterization of curves of maximum slope in terms of differential inclusions. Following [2, Ch. 1], for a functional $E : \mathcal{X} \to (-\infty, \infty]$, we employ the Fréchet subdifferential $\partial E \subset \mathcal{X}^*$, where for $x \in \mathrm{dom}\,(E)$, we define

$$\xi \in \partial E(x) \Leftrightarrow \liminf_{z\to x} \frac{E(z) - E(x) - \langle \xi, z - x\rangle}{\|z - x\|} \ge 0 \tag{3.1}$$

with $\mathrm{dom}\,(\partial E) = \{x \in \mathcal{X} : \partial E(x) \ne \emptyset\}$. Assuming that $\partial E(x)$ is weakly* closed for every $x \in \mathrm{dom}\,(\partial E)$ – which holds true in particular, if $\mathcal{X}$ is reflexive or E is a so called C^1-perturbation of a convex function (see Proposition 3.1) – we furthermore define

$$\partial^\circ E(x) := \operatorname*{arg\,min}_{\xi \in \partial E(x)} \|\xi\|_* \subset \partial E(x).$$

https://doi.org/10.1017/S0956792525100120 Published online by Cambridge University Press

Note that $\partial^\circ E(x)$ is still potentially multivalued; however, all elements have the same dual norm. This justifies using the notation $\|\partial^\circ E(x)\|_* = \min\{\|\xi\|_* : \xi \in \partial E(x)\}$ in the following.

3.1. On C^1-perturbations of convex functions

Functions that can be split into a convex function E^{c} and a differentiable part E^{d}, i.e., $E = E^{\mathrm{c}} + E^{\mathrm{d}}$, are called C^1-perturbations of convex functions. This particular class of functions exhibits a variety of useful properties. We collect the ones that are relevant for our setting in the following proposition, which is a combination of Corollary 1.4.5 and Lemma 2.3.6 in [2].

Proposition 3.1 (C^1-perturbations of convex functions). *If $E : \mathcal{X} \to (-\infty, +\infty]$ admits a decomposition $E = E^{\mathrm{c}} + E^{\mathrm{d}}$, into a proper, lower semicontinuous convex function E^{c} and a C^1-function E^{d}, then*

(i) $\partial E = \partial E^{\mathrm{c}} + DE^{\mathrm{d}}$,

(ii)

$$\left.\begin{aligned} \xi^n \in \partial E(x^n),\\ x^n \to x \in \operatorname{dom}(\partial E),\\ \xi^n \rightharpoonup^* \xi \end{aligned}\right\} \Rightarrow \left\{\begin{aligned} \xi \in \partial E(x),\\ E(x^n) \to E(x), \end{aligned}\right.$$

(iii) $|\partial E|(x) = \|\partial^\circ E(x)\|_* \quad \forall x \in \mathcal{X}$,

(iv) $|\partial E|$ *is* $\|\cdot\|$*-lower semicontinuous,*

(v) $|\partial E|$ *is a strong upper gradient of* E.

Considering Banach spaces that fulfil Assumptions 1.a and 1.b with their strong topology and energies that are C^1 perturbations, the existence of ∞-curves of maximum slope follows directly by Theorem 2.11.

An important example of such a Banach space $\mathcal{X}$ is the Euclidean space, since our motivating application, namely adversarial attacks, usually employs a finite-dimensional image space. We formulate this result in the following corollary.

Corollary 3.2 (Existence for C^1-perturbations in finite dimensions). *Let $\mathcal{X} = (\mathbb{R}^d, \|\cdot\|)$ and $E : \mathbb{R}^d \to (-\infty, +\infty]$ admit a decomposition $E = E^{\mathrm{c}} + E^{\mathrm{d}}$ into a proper, lower semicontinuous convex function E^{c} and a C^1-function E^{d}. For every $x^0 \in \operatorname{dom}(E)$, there exists at least one curve of maximal slope in the sense of Definition 2.1 with $u(0) = x^0$. Further, this curve satisfies the energy dissipation equality (2.10).*

Proof. We choose σ to be the norm topology, such that Assumptions 1.a and 1.b are fulfilled and E fulfils Assumption 2.a. By Proposition 3.1, $|\partial E|$ is lower semicontinuous and a strong upper gradient. Therefore, also Assumption 2.b is fulfilled and the application of Theorem 2.11 yields the desired result. □

In the infinite-dimensional case, existence is harder to prove. Usually, σ is chosen as the weak or weak$*$ topology, such that when $\mathcal{X}$ is reflexive or a dual space, the Banach–Alaoglu theorem yields compactness and that Assumptions 1.a and 1.b are fulfilled. A desirable property for the energy functional is the so-called σ-weak$*$ closure property

$$\left.\begin{aligned} \xi^n \in \partial E(x^n),\\ x^n \xrightarrow{\sigma} x \in \operatorname{dom}(\partial E),\\ \xi^n \rightharpoonup^* \xi \end{aligned}\right\} \Rightarrow \left\{\begin{aligned} \xi \in \partial E(x),\\ E(x^n) \to E(x) \end{aligned}\right.$$

of its subdifferential, c.f. Item 2. The σ-lower semicontinuity of the slope Assumption 2.b and (3.7) are almost immediate consequences of the closure property, as was shown in [2, Lemma 2.3.6, Theorem 2.3.8].

https://doi.org/10.1017/S0956792525100120 Published online by Cambridge University Press

Example 2. *As an application of Corollary 3.2, we consider the finite-dimensional adversarial setting introduced in section 1, i.e., we choose* $\mathcal{X} = \mathbb{R}^d$. *Let*

$$E(x) := \underbrace{-\ell(h(x), y)}_{=E^{\mathrm{d}}} + \underbrace{\chi_{\overline{B_\varepsilon}(x^0)}}_{E^{\mathrm{c}}},$$

then by the chain rule $E^{\mathrm{d}} \in C^1(\mathcal{X})$, *if* $h \in C^1(\mathcal{X};\mathcal{Y})$ *and* $\ell \in C^1(\mathcal{Y} \times \mathcal{Y})$. *We consider a neural network* $h = \phi^L \circ \ldots \circ \phi^1$ *with the lth layer being given as*

$$\phi^l : \mathbb{R}^{d^l} \to \mathbb{R}^{d^{l+1}}, \phi^l(z) := \alpha(Wz + b),$$

for a weight matrix $W \in \mathbb{R}^{d^{l+1},d^l}$, *bias* $b \in \mathbb{R}^{d^{l+1}}$ *and activation function* $\alpha : \mathbb{R} \to \mathbb{R}$, *which is applied entry-wise. Therefore, the network h is* C^1 *if its activation function is in* $C^1(\mathbb{R})$. *Typical examples that fulfil this assumption are the Sigmoid function and smooth approximations to ReLU [42], such as GeLU [51], see also Appendix H for more details on such activation functions. Furthermore, many popular loss functions are in* $C^1(\mathcal{Y} \times \mathcal{Y})$, *like the mean squared error (MSE) or Cross-Entropy paired with a Softmax layer [9, 29, 45]. On the other hand, the root MSE is not differentiable whenever a component is* 0.

Lemma 2.8 provides an alternative characterization of the metric slope, employing a lim sup formulation. The next two lemmas show that C^1-perturbations are regular enough, such that the limit superior can be replaced by a standard limit. This is used in Lemma 4.17. The first lemma establishes the fact that for convex functionals, there is a minimizing sequence for the value of $E_\tau(x)$ that lies on the boundary $\partial B_\tau(x)$.

Remark 3.3. Similar to [2, section 3.1], we remark some properties of $E_\tau(x) = \inf_{\tilde{x} \in \overline{B_\tau}(x)} E(\tilde{x})$ in the case, when E is convex. Since $E_\tau(x)$ is defined via the infimal convolution, see Remark 2.6, we can directly infer convexity in the x argument, if E was already convex. Furthermore, we also have convexity in τ, which can be seen as follows. Let $\tau_1, \tau_2 \geq 0$ be arbitrary, where we also allow them to attain 0. For any $z_1 \in \overline{B_{\tau_1}}(x)$, $z_2 \in \overline{B_{\tau_2}}(x)$, we have that $\lambda z_1 + (1-\lambda)z_2 \in \overline{B_{\tilde{\tau}}}(x)$ with $\tilde{\tau} = \lambda\tau_1 + (1-\lambda)\tau_2$ for any $\lambda \in [0,1]$. The definition of E_τ and the convexity of E yield

$$E_{\tilde{\tau}}(x) \leq E(\lambda z_1 + (1-\lambda)z_2) \leq \lambda E(z_1) + (1-\lambda)E(z_2)$$

and since $z_1 \in \overline{B_{\tau_1}}(x)$, $z_2 \in \overline{B_{\tau_2}}(x)$ were arbitrary, we obtain

$$E_{\tilde{\tau}}(x) \leq \lambda E_{\tau_1}(x) + (1-\lambda)E_{\tau_2}(x).$$

If $x \in \operatorname{dom}(E)$), we have that $\operatorname{dom}(\tau \mapsto E_\tau(x)) = [0, \infty)$ and thus $\tau \to E_\tau$ is continous on $(0, \infty)$. If E is lower semicontinuous, we also obtain continuity at 0.

Lemma 3.4. *If E is a proper, convex, lower semicontinuous function, then for all* $x \in \operatorname{dom}(E)$ *with* $|\partial E(x)| \neq 0$, *there is an* $\epsilon > 0$ *such that for all* $0 < \tau < \epsilon$, *there exists a sequence* $(x^n)_{n\in\mathbb{N}}$ *with*

$$E(x^n) \to E_\tau(x) \quad \textit{and} \quad \|x - x^n\| = \tau \ \forall n \in \mathbb{N}. \tag{3.2}$$

If in addition the Banach space $\mathcal{X}$ *is reflexive, then there exists* $x_\tau \in \mathcal{X}$ *with*

$$E(x_\tau) = E_\tau(x) \quad \textit{and} \quad \|x - x_\tau\| = \tau.$$

Proof. Let $x \in \operatorname{dom}(E)$ with $|\partial E(x)| \neq 0$, then the mapping $\tau \mapsto E_\tau(x)$ is non-increasing and not constant. Therefore, we can find an $\epsilon > 0$ such that $E_\tau(x) > E_\epsilon(x)$ for all $0 < \tau < \epsilon$. Let $(\tilde{x}^n)_{n\in\mathbb{N}}$ be a sequence such that

$$\lim_{n\to\infty} E(\tilde{x}_n) = E_\tau(x).$$

Since $E_\tau(x) > E_\epsilon(x)$, we can find an element $\hat{x}$ that fulfils

$$E(\tilde{x}_n) > E(\hat{x}) \qquad \text{for every } n \in \mathbb{N} \qquad \text{and} \qquad \tau < \left\|x - \hat{x}\right\| \leq \epsilon.$$

Since $\hat{x} \notin \overline{B_\tau}(x)$ and $\tilde{x}_n \in \overline{B_\tau}(x)$, the line between each pair $(\hat{x}, \tilde{x}_n)$,

$$c_n : t \in [0,1] \mapsto t\hat{x} + (1-t)\tilde{x}_n$$

https://doi.org/10.1017/S0956792525100120 Published online by Cambridge University Press

has to intersect the sphere $\partial B_\tau(x)$ at some point $t_n \in [0,1)$, where we define the intersection point as $x_n = c_n(t_n) \in \partial B_\tau(x)$. Due to convexity, we obtain

$$E_\tau(x) \le E(x_n) = E(t_n\hat{x} + (1-t_n)\tilde{x}_n) \le t_n E(\hat{x}) + (1-t_n)E(\tilde{x}_n) \le E(\tilde{x}_n).$$

Note that the last inequality would only be strict if $t_n \neq 0$; however, since $\tilde{x}_n$ might already be lying on the sphere, we only obtain the weak inequality. The sequence x_n now is the desired sequence in (3.2).

In the reflexive case, the weak compactness of the unit ball guarantees weak convergence of a subsequence of $(x^n)_{n\in\mathbb{N}}$ to some $x_\tau \in \overline{B_\tau}(x)$. Lower semicontinuity and convexity imply weak lower semicontinuity of E and thus

$$E_\tau(x) \le E(x_\tau) \le \liminf_{n\to\infty} E(x_n) = E_\tau(x).$$

As above, we can choose an element $\hat{x}$ with $\|\hat{x} - x\| > \tau$ with $E(x_\tau) > E(\hat{x})$. Applying the same argument as above, there is some $t \in [0,1)$ such that $t\hat{x} + (1-t)x_\tau$ intersects $\partial B_\tau(x)$. As above, if $t \neq 0$, convexity yields

$$E_\tau(x) \le E(t\hat{x} + (1-t)x_\tau) < E(x_\tau), \tag{3.3}$$

which contradicts the fact that $E_\tau(x) = E(x_\tau)$ and thus x_τ must have already been on the boundary. □

Using the previous lemma, we can now show that for C^1-perturbations of convex functions, we can replace the lim sup in Lemma 2.8 by a normal limit.

Lemma 3.5. *Let $E : \mathcal{X} \to (-\infty, +\infty]$ admit a decomposition $E = E^c + E^d$, into a proper, lower semicontinuous convex function E^c and a C^1-function E^d, then for all $x \in \operatorname{dom}(E)$, we have*

$$|\partial E|(x) = \lim_{\tau\to 0^+} \frac{E(x) - E_\tau(x)}{\tau}. \tag{3.4}$$

Proof. Step 1: The convex case.

We first assume that E is convex. We choose τ small enough such that by Lemma 3.4, we obtain a sequence $\{x_n\}_n$ with $\|x - x_n\| = \tau$ and $\lim_{n\to\infty} E(x_n) = E_\tau(x)$. For each $n \in \mathbb{N}$, we consider the line

$$c_n(t) := t\,x_n + (1-t)\,x$$

evaluated at $\tilde{t} = \tilde{\tau}/\tau$ for some $0 < \tilde{\tau} < \tau$, which yields $\|x - c_n(\tilde{t})\| = \tilde{\tau}/\tau\ \|x - x_n\| = \tilde{\tau}$. Due to convexity, we obtain

$$E(c_n(\tilde{t})) \le \tilde{t}\,E(x_n) + \left(1-\tilde{t}\right)\,E(x) \qquad \Rightarrow \qquad E(x) - E(c_n(\tilde{t})) \ge \tilde{t}\ \left(E(x) - E(x_n)\right).$$

Using the fact that $E_{\tilde{\tau}}(x) \le E(c_n(\tilde{t}))$ and dividing by $\tilde{\tau}$ in the above inequality yields

$$\frac{E(x) - E_{\tilde{\tau}}(x)}{\tilde{\tau}} \ge \frac{E(x) - E(c_n(\tilde{t}))}{\tilde{\tau}} \ge \frac{E(x) - E(x_n)}{\tau}.$$

Considering the limit $n \to \infty$, we obtain the following inequality,

$$\frac{E(x) - E_{\tilde{\tau}}(x)}{\tilde{\tau}} \ge \limsup_{n\to\infty} \frac{E(x) - E(c_n(\tilde{t}))}{\tilde{\tau}} \ge \lim_{n\to\infty} \frac{E(x) - E(x_n)}{\tau} = \frac{E(x) - E_\tau(x)}{\tau}.$$

This shows that $\tau \mapsto Q(\tau) := \frac{E(x) - E_\tau(x)}{\tau}$ is decreasing in τ, and therefore, for a null sequence $\tau_n \to 0$, $Q(\tau_n)$ is an increasing sequence. The monotone convergence theorem together with Lemma 2.8 shows (3.4).

Step 2: Extension to C^1-perturbations.

We now assume that E is a C^1-perturbation of a convex function. By the definition of differentiability, we can write

$$E(z) = \underbrace{E^c(z) + E^d(x) - \langle DE^d(x), x - z\rangle}_{=:F(x)} + R(x, x-z),$$

https://doi.org/10.1017/S0956792525100120 Published online by Cambridge University Press

with $R(x, x-z) \in o(|x-z|)$ for every $z \in \operatorname{dom}(E)$. We observe that F is again a convex function. Let $\epsilon > 0$, then we denote by $x^E_{\tau,\epsilon}, x^F_{\tau,\epsilon} \in \overline{B_\tau}(x)$ the quasi-minimizers that fulfil

$$E(x^E_{\tau,\epsilon}) - E_\tau(x) \le \tau\epsilon \quad \text{and} \quad F(x^F_{\tau,\epsilon}) - F_\tau(x) \le \tau\epsilon \quad \text{respectively.}$$

We use the estimate

$$\begin{aligned} E_\tau(x) - F_\tau(x) &\le E_\tau(x) - F(x^F_{\tau,\epsilon}) + \tau\epsilon \\ &= \underbrace{E_\tau(x) - E(x^F_{\tau,\epsilon})}_{\le 0} + R(x, x - x^F_{\tau,\epsilon}) + \tau\epsilon \le |R(x, x - x^F_{\tau,\epsilon})| + \tau\epsilon \end{aligned}$$

and analogously

$$\begin{aligned} F_\tau(x) - E_\tau(x) &\le F_\tau(x) - E(x^E_{\tau,\epsilon}) + \tau\epsilon \\ &= \underbrace{F_\tau(x) - F(x^E_{\tau,\epsilon})}_{\le 0} - R(x, x - x^E_{\tau,\epsilon}) + \tau\epsilon \le |R(x, x - x^E_{\tau,\epsilon})| + \tau\epsilon, \end{aligned}$$

to obtain

$$|E_\tau(x) - F_\tau(x)| \le \max\left\{|R(x, x - x^E_{\tau,\epsilon})|, |R(x, x - x^F_{\tau,\epsilon})|\right\} + \tau\epsilon. \tag{3.5}$$

Using that $E(x) = F(x)$ and dividing by τ in (3.5) yields the inequality

$$\left|\frac{E(x) - E_\tau(x)}{\tau} - \frac{F(x) - F_\tau(x)}{\tau}\right| \le \underbrace{\frac{\max\{|R(x, x - x^E_\tau(\epsilon))|, |R(x, x - x^F_\tau(\epsilon))|\}}{\tau}}_{:= r(\tau)} + \epsilon. \tag{3.6}$$

Since $\left|x - x^E_\tau(\epsilon)\right| = \left|x - x^F_\tau(\epsilon)\right| \le \tau$, it holds $\lim_{\tau\to 0} r(\tau) = 0$. Taking the lim sup of (3.6) and sending ϵ to zero then yields,

$$\lim_{\tau\to 0^+} \left|\frac{E(x) - E_\tau(x)}{\tau} - \frac{F(x) - F_\tau(x)}{\tau}\right| = 0.$$

Therefore, the limit in (3.4) exists,

$$\begin{aligned} \lim_{\tau\to 0^+} \frac{E(x) - E_\tau(x)}{\tau} &= \lim_{\tau\to 0^+} \frac{E(x) - E_\tau(x)}{\tau} - \frac{F(x) - F_\tau(x)}{\tau} + \frac{F(x) - F_\tau(x)}{\tau} \\ &= \lim_{\tau\to 0^+} \frac{F(x) - F_\tau(x)}{\tau} = |\partial F|(x), \end{aligned}$$

where in the last step, we used that F is convex together with **Step 1**. □

3.2. Differential inclusions

Similar to [2, Proposition 1.4.1] for finite p, we now give a characterization of ∞-curves of maximal slope via differential inclusions, whenever the slope of the energy E can be written as

$$|\partial E|(x) = \min\{\|\xi\|_* : \xi \in \partial E(x)\} = \|\partial^\circ E(x)\|_* \quad \forall x \in \mathcal{X}. \tag{3.7}$$

By Proposition 3.1, this is, e.g., the case for C^1-perturbations. Let us start by defining a degenerate duality mapping $\mathcal{J}_\infty : \mathcal{X} \to 2^{\mathcal{X}^*}$,

$$\mathcal{J}_\infty(x) := \begin{cases} \{\xi \in \mathcal{X}^* : \langle \xi, u\rangle = \|\xi\|_*\} & \text{if } \|x\| = 1, \\ \{0\} & \text{if } \|x\| < 1, \\ \emptyset & \text{if } \|x\| > 1, \end{cases}$$

as the limit case of the classical p-duality mapping [93, Definition 2.27]

$$\mathcal{J}_p(x) := \left\{\zeta \in \mathcal{X}^* : \ \langle \zeta, u\rangle = \|x\|\|\zeta\|_*, \ \|\zeta\|_* = \|x\|^{p-1}\right\}.$$

This definition allows us to extend the classical Asplund theorem [93, Theorem 2.28] to the limit case.

https://doi.org/10.1017/S0956792525100120 Published online by Cambridge University Press

Theorem 3.6 (Asplund theorem for $p = \infty$). *The following identity holds true,*

$$\mathcal{J}_\infty = \partial \chi_{\overline{B_1}}.$$

Proof. For $x \in \mathcal{X}$ with $\|x\| \neq 1$, the equality holds trivially. Therefore, we consider $\|x\| = 1$.

Step 1: $\mathcal{J}_\infty(x) \subset \partial \chi_{\overline{B_1}}(x)$.

Let $\xi \in \mathcal{J}_\infty(x)$, which means $\langle \xi, x\rangle = \|\xi\|_*$, and consider an arbitrary $z \in \mathcal{X}$. If $\|z\| \leq 1$, we obtain

$$\chi_{\overline{B_1}}(z) - \langle \xi, z - x\rangle = -\langle \xi, z\rangle + \|\xi\|_* \geq \|\xi\|_* (1 - \|z\|) \geq 0 = \chi_{\overline{B_1}}(x),$$

while for $\|z\| > 1$, the inequality holds trivially, thus we have $\xi \in \partial \chi_{\overline{B_1}}(x)$.

Step 2: $\mathcal{J}_\infty(x) \supset \partial \chi_{\overline{B_1}}(x)$.

Let $\xi \in \partial \chi_{\overline{B_1}}(x)$, then for all $z \in \overline{B_1}$ we get

$$\underbrace{\partial \chi_{\overline{B_1}}(z)}_{=0} \geq \underbrace{\partial \chi_{\overline{B_1}}(x)}_{=0} + \langle \xi, z - x\rangle \iff \langle \xi, z\rangle \leq \langle \xi, x\rangle \leq \|\xi\|_*.$$

Taking the supremum over all $z \in \overline{B_1}$ yields the equality $\langle \xi, x\rangle = \|\xi\|_*$ and thus $\xi \in \mathcal{J}_\infty(x)$. □

Next, we are interested in the behaviour of the energy along curves of maximal slope. We derive a more general chain rule for subdifferentiable energies that only requires differentiability along curves.

Lemma 3.7 (Chain rule). *Let $u : [0, T] \to \mathrm{dom}\,(E)$ be a curve, then at each point t where u and $E \circ u$ are differentiable and $\partial E(u(t)) \neq \emptyset$, we have*

$$\frac{d}{dt} E(u(t)) = \langle \xi, u'(t)\rangle \quad \forall \xi \in \partial E(u(t)). \tag{3.8}$$

Proof. Let $t \in [0, T]$ be a point, where u and $E \circ u$ are differentiable, then we use the definition of the derivative to obtain

$$\frac{\mathrm{d}}{\mathrm{d}t} E(u(t)) - \langle \xi, u'(t)\rangle = \lim_{n\to\infty} \frac{E(u(t + h_n)) - E(u(t)) - \langle \xi, u(t + h_n) - u(t)\rangle}{h_n} =: (\spadesuit),$$

where $\{h_n\}_n$ is a null sequence. We first consider only positive null sequences $h_n > 0$, where we want to ensure that $u(t + h_n) \neq u(t)$. If such a sequence does not exist, we infer that

$$\frac{\mathrm{d}}{\mathrm{d}t} E(u(t)) = 0 = u'(t)$$

and (3.8) holds. Now assuming that there exists a sequence with $u(t + h_n) \neq u(t)$ we continue,

$$(\spadesuit) = \lim_{n\to\infty} \underbrace{\frac{E(u(t + h_n)) - E(u(t)) - \langle \xi, u(t + h_n) - u(t)\rangle}{\|u(t + h_n) - u(t)\|}}_{=: l_n} \cdot \underbrace{\frac{\|u(t + h_n) - u(t)\|}{h_n}}_{r_n}.$$

Note that $r_n \geq 0$ for all $n \in \mathbb{N}$ since we only allowed positive null sequences. Since u is differentiable and in particular continuous at t and since $\xi \in \partial E(u(t))$ (3.1) yields

$$\liminf_{n\to\infty} l_n \geq 0,$$

i.e., for every null sequence $\{h_n\}_n$, we can find a subsequence $\{h_n\}_n$ such that l_n either converges to some limit $l \geq 0$ or diverges to $+\infty$. In the convergent case, we obtain

$$(\spadesuit) = l \cdot \|u'(t)\| \geq 0.$$

In the divergent case, we also have $(\spadesuit) \geq 0$, since we can find a n_0 such that l_n is non-negative for all $n \geq N$. Using the same arguments as above, but only allowing negative null sequences $h_n < 0$, we instead obtain $(\spadesuit) \leq 0$. This finally yields

$$\frac{\mathrm{d}}{\mathrm{d}t} E(u(t)) - \langle \xi, u'(t)\rangle = 0.$$

□

https://doi.org/10.1017/S0956792525100120 Published online by Cambridge University Press

The chain rule from Lemma 3.7, together with the characterization of the metric slope (3.7), enables us to show that energy dissipation inequality (InfFlow) can be equivalently characterized via a differential inclusion.

Theorem 3.8. *Let $E{:}\mathcal{X} \to (-\infty, +\infty]$ satisfy (3.7) and $u : [0, 1] \to \mathcal{X}$ be an a.e. differentiable Lipschitz curve. Let further $E \circ u$ be a.e. equal to a non-increasing function ψ, then the following are equivalent:*

(i) $|u'|(t) \le 1$ *and* $\psi'(t) \le -|\partial E|(u(t))$ *for a.e.* $t \in [0, T]$,

(ii) $\mathcal{J}_\infty(u'(t)) \supset -\partial^\circ E(u(t)) \neq \emptyset$ *for a.e.* $t \in [0, 1]$,

(iii) $u'(t) \in \partial \|\cdot\|_*(-\xi) \cap \mathcal{X} = -\arg\max_{x \in \overline{B_1}} \langle \xi, x\rangle$ *for all* $\xi \in \partial^\circ E(u(t)) \neq \emptyset$, *and a.e.* $t \in (0, T)$.

Proof. Step 1: (i) $\Leftrightarrow$ (iii).

Since ψ is a monotone function, it is differentiable a.e., and thus we can find a Lebesgue null set $N \subset [0, T]$, such that u and ψ are differentiable and $E(u(t)) = \psi(t)$ for every $t \in [0, T] \setminus N$. Using Lemma 3.7 and (3.7) for $t \in [0, 1] \setminus N$ we obtain,

$$\left.\begin{matrix}\psi'(t) \le -|\partial E|(u(t)) \\ |u'|(t) \le 1\end{matrix}\right\} \Leftrightarrow \begin{cases}\langle \xi, u'(t)\rangle = \psi'(t) \le -\|\xi\|_* & \text{for all} \quad \xi \in \partial^\circ E(u(t)) \\ |u'|(t) \le 1\end{cases}$$
$$\Leftrightarrow \langle \xi, u'(t)\rangle \le -\|\xi\|_* - \chi_{\overline{B_1}}(u'(t)) \quad \text{for all} \quad \xi \in \partial^\circ E(u(t)).$$

For each $\xi \in \partial^\circ E(u(t))$, the last statement is Item 2 with $f = \chi_{\overline{B_1}}$ and $f^* = \|\cdot\|_*$, which is equivalent to Item 1, i.e.,

$$\langle \xi, u'(t)\rangle \le -\|\xi\|_* - \chi_{\overline{B_1}}(u'(t)) \quad \Leftrightarrow \quad u'(t) \in \partial \|\cdot\|_*(-\xi), \tag{3.9}$$

and thus we have shown (i) $\Leftrightarrow$ (iii). The set identity in (*iii*),

$$\partial \|\cdot\|_*(-\xi) \cap \mathcal{X} = -\arg\max_{x \in \overline{B_1}} \langle \xi, x\rangle,$$

follows from Corollary A.5.

Step 2: (i) $\Leftrightarrow$ (ii).

Using the equivalence of Item 2 and Item 1 in (3.9), we also obtain that for a.e. $t \in [0, T]$ and all $\xi \in \partial E^\circ(u(t))$

$$(i) \quad \Leftrightarrow \quad -\xi \in \partial \chi_{\overline{B_1}}(u'(t)).$$

From Asplund's theorem (Theorem 3.6), we have that

$$-\xi \in \partial \chi_{\overline{B_1}}(u'(t)) \Longleftrightarrow -\xi \in \mathcal{J}_\infty(u'(t))$$

which thus implies (*i*) $\Leftrightarrow$ (*ii*). □

3.3. Semi-implicit time stepping

The minimizing movement scheme in (MinMove) can be considered as an implicit time stepping scheme, which is often computationally intractable in practice. Therefore, one may want to instead employ an explicit scheme. In this regard, we are interested in minimizing movement schemes of the semi-implicit energy, which in many cases can be computed explicitly. We consider a Banach space $\mathcal{X}$ that fulfils Assumptions 1.a and 1.b and a C^1-perturbation of a convex function $E = E^{\mathrm{d}} + E^{\mathrm{c}}$, fulfilling assumptions Assumptions 2.a and 2.b. Furthermore, we assume:

Assumption 3.a (Lipschitz continuous differentiability). *The differentiable part E^{d} has a Lipschitz continuous first derivative.*

https://doi.org/10.1017/S0956792525100120 Published online by Cambridge University Press

We can linearize the differentiable part of the energy around a point z and define the linearized energy by

$$E^{\mathrm{sl}}(x;z) := E^{\mathrm{d}}(z) + \langle DE^{\mathrm{d}}(z), x - z\rangle + E^{\mathrm{c}}(x).$$

To ensure that the minimizers in (3.10) are obtained, we assume:

Assumption 3.b (Lower semi-continuity). *The semi linearization $x \mapsto E^{\mathrm{sl}}(x;z)$ is σ-lower semicontinuous for every $z \in \mathcal{X}$.*

Remark 3.9. In reflexive spaces, this is a very mild assumption, as the σ-topology is often chosen to be the weak topology. In this case, we only need an assumption on the convex part E^{c}, namely lower semicontinuity, which together with convexity implies weak lower semicontinuity. The linearized part $x \mapsto E^{\mathrm{d}}(z) + \langle DE^{\mathrm{d}}(z), x - z\rangle$ is even weakly continuous and therefore, we do not need additional assumptions.

Definition 3.10 (Semi-implicit Scheme). *For $x^0 \in \mathrm{dom}\,(E^{\mathrm{c}})$, we define the semi-implicit scheme as*

$$x^{k+1}_{\mathrm{si},\tau} \in \operatorname*{arg\,min}_{x \in \overline{B_\tau}(x^k_{\mathrm{si},\tau})} E^{\mathrm{sl}}\big(x; x^k_{\mathrm{si},\tau}\big), \tag{3.10}$$

for $k \in \mathbb{N}$ with $x^0_{\mathrm{si},\tau} = x^0$. We define the step function $\bar{x}_{\mathrm{si},\tau}$ by

$$\bar{x}_{\mathrm{si},\tau}(0) = x^0, \quad \bar{x}_{\mathrm{si},\tau}(t) = x^k_{\mathrm{si},\tau} \qquad \text{if} \qquad t \in \big(t^{k-1}_\tau, t^k_\tau\big], k \geq 1.$$

Furthermore, we define

$$|x'_{\mathrm{si},\tau}|(t) := \frac{d\big(x^k_{\mathrm{si},\tau}, x^{k-1}_{\mathrm{si},\tau}\big)}{t^k_\tau - t^{k-1}_\tau} \text{ if } t \in \big(t^{k-1}_\tau, t^k_\tau\big)$$

as the metric derivative of the corresponding piecewise affine linear interpolation.

Remark 3.11. The above scheme can also be recovered via the theory of doubly non-linear equations developed in [69]. Namely, by considering the state-dependent dissipation potential

$$\Psi_z(v) := \chi_{\overline{B_1}}(v) + E^{\mathrm{d}}(z) + \langle DE^{\mathrm{d}}(z), v\rangle$$

the minimizing movement scheme defined in [69, Eq. (4.9)] is given as

$$x^{k+1}_{\mathrm{si},\tau} \in \operatorname*{arg\,min}_{x \in \mathcal{X}} \left\{ \tau \Psi_{x^k_{\mathrm{si},\tau}}\left(\frac{x - x^k_{\mathrm{si},\tau}}{\tau}\right) + E^{\mathrm{c}}(x) \right\}$$

which exactly recovers the scheme defined in Definition 3.10. The authors show convergence of this scheme towards solution of the equation

$$\partial \Psi_{u(t)}(u'(t)) + \partial E^{\mathrm{c}}(u(t)) \ni 0$$

which corresponds to the inclusion derived in Theorem 3.8. However, we cannot directly apply the results of [69] since the choice of dissipation potential as above violates condition $(2.\Psi_1)$, since $\mathrm{dom}\,(\Psi) \neq \mathcal{X}$, $(2.\Psi_2)$ since in general $\Psi_u(0) \neq 0$ and the growth condition on the Fenchel conjugate $\Psi^*_z(\xi) = \big\|\xi - DE^{\mathrm{d}}(z)\big\|_* - E^{\mathrm{d}}(z)$ is not fulfilled and also $(2.\Psi_3)$. In fact, a more detailed study on how these assumptions could be relaxed would be very interesting, which we, however, leave for future work.

An important special case of the above scheme is a reflexive Banach space $\mathcal{X}$ together with a C^1 energy E, i.e., we can choose $E^{\mathrm{c}} = 0$. In this case, the scheme is fully explicit, as the following lemma shows.

Lemma 3.12. *If the Banach space $\mathcal{X}$ is reflexive and $E \in C^1(\mathcal{X})$, then we can explicitly compute the iterates in Definition 3.10 as*

$$x^{k+1}_{\mathrm{si},\tau} \in x^k_{\mathrm{si},\tau} - \tau\, \partial \|\cdot\|_*(DE(x^k_{\mathrm{si},\tau})).$$

https://doi.org/10.1017/S0956792525100120 Published online by Cambridge University Press

Proof. We compute

$$\begin{aligned} x^{k+1}_{\mathrm{si},\tau} &\in \operatorname*{arg\,min}_{x\,:\,\|x-x^k_{\mathrm{si},\tau}\|\le\tau} E(x^k_{\mathrm{si},\tau}) + \langle DE(x^k_{\mathrm{si},\tau}), x - x^k_{\mathrm{si},\tau}\rangle \\ &= \operatorname*{arg\,min}_{x\,:\,\|x-x^k_{\mathrm{si},\tau}\|\le\tau} \langle DE(x^k_{\mathrm{si},\tau}), x\rangle \\ &= -\operatorname*{arg\,max}_{x\,:\,\|x-x^k_{\mathrm{si},\tau}\|\le\tau} \langle DE(x^k_{\mathrm{si},\tau}), x\rangle \\ &= x^k_{\mathrm{si},\tau} - \tau \operatorname*{arg\,max}_{x\in\overline{B_1}} \langle DE(x^k_{\mathrm{si},\tau}), x\rangle \\ &= x^k_{\mathrm{si},\tau} - \tau\, \partial\|\cdot\|_*(DE(x^k_{\mathrm{si},\tau})), \end{aligned}$$

where for the last identity, we used A.5. □

In section 5, we consider a case where $E^{\mathrm{c}} \neq 0$, but the scheme can still be computed explicitly. In fact, the iteration then coincides with (IFGSM), which ultimately yields the desired convergence result.

It is easy to see that the metric slope of E and its semi linearization $E^{\mathrm{sl}}(\,\cdot\,;z)$ coincide in the point of linearization z, i.e. $|\partial E|(z) = |\partial E^{\mathrm{sl}}(\,\cdot\,;z)|(z)$. The next lemma estimates the difference of their slope when u is not the point of linearization.

Lemma 3.13. *Let E be a C^1-perturbation of a convex function satisfying Assumption 3.a, then for each $z, x \in \mathcal{X}$, we have the following estimate*

$$\left||\partial E|(x) - |\partial E^{\mathrm{sl}}(\,\cdot\,;z)|(x)\right| \le \mathrm{Lip}(DE^{\mathrm{d}})\|z - x\|. \tag{3.11}$$

Proof. Let $z, x \in \mathcal{X}$, from Item 1 we know

$$\begin{aligned} \partial E(x) &= \partial E^{\mathrm{c}}(x) + DE^{\mathrm{d}}(x), \\ \partial E^{\mathrm{sl}}(x;z) &= \partial E^{\mathrm{c}}(x) + DE^{\mathrm{d}}(z), \end{aligned}$$

and then Item 3 implies that there exists $\xi_1, \xi_2 \in \partial E^{\mathrm{c}}(x)$ such that

$$\begin{aligned} |\partial E|(x) &= \min\left\{\left\|\xi + DE^{\mathrm{d}}(x)\right\|_* : \xi \in \partial E^{\mathrm{c}}(x)\right\} = \left\|\xi_1 + DE^{\mathrm{d}}(x)\right\|_*, \\ \left|\partial E^{\mathrm{sl}}(\,\cdot\,;z)\right|(x) &= \min\left\{\left\|\xi + DE^{\mathrm{d}}(z)\right\|_* : \xi \in \partial E^{\mathrm{c}}(x)\right\} = \left\|\xi_2 + DE^{\mathrm{d}}(z)\right\|_*. \end{aligned}$$

We can then estimate

$$\begin{aligned} |\partial E|(x) &\le \left\|DE^{\mathrm{d}}(x) + \xi_2\right\|_* \le \|DE^{\mathrm{d}}(x) - DE^{\mathrm{d}}(z)\|_* + \|DE^{\mathrm{d}}(z) + \xi_2\|_* \\ &\le \mathrm{Lip}(DE^{\mathrm{d}})\|x - z\| + \left|\partial E^{\mathrm{sl}}(\,\cdot\,;z)\right|(x), \end{aligned}$$

and therefore

$$|\partial E|(x) - \left|\partial E^{\mathrm{sl}}(\,\cdot\,;z)\right|(x) \le \mathrm{Lip}(DE^{\mathrm{d}})\|x - z\|.$$

Analogously, we estimate

$$\begin{aligned} \left|\partial E^{\mathrm{sl}}(\cdot, z)\right|(x) &\le \|DE^{\mathrm{d}}(z) + \xi_1\| \le \|DE^{\mathrm{d}}(z) - DE^{\mathrm{d}}(x)\| + \|DE(x) + \xi_1\| \\ &\le \mathrm{Lip}(DE^{\mathrm{d}})\|x - z\| + |\partial E|(x). \end{aligned}$$

and therefore

$$\left|\partial E^{\mathrm{sl}}(\,\cdot\,;z)\right|(x) - |\partial E|(x) \le \mathrm{Lip}(DE^{\mathrm{d}})\|x - z\|.$$

This concludes the proof. □

In the following, we want to define a variational interpolation similar to Definition 2.10. Therefore, we consider

$$E^{\mathrm{sl}}_\tau(x;z) = \min_{\tilde{x}\in\overline{B_\tau}(x)} E^{\mathrm{sl}}(\tilde{x};z).$$

For better readability, if z and x coincide above, we set

$$E^{\mathrm{sl}}_\tau(x) := E^{\mathrm{sl}}_\tau(x;x) = \min_{\tilde{x}\in\overline{B_\tau}(x)} E^{\mathrm{sl}}(\tilde{x};x).$$

https://doi.org/10.1017/S0956792525100120 Published online by Cambridge University Press

Definition 3.14 (Semi-implicit variational interpolation). *We denote by* $\tilde{x}_{\mathrm{si},\tau} : [0,T] \to \mathcal{X}$ *any interpolation of the discrete values satisfying*

$$\tilde{x}_{\mathrm{si},\tau}(t) \in \arg\min_x \left\{ E^{\mathrm{sl}}\left(x; x^{k-1}_{\mathrm{si},\tau}\right) : d\left(x, x^{k-1}_{\mathrm{si},\tau}\right) \le t - t^{k-1}_\tau \right\}$$

if $t \in (t^{k-1}_\tau, t^k_\tau]$ *and* $k \ge 1$. *Furthermore, we define*

$$\mathcal{D}_\tau(t) := \frac{d}{dt} E^{\mathrm{sl}}_{(t-t^{k-1}_\tau)}(x^{k-1}_{\mathrm{si},\tau}). \tag{3.12}$$

The following Lemma shows that the variational interpolation of the semi-implicit minimizing movement scheme satisfies the same properties, (2.8) and (2.9), as the *De Giorgi variational interpolation*, up to an error in $\mathcal{O}(\tau)$.

Lemma 3.15. *We have that*

$$\mathcal{D}_\tau(t) = \frac{d}{dt} E^{\mathrm{sl}}_{\left(t-t^{k-1}_\tau\right)}\left(x^{k-1}_{\mathrm{si},\tau}\right) \le -|\partial E^{\mathrm{sl}}\left(\,\cdot\,; x^{k-1}_{\mathrm{si},\tau}\right)|(\tilde{x}_{\mathrm{si},\tau}(t)) = -|\partial E|(\tilde{x}_{\mathrm{si},\tau}(t)) + \mathcal{O}(\tau) \text{ if } t \in (t^{k-1}_\tau, t^k_\tau] \tag{3.13}$$

and

$$E(\tilde{x}_{\mathrm{si},\tau}(s)) + \int_s^t D_\tau(r)\, dr \ge E(\tilde{x}_{\mathrm{si},\tau}(t)) + \mathcal{O}(\tau) \quad \forall\, 0 \le s \le t \le T. \tag{3.14}$$

Proof. For (3.13), we apply Lemma 2.9 to the mapping $x \mapsto E^{\mathrm{sl}}(x; x^{k-1}_{\mathrm{si},\tau})$ to obtain

$$\frac{\mathrm{d}}{\mathrm{d}t} E^{\mathrm{sl}}_{\left(t-t^{k-1}_\tau\right)}\left(x; x^{k-1}_{\mathrm{si},\tau}\right) \le |\partial E^{\mathrm{sl}}\left(\cdot, x^{k-1}_{\mathrm{si},\tau}\right)|\left(x_{\min,t-t^{k-1}_\tau}\right),$$

where $x_{\min,t-t^{k-1}_\tau} \in \arg\min_{\tilde{x}} \{E^{\mathrm{sl}}(\tilde{x}; x^{k-1}_{\mathrm{si},\tau}) : \tilde{x} \in \overline{B_\tau}(x)\}$. Choosing $v = x^{k-1}_{\mathrm{si},\tau}$ then yields

$$\frac{\mathrm{d}}{\mathrm{d}t} E^{\mathrm{sl}}_{(t-t^{k-1}_\tau)}(x^{k-1}_{\mathrm{si},\tau}) \le -|\partial E^{\mathrm{sl}}(\,\cdot\,; x^{k-1}_{\mathrm{si},\tau})|(\tilde{x}_{\mathrm{si},\tau}(t)).$$

The last equality of (3.13) follows by Lemma 3.13. To show (3.14), we again use Lemma 2.9 and get

$$E^{\mathrm{sl}}\left(\tilde{x}_{\mathrm{si},\tau}(s); x^k_{\mathrm{si},\tau}\right) + \int_s^t \mathcal{D}_\tau(r)\,\mathrm{d}r \ge E^{\mathrm{sl}}\left(\tilde{x}_{\mathrm{si},\tau}(t), x^k_{\mathrm{si},\tau}\right) \quad \text{for all} \quad t^k_\tau \le s \le t \le t^{k+1}_\tau.$$

Due to Theorem C.1

$$\begin{aligned}
\left|E^{\mathrm{sl}}\left(\tilde{x}_{\mathrm{si},\tau}(s); x^k_{\mathrm{si},\tau}\right) - E(\tilde{x}_{\mathrm{si},\tau}(s))\right| &= \left|E^{\mathrm{d}}\left(x^k_{\mathrm{si},\tau}\right) + \langle DE^{\mathrm{d}}(x^k_{\mathrm{si},\tau}), \tilde{x}_{\mathrm{si},\tau}(s) - x^k_{\mathrm{si},\tau}\rangle - E^{\mathrm{d}}(\tilde{x}_{\mathrm{si},\tau}(s))\right| \\
&\le \left| \int_0^1 \left\langle DE^{\mathrm{d}}\left(x^k_{\mathrm{si},\tau} + r(\tilde{x}_{\mathrm{si},\tau}(s) - x^k_{\mathrm{si},\tau})\right), x^k_{\mathrm{si},\tau} - \tilde{x}_{\mathrm{si},\tau}(s)\right\rangle \right. \\
&\qquad \left. - \langle DE^{\mathrm{d}}(x^k_{\mathrm{si},\tau}), \tilde{x}_{\mathrm{si},\tau}(t) - x^k_{\mathrm{si},\tau}\rangle\, \mathrm{d}r \right| \\
&\le \int_0^1 r\mathrm{Lip}(DE^{\mathrm{d}})\|\tilde{x}_{\mathrm{si},\tau}(s) - x^k_{\mathrm{si},\tau}\|^2\, \mathrm{d}r \\
&\le \frac{1}{2}\mathrm{Lip}(DE^{\mathrm{d}})\|\tilde{x}_{\mathrm{si},\tau}(s) - x^k_{\mathrm{si},\tau}\|^2 \le \frac{1}{2}\mathrm{Lip}(DE^{\mathrm{d}})\tau^2
\end{aligned}$$

and analogously $|E^{\mathrm{sl}}(\tilde{x}_{\mathrm{si},\tau}(t), x^k_{\mathrm{si},\tau}) - E(\tilde{x}_{\mathrm{si},\tau}(t))| \le \frac{1}{2}\mathrm{Lip}(DE^{\mathrm{d}})\tau^2$. Therefore, for all $t^k_\tau \le s \le t \le t^{k+1}_\tau$, we have that

$$\begin{aligned}
E(\tilde{x}_{\mathrm{si},\tau}(s)) + \int_s^t \mathcal{D}_\tau(r)\,\mathrm{d}r &\ge E^{\mathrm{sl}}(\tilde{x}_{\mathrm{si},\tau}(s); x^k_\tau) + \int_s^t \mathcal{D}_\tau(r)\,\mathrm{d}r - \frac{1}{2}\mathrm{Lip}(DE^{\mathrm{d}})\tau^2 \\
&\ge E^{\mathrm{sl}}(\tilde{x}_{\mathrm{si},\tau}(t); x^k_\tau) - \frac{1}{2}\mathrm{Lip}(DE^{\mathrm{d}})\tau^2 \\
&\ge E(\tilde{x}_{\mathrm{si},\tau}(t)) - \mathrm{Lip}(DE^{\mathrm{d}})\tau^2.
\end{aligned} \tag{3.15}$$

https://doi.org/10.1017/S0956792525100120 Published online by Cambridge University Press

Now for $s \in [t_\tau^m, t_\tau^{m+1}]$ and $t \in [t_\tau^k, t_\tau^{k+1}]$ with $m \le k$, we add up (3.15) to obtain

$$\begin{aligned} &E(\tilde{x}_{\mathrm{si},\tau}(s)) + \int_s^{t_\tau^{m+1}} \mathcal{D}_\tau(r)\,\mathrm{d}r + \sum_{i=m+1}^{k-1} \int_{t_\tau^i}^{t_\tau^{i+1}} \mathcal{D}_\tau(r)\,\mathrm{d}r + \int_{t_\tau^k}^{t} \mathcal{D}_\tau(r)\,\mathrm{d}r \\ &\quad \ge E(\tilde{x}_{\mathrm{si},\tau}(t)) - \sum_{i=m}^{k} \mathrm{Lip}(DE^{\mathrm{d}})\tau^2 \\ &\quad = E(\tilde{x}_{\mathrm{si},\tau}(t)) - (k-m)\mathrm{Lip}(DE^{\mathrm{d}})\tau^2 \\ &\quad \ge E(\tilde{x}_{\mathrm{si},\tau}(t)) - T\mathrm{Lip}(DE^{\mathrm{d}})\tau \end{aligned}$$

such that we finally obtain (3.14). □

As an immediate consequence of Lemma 3.15, we can replace the minimizing movement scheme in the proof of Theorem 2.11 by the semi-implicit scheme, as the error terms are of order $\mathcal{O}(\tau)$ and vanish during the limiting process $\tau \to 0$. Then $\bar{x}_{\mathrm{si},\tau}$ σ-converges up to a subsequence to a ∞-curve of maximal slope.

Theorem 3.16. *Let E be a C^1-perturbation of a convex function. Under Assumptions 1.a to 3.b, there exists a ∞-curve of maximal slope $u(t)$, with respect to the energy E and its upper gradient $|\partial E|$, and a subsequence of $\tau_n = T/n$ such that*

$$\bar{x}_{\mathrm{si},\tau_n}(t) \xrightarrow{\sigma} u(t) \text{ as } n \to \infty \quad \forall t \in [0,T].$$

Proof. We simply replace the minimizing movement scheme in Definition 2.5 and De Giorgis variational interpolation (see Definition 2.10) by the semi-implicit scheme in Definition 3.10 and its corresponding variational interpolation of Lemma 3.15. Proceeding similarly as in the proof of Theorem 2.11, we use Proposition B.1 to show

$$\bar{x}_{\mathrm{si},\tau_n}(t) \xrightarrow{\sigma} u(t) \text{ as } n \to \infty \quad \forall t \in [0,T]$$

for a subsequence τ_n, where u is a 1-Lipschitz curve with $u(0) = x^0$. Then the same holds true for $\tilde{x}_{\mathrm{si},\tau_n}(t)$.

Taking for τ_n the limes inferior for $n \to \infty$ of (3.14) and using Assumption 2.a, Assumption 2.b and (3.13) we again obtain

$$E(u(0)) \ge E(u(t)) + \int_0^t |\partial E|(u(r))\,\mathrm{d}r \quad \text{for all } t \in [0,T].$$

Since on the other hand $|\partial E|$ is a strong upper gradient, equality in the above equation must hold. □

Remark 3.17. Let τ_n be any sequence such that $\tau_n \to 0$. If the ∞-curve of maximal slope u is unique, we can apply Theorem 3.16 to every subsequence of τ_n and find a further subsequence $\tilde{\tau}_n$ such that

$$\bar{x}_{\mathrm{si},\tilde{\tau}_n}(t) \xrightarrow{\sigma} u(t) \text{ as } n \to \infty \quad \forall t \in [0,T].$$

This implies that already for τ_n

$$\bar{x}_{\mathrm{si},\tau_n}(t) \xrightarrow{\sigma} u(t) \text{ as } n \to \infty \quad \forall t \in [0,T].$$

and the semi-implicit scheme converges.

4. Wasserstein infinity flows

The previous sections consider a "single particle", $x \in \mathcal{X}$, trying to minimize an energy $\mathcal{E}$, by following an ∞-curve of maximal slope. This single particle may be drawn from a probability distribution $\mu_0 \in \mathcal{P}(\mathcal{X})$, which over time also minimizes an energy $\mathcal{E}$ defined on the space of probabilities. In this section, we choose the underlying metric space $\mathcal{S}$ to be the space of Borel probability measures with bounded support $\mathcal{P}_\infty(\mathcal{X})$, and equip it with the ∞-Wasserstein distance. We show that for potential energies, ∞-curves of maximal slope can be expressed via a probability measures η on the space $C(0,T;\mathcal{X})$, which

https://doi.org/10.1017/S0956792525100120 Published online by Cambridge University Press

is concentrated on ∞-curves of maximal slope on the underlying Banach space $\mathcal{X}$. From η, we can then derive a corresponding continuity equation which those ∞-curves of maximal slope have to fulfil.

This concept is commonly referred to as the "superposition principle", where our approach directly follows the setup of [2, 62, 63]. We refer to [99] for an overview of different works in this direction, as well as results that hold true in a much more general setting.

4.1. Preliminaries on Wasserstein spaces

We give a brief introduction to the basic properties of Wasserstein spaces. For more details, we refer to [2, 44, 104]. In the following, $(\mathcal{X}, \|\cdot\|)$ is a separable Banach space. We denote by $\mathcal{P}(\mathcal{X})$ the space of Borel probability measures on $\mathcal{X}$. For $1 \le p < \infty$, $\mathcal{P}_p(\mathcal{X}) \subset \mathcal{P}(\mathcal{X})$ is the subset of measures with finite p-momentum, while $\mathcal{P}_\infty(\mathcal{X}) \subset \mathcal{P}(\mathcal{X})$ is the subset of measures with bounded support. For $1 \le p < \infty$ and $\mu, \nu \in \mathcal{P}_p(\mathcal{X})$, we define the p-Wasserstein distance as

$$W_p^p(\mu, \nu) := \inf_{\gamma \in \Gamma(\mu,\nu)} \int \|x - z\|^p \mathrm{d}\gamma(x, z).$$

Here,

$$\Gamma(\mu, \nu) := \{\gamma \in \mathcal{P}(\mathcal{X} \times \mathcal{X}) : \pi^1_\# \gamma = \mu, \pi^2_\# \gamma = \nu\}, \tag{4.1}$$

is the set of admissible transport plans and $\pi^1(x, z) = x$, $\pi^2(x, z) = z$ denote the projection on the first and second component. For $\mu, \nu \in \mathcal{P}_\infty(\mathcal{X})$, the ∞-Wasserstein distance is given by

$$W_\infty(\mu, \nu) := \inf_{\gamma \in \Gamma(\mu,\nu)} \gamma - \operatorname{ess\,sup} \|x - z\|. \tag{4.2}$$

In both cases, the minimum of (4.1) and (4.2) is obtained (see, e.g., [2, 104] and [44, Proposition 1] for the case $p = \infty$) and $\Gamma_0(\mu, \nu)$ denotes the set of optimal transport plans where the minimum is reached.

Proposition 4.1 [44, Proposition 6.]. *For $p \in [1, \infty]$, $\mathcal{W}_p = (\mathcal{P}_p(\mathcal{X}), W_p)$, i.e., $\mathcal{P}_p(\mathcal{X})$ equipped with the p-Wasserstein distance, is a complete metric space. For $p < \infty$, $\mathcal{W}_p$ is separable.*

The following lemma shows that Wasserstein distances are ordered in such a way that they get stronger by increasing p, see [44, Proposition 3.].

Lemma 4.2 [44, Proposition 3.]. *For $1 \le p \le q \le \infty$ and $\mu, \nu \in \mathcal{P}(\mathcal{X})$*

$$W_p(\mu, \nu) \le W_q(\mu, \nu) \tag{4.3}$$

and in particular

$$W_\infty(\mu, \nu) = \sup_p W_p(\mu, \nu) = \lim_{p \to \infty} W_p(\mu, \nu).$$

Let now σ denote the narrow topology, namely, $\mu^n \xrightarrow{\sigma} \mu$ iff,

$$\int_{\mathcal{X}} \varphi \, \mathrm{d}\mu^n \to \int_{\mathcal{X}} \varphi \, \mathrm{d}\mu \quad \forall \varphi \in C_b(\mathcal{X}), \tag{4.4}$$

where $C_b(\mathcal{X})$ denotes the space of bounded and continuous functions on $\mathcal{X}$. The next lemma is helpful, when we are considering limits in (4.4) with φ being unbounded or only lower semicontinuous.

Lemma 4.3 [2, Lemma 5.1.7.]. *Let $(\mu^n)_{n \in \mathbb{N}}$ be a sequence in $\mathcal{P}(\mathcal{X})$ narrowly converging to $\mu \in \mathcal{P}(\mathcal{X})$. If $g : \mathcal{X} \to (-\infty, +\infty]$ is lower semicontinuous and its negative part $g^- = -\min\{g, 0\}$ is uniformly integrable w.r.t. the set $\{\mu^n\}_{n \in \mathbb{N}}$, then*

$$\liminf_{n \to \infty} \int_{\mathcal{X}} g(x) d\mu^n(x) \ge \int_{\mathcal{X}} g(x) \, d\mu(x) > -\infty.$$

When working with probability measures, Prokhorov's theorem ([7, Theorems 5.1–5.2], repeated for convenience in the appendix, Theorem D.1) is useful since it characterizes relatively compact sets with

https://doi.org/10.1017/S0956792525100120 Published online by Cambridge University Press

respect to the narrow topology. In certain situations, the assumption (D.1) of this theorem, i.e.,

$$\forall \epsilon > 0 \quad \exists K_\epsilon \text{ compact in } \mathcal{X} \text{ such that } \mu(\mathcal{X} \setminus K_\epsilon) \leq \epsilon \quad \forall \mu \in \mathcal{K},$$

can only be shown for bounded and not compact sets. There, we use the observation in the following remark to still obtain some sort of relative compactness. In the following, we denote by $\mathcal{X}_\omega$, the space $\mathcal{X}$ equipped with the weak topology $\sigma(\mathcal{X}, \mathcal{X}^*)$.

Remark 4.4. If $\mathcal{X}$ is separable and reflexive, then so is its dual. For a countable dense subset $\{x_n^*\}_{n\in\mathbb{N}}$ of $\overline{B_1^{\mathcal{X}^*}}$, we can define the norm

$$\|x\|_\omega = \sum_{n=1}^{\infty} \frac{1}{n^2} |\langle x_n^*, x\rangle|,$$

which induces the weak topology $\sigma(\mathcal{X}, \mathcal{X}^*)$ on bounded sets [73, Lemma 3.2]. This norm is a so-called Kadec norm. In particular, we have that the Borel sigma algebra $\mathcal{B}(\mathcal{X})$, generated by the norm topology, and the one generated by the weak topology $\mathcal{B}(\mathcal{X}_\omega)$ coincide and thus $\mathcal{P}(\mathcal{X}) = \mathcal{P}(\mathcal{X}_\omega)$, see [37, Theorem 1.1]. Now, let us assume that for a set $\mathcal{K} \subset \mathcal{P}(\mathcal{X}) = \mathcal{P}(\mathcal{X}_\omega)$, we have that

$$\forall \epsilon > 0 \quad \exists K_\epsilon \ \|\cdot\|\text{-bounded in } \mathcal{X}, \text{ such that } \mu(\mathcal{X} \setminus K_\epsilon) \leq \epsilon \quad \forall \mu \in \mathcal{K}. \tag{4.5}$$

Since bounded sets are subsets of $\overline{B_\epsilon}$ for ϵ large enough and $\overline{B_\epsilon}$ is compact in the weak topology $\sigma(\mathcal{X}, \mathcal{X}^*)$, Prokhorov's theorem can be applied for $\mathcal{X}_\omega$. We obtain that there exists a subsequence $(\mu^n)_{n\in\mathbb{N}} \subset \mathcal{K}$ and a limit $\mu \in \mathcal{P}(\mathcal{X}) = \mathcal{P}(\mathcal{X}_\omega)$ such that

$$\int_{\mathcal{X}} \varphi \, \mathrm{d}\mu^n \to \int_{\mathcal{X}} \varphi \, \mathrm{d}\mu \quad \forall \varphi \in C_b^\omega(\mathcal{X}), \tag{4.6}$$

where $C_b^\omega(\mathcal{X})$ now denotes the set of weakly continuous bounded functions.

The next lemma follows from [104, Theorem 6.9, Corollary 6.11, Remark 6.12] together Lemma 4.2, i.e., [44, Proposition 3].

Lemma 4.5 (Compatibility). *The narrow topology σ is weaker than the topology induced by $W_p(\cdot,\cdot)$ on $\mathcal{P}_p(\mathcal{X})$, for every $p \in [1, \infty]$. Furthermore, W_p is lower semicontinuous with respect to the narrow topology σ, i.e., for every $p \in [1, \infty]$:*

$$\left.\begin{array}{l} \mu^n \xrightarrow{\sigma} \mu \\ \nu^n \xrightarrow{\sigma} \nu \end{array}\right\} \Longrightarrow W_p(\mu, \nu) \leq \liminf_{n\to\infty} W_p(\mu^n, \nu^n).$$

Remark 4.6. For $1 \leq p < \infty$, convergence in W_p is equivalent to narrow convergence and convergence of the p-th moment [104, Theorem 6.9]. This equality is lost for the ∞-Wasserstein distance, as convergence in the narrow topology ($\mu^n \xrightarrow{\sigma} \mu$) together with $\bigcup_n \operatorname{supp}(\mu^n)$ being bounded or relatively compact no longer guarantees convergence in W_∞, as Example 3 demonstrates.

Example 3. *We consider the sequence*

$$\mu^n = \frac{n-1}{n} \delta_0 + \frac{1}{n} \delta_1,$$

where δ_t denotes the Dirac measure at $t \in \mathbb{R}$. Then we have that

$$\int_{\mathbb{R}} \varphi \, d\mu^n = \frac{n-1}{n} \varphi(0) + \frac{1}{n} \varphi(1) \xrightarrow{n\to\infty} \varphi(0) = \int_{\mathbb{R}} \varphi d\delta_0$$

for every $\varphi \in C_b(\mathbb{R})$ and thus $\mu^n \xrightarrow{\sigma} \delta_0$. However, we see that

$$W_\infty(\mu^n, \delta_0) = \min_{\gamma \in \Gamma(\mu^n, \delta_0)} \gamma - \operatorname{ess\,sup} |x - z| = |1 - 0| = 1,$$

for every $n \in \mathbb{N}$ and thus we have no convergence in W_∞.

https://doi.org/10.1017/S0956792525100120 Published online by Cambridge University Press

4.2. Absolutely continuous curves in Wasserstein spaces and the superposition principle

In this section, we employ the superposition principle to obtain alternative characterizations of absolutely continuous curves in Wasserstein spaces.

In [62], Lisini shows that for $p \in (1, \infty)$, p-absolutely continuous curves $\mu : [0, T] \to W_p$ can be written as a push forward of a Borel probability measure over the space of continuous curves. Using this statement, the author was able to derive a well-known characterization of absolutely continuous curves via solutions of continuity equations, when the underlying space of W_p is Banach. In [63], the first result was extended to Wasserstein–Orlicz spaces, which also covers the W_∞ case.

In [99, Section 4], the authors were able to derive a refined version of the result obtained in [62] that also includes the case $p = \infty$. For completeness, we state the corresponding theorems in this section and provide the proofs that specifically adapt the arguments of [62] to our setting in Appendix E.

Connected to this, we also refer to the discussion in the book by [91, Ch. 5.5.1] and the associated paper [10], where this topic was discussed as the limit $p \to \infty$, for $\mathcal{X} = \mathbb{R}^d$. We further discuss difficulties arising when the norm of underlying Banach space is not strictly convex.

Let $\mathcal{P}(C(0, T; \mathcal{X}))$ denote the space of Borel probability measures on the Banach space of continuous functions on the interval $[0, T]$. We define the evaluation map $e_t{:}C(0, T; \mathcal{X}) \to \mathcal{X}$ by

$$e_t(u) = u(t).$$

Then absolutely continuous curves in Wasserstein spaces can be represented by a Borel probability measure on $C(0, T; \mathcal{X})$ concentrated on the set of absolutely continuous curves in $\mathcal{X}$, as the following theorem from [63] shows. Here, $\mathrm{AC}^p(0, T; W_p)$ denotes the set of p-absolutely continuous curves $\mu : [0, T] \to W_p$.

Theorem 4.7 [63, Theorem 3.1]. *Let $\mathcal{X}$ be separable. For $p \in (1, \infty]$, if $\mu \in \mathrm{AC}^p(0, T; \mathcal{W}_p)$, then there exists $\eta \in \mathcal{P}(C(0, T; \mathcal{X}))$ such that*

- *η is concentrated on $\mathrm{AC}^p(0, T; \mathcal{X})$,*
- *$e_{t\#}\eta = \mu_t \quad \forall t \in [0, T]$,*
- *for a.e. $t \in [0, T]$ the metric derivative $|u'|(t)$ exists for η-a.e. $u \in C(0, T; \mathcal{X})$ and it holds the equality*

$$|\mu'|(t) = \| |u'|(t) \|_{L_p(\eta)}.$$

For a Banach space $(\mathcal{X}, \|\cdot\|)$ and a finite measure space $(\Omega, \mathcal{A}, \mu)$, we denote for $1 \le p \le \infty$ the Lebesgue–Bochner space by $L^p(\mu; \mathcal{X})$. A function $f{:}\Omega \to \mathcal{X}$ belongs to $L^p(\mu; \mathcal{X})$ if it is μ-Bochner integrable and its norm

$$\|f\|^p_{L_p(\mu;\mathcal{X})} := \int_\Omega \|f\|^p \, \mathrm{d}\mu \quad \text{for} \quad 1 \le p < \infty,$$
$$\|f\|_{L_\infty(\mu;\mathcal{X})} := \mu - \operatorname{ess\,sup} \|f\| \quad p = \infty,$$

is finite, see [34]. For a narrowly continuous curve $\mu{:}[0, T] \to \mathcal{P}_p(\mathcal{X})$, we define $\bar{\mu} \in \mathcal{P}([0, T] \times \mathcal{X})$ by

$$\int_{[0,T]\times\mathcal{X}} \varphi(t, x) d\bar{\mu} := \frac{1}{T} \int_{[0,T]} \int_{\mathcal{X}} \varphi(t, x) \, \mathrm{d}\mu_t(x) \, \mathrm{d}t$$

for every bounded Borel function $\varphi{:}[0, T] \times \mathcal{X} \to \mathbb{R}$. Let $v : [0, T] \times \mathcal{X} \to \mathcal{X}$ be a time dependent velocity field belonging to $L^p(\bar{\mu}, \mathcal{X})$, then we say (μ, v) satisfies the continuity equation

$$\partial \mu_t + \operatorname{div}(v_t \mu_t) = 0, \tag{CE}$$

if the relation

$$\frac{\mathrm{d}}{\mathrm{d}t} \int_{\mathcal{X}} \varphi \, \mathrm{d}\mu_t = \int_{\mathcal{X}} \langle D\varphi, v_t \rangle \, \mathrm{d}\mu_t \quad \forall \varphi \in C^1_b(\mathcal{X})$$

https://doi.org/10.1017/S0956792525100120 Published online by Cambridge University Press

holds in the sense of distributions in $(0,T)$. Here, $C_b^1(\mathcal{X})$ denotes the space of bounded, Fréchet-differentiable functions $\varphi : \mathcal{X} \to \mathbb{R}$, such that $D\varphi : \mathcal{X} \to \mathcal{X}^*$ is continuous and bounded. Using this notion, we define,

$$\mathrm{EC}^p(\mathcal{X}) := \left\{ (\mu, v) : \begin{array}{ll} \mu : [0,T] \to \mathcal{P}_p(\mathcal{X}) & \text{is narrowly continous, } v \in L^p(\bar{\mu};\mathcal{X}), \\ (\mu, v) & \text{satisfies the continuity equation} \end{array} \right\}.$$

As the next theorem shows, the curves contained in the support of η in Theorem 4.7 can be understood as the "characteristics" of a corresponding transport equation. The statement is an extension of [62, Theorem 7] to the case $p = \infty$. For completeness, we give an adapted proof in Appendix E. Here we assume that the Banach space also has the Radon–Nikodým property, see, e.g., [89, Ch. 5], which we recall in the following. In particular, every reflexive Banach space has this property, see [89, Corollary 5.45].

Definition 4.8 (Radon–Nikodým property). *We say that a Banach space $\mathcal{X}$ has the Radon–Nikodým property, if for every vector measure μ of bounded variation defined over a σ-algebra Σ over $\mathcal{X}$, that is absolutely continuous with respect to a finite, positive measure λ, there exists a λ-Bochner integrable function f such that $\mu(A) = \int_A f \, d\lambda$ for all $A \in \Sigma$.*

Theorem 4.9. *Let $\mathcal{X}$ be separable and satisfy the Radon–Nikodým property. If $\mu \in \mathrm{AC}^\infty([0,T];\mathcal{W}_\infty)$, then there exists a vector field $v : [0,T] \times \mathcal{X} \to \mathcal{X}$ such that $(\mu, v) \in \mathrm{EC}^\infty(\mathcal{X})$ and*

$$\|v_t\|_{L^\infty(\mu_t;\mathcal{X})} \le |\mu'|(t) \quad \text{for a.e. } t \in (0,T). \tag{4.7}$$

If in addition $\mathcal{X}$ satisfies the bounded approximation property (BAP), then the following Theorem 4.11 acts as the counterpart of Theorem 4.9 and states that solutions of the continuity equation are absolutely continuous curves. In particular, for a specific $\mu \in \mathrm{AC}^p([0,T];\mathcal{W}_p)$, the velocity v field obtained in Theorem 4.9 is minimal in the sense that

$$\|v_t\|_{L^p(\mu;\mathcal{X})} = |\mu'|(t) \le \|\tilde{v}_t\|_{L^p(\mu;\mathcal{X})} \quad \text{for a.e. } t \in (0,T) \text{ and for all } \tilde{v} \text{ satisfying } (\mu, \tilde{v}) \in \mathrm{EC}^p(\mathcal{X}).$$

We briefly recall the (BAP) and then state Theorem 4.11, which is an extension of [62, Theorem 8] to $p = \infty$. For completeness, the proof (which again is a slight modification of [62]) is provided in Appendix E.

Definition 4.10 (BAP). *A separable Banach space $\mathcal{X}$ satisfies the bounded approximation property (BAP), if there exists a sequence of finite rank linear operators $T_n : \mathcal{X} \to \mathcal{X}$ such that*

$$\lim_{n\to\infty} \|T_n x - x\| = 0.$$

In particular, every Hilbert space and every Banach space with a Schauder basis fulfils this property, see [92, Ch. 9].

Theorem 4.11. *Assume that $\mathcal{X}$ is separable and satisfies the Radon–Nikodým property as well as the bounded approximation property (BAP). If $(\mu, v) \in \mathrm{EC}^\infty(\mathcal{X})$, then $\mu \in \mathrm{AC}^\infty([0,T];\mathcal{W}_\infty)$ and*

$$|\mu'|(t) \le \|v_t\|_{L^\infty(\mu_t;\mathcal{X})} \quad \text{for a.e. } t \in (0,T).$$

Remark 4.12 (Uniqueness of the velocity field). As mentioned before, if $\mathcal{X}$ satisfies the bounded approximation property, the velocity field obtained in Theorem 4.9 is minimal. For the case that $p \in (1, +\infty)$ and the norm of the underlying Banach space $\mathcal{X}$ is strictly convex, then $\|\cdot\|_{L^p(\mu_t;\mathcal{X})}$ is also strictly convex. Then the uniqueness of the minimal velocity field follows. In the other cases, the uniqueness is lost.

Remark 4.13. Whenever Theorem 4.11 is applicable, $\|v_t\|_{L^\infty(\mu_t;\mathcal{X})} = |\mu'|(t)$ for a.e. $t \in (0,T)$ and thus (E.1) is actual an equality. For the Wasserstein spaces $p \in (1, +\infty)$, we obtain

$$\int_{\mathcal{X}} \left\| \int_{C(0,T;\mathcal{X})} u'(t) d\bar{\eta}_{x,t} \right\|^p \mathrm{d}\mu_t = \int_{\mathcal{X}} \int_{C(0,T;\mathcal{X})} \left\| u'(t) \right\|^p d\bar{\eta}_{x,t} \, \mathrm{d}\mu_t \quad \text{for a.e. } t \in (0,T)$$

https://doi.org/10.1017/S0956792525100120 Published online by Cambridge University Press

or equivalently

$$\left\| \int_{C(0,T;\mathcal{X})} u'(t) d\bar{\eta}_{x,t} \right\|^p = \int_{C(0,T;\mathcal{X})} \left\| u'(t) \right\|^p d\bar{\eta}_{x,t} \quad \text{for } \bar{\mu}\text{-a.e. } (t,x) \in [0,T] \times \mathcal{X}.$$

from corresponding calculations [62, Theorem 7]. Notice that this is the equality case of Jensen's inequality. For a strictly convex norm $\|\cdot\|$, this equality can only hold when $u'(t)$ is constant $\bar{\eta}_{x,t}$-a.e. Thus, heuristically spoken, all curves passing through a point $x \in \mathcal{X}$ at time t have the same derivative. This is in particular the reason why on an infinitesimal level optimal transport plans $\gamma_h \in \Gamma(\mu_t, \mu_{t+h})$ behave like classical optimal transport, i.e., for a.e. $t \in (0,T)$ (see [2, Proposition 8.4.6]),

$$\lim_{h\to 0} \left(\pi^1, \frac{1}{h}(\pi^2 - \pi^1) \right)_{\#} \gamma_h = (Id \times v_t)_{\#} \mu_t \quad \text{in } \mathcal{P}(\mathcal{X} \times \mathcal{X}).$$

This argument fails in the case W_∞ or when the norm $\|\cdot\|$ is not strictly convex.

4.3. Curves of maximal slope of potential energies

In addition to being separable, we now assume $\mathcal{X}$ to be reflexive, and we need the following assumption on the potential E.

Assumption 4.a. *Let $E : \mathcal{X} \to (-\infty, +\infty]$ be weakly continuous on its domain, which we assume to be closed and convex.*

The potential energy $\mathcal{E} : \mathcal{P}_\infty(\mathcal{X}) \to (-\infty, +\infty]$ is defined as

$$\mathcal{E}(\mu) := \int E(x)\, \mathrm{d}\mu(x).$$

As in section 2, we consider a minimizing movement scheme, approximating curves of maximal slope, where in each step the following minimization problem arises,

$$\operatorname*{arg\,min}_{\tilde{\mu}: W_\infty(\tilde{\mu},\mu) \le \tau} \int E(x) d\tilde{\mu}(x). \tag{4.8}$$

Notably, the ∞-Wasserstein distance in (4.8) restricts the movement of mass uniformly. Intuitively, this means that for every point $x \in \mathcal{X}$ we need to solve the local problem

$$r_\tau(x) := \operatorname*{arg\,min}_{\tilde{x} \in \overline{B_\tau(x)}} E(\tilde{x}), \tag{4.9}$$

where $r_\tau(x) : \mathcal{X} \rightrightarrows \mathcal{X}$ is a possibly multivalued correspondence, see Appendix F. Then a possible optimal transport plan between μ and a minimizer of (4.8), $\mu_{\min}$, should transport the mass from some point x to a minimizing point in $r_\tau(x)$. In this regard, we employ the measurable maximum theorem ([24, Theorem 18.19], repeated for convenience in the appendix as Theorem F.3) . This theorem guarantees the measurability of the "argmin" correspondence in (4.9). Definitions of (weak) measurability for correspondences are repeated in Appendix F, where we refer to [24] for a detailed overview over the topic. In order to apply the mentioned theorems to the problem in (4.9), we need to check the underlying correspondence for weak measurability. Let us define

$$\mathrm{dom}_\tau\,(E) := \{x \in \mathcal{X} : \|x - z\| \le \tau \text{ for a } z \in \mathrm{dom}\,(E)\}.$$

Lemma 4.14. *For $\tau \ge 0$, the correspondence $\varphi_\tau : (\mathcal{X} \cap \mathrm{dom}_\tau\,(E), \|\cdot\|) \rightrightarrows (\mathcal{X} \cap \mathrm{dom}\,(E), \|\cdot\|_\omega)$ given by $\varphi_\tau : x \mapsto \overline{B_\tau}(x)$ is weakly measurable and has nonempty weakly compact values.*

Proof. Every weakly open set $G \subset \mathcal{X} \cap \mathrm{dom}\,(E)$ is strongly open as well. And for strongly open sets G, the lower inverse as defined in (F.1) is given as

$$\varphi_\tau^l(G) = \{s \in \mathcal{X} \mid \exists\, x \in G \text{ with } \|s - x\| \le \tau\}.$$

https://doi.org/10.1017/S0956792525100120 Published online by Cambridge University Press

Since G is strongly open, this set is again strongly open, and thus in $\Sigma = \mathcal{B}(\mathcal{X})$, yielding weak measurability of φ_τ. To conclude, we observe that $\overline{B_\tau}(x)$ is nonempty and weakly compact. □

The next corollary now follows immediately from the measurable maximum theorem.

Corollary 4.15. *Let $\mathcal{X}$ be a reflexive, separable Banach space and let E fulfil Assumption 4.a, then for $\tau \geq 0$*

$$E_\tau(x) := \min_{\tilde{x} \in \overline{B_\tau}(x)} E(\tilde{x}) \tag{4.10}$$

is $\mathcal{B}(\mathcal{X})$-measurable. The correspondence $r_\tau : \mathcal{X} \rightrightarrows \mathcal{X}$

$$r_\tau(x) := \arg\min_{\tilde{x} \in \overline{B_\tau}(x)} E(\tilde{x}) \tag{4.11}$$

has nonempty and compact values, it is measurable and admits a $\mathcal{B}(\mathcal{X})$-measurable selector.

Proof. As mentioned in remark Remark 4.4, $\mathcal{B}(\mathcal{X})$ and $\mathcal{B}(\mathcal{X}_\omega)$ coincide in this particular setting. We choose the correspondence φ_τ from Lemma 4.14 and set $f(s, x) = -E(x)$. Since Assumption 4.a guarantees that $f(s, x) = -E(x)$ is a Carathéodory function the application of Theorem F.3 yields this corollary, but only restricted to $\mathcal{X} \cap \mathrm{dom}_\tau\,(E)$. However, we can extend $E_\tau(x)$ and r_τ measurably by setting them to $+\infty$ and $\overline{B}_\tau(x)$ on $\mathrm{dom}_\tau\,(E)^c$ respectively. □

Theorem 4.16. *Let $\mathcal{X}$ be a reflexive, separable Banach space and let E fulfil Assumption 4.a, then*

$$\mu_\tau := (r_\tau)_\# \mu \in \arg\min_{\tilde{\mu}: W_\infty(\tilde{\mu}, \mu) \leq \tau} \int E(x) d\tilde{\mu}(x)$$

for every measurable selection r_τ of r_τ from (4.11).

Proof. Corollary 4.15 ensures the existence of measurable selectors of (4.11). We take $\tilde{\mu}$, such that $W_\infty(\mu, \tilde{\mu}) \leq \tau$ and $\gamma \in \Gamma_0(\mu, \tilde{\mu})$, then by disintegration we get

$$\int E(x) \mathrm{d}\tilde{\mu}(x) = \int E(x) \mathrm{d}\gamma(z, x) = \int \int E(x) \mathrm{d}\rho_z(x)\, \mathrm{d}\mu(z)$$

with a Borel family of probability measures $\{\rho_z\}_{z \in \mathcal{X}_1} \subset \mathcal{P}(\mathcal{X})$ and $\mathrm{supp}\,(\rho_z) \subset \overline{B_\tau}(z)$. We further estimate,

$$\int \int E(x) \mathrm{d}\rho_z(x) \mathrm{d}\mu(z) \geq \int E(r_\tau(z)) \mathrm{d}\mu(z) = \int E(z)\, \mathrm{d}(r_\tau)_\# \mu(z),$$

and since $\tilde{\mu}$ was arbitrary, this concludes the proof. □

In order to proceed with the following lemma, we also need the assumption that the potential is a C^1-perturbation of a convex function and is Lipschitz continuous.

Assumption 4.b. *Let $E : \mathcal{X} \to (-\infty, +\infty]$ be a C^1-perturbation of a proper, convex lower semicontinuous function. Further, let the differentiable part E^d be globally Lipschitz.*

Then the relation between the slope of $\mathcal{E}$ and the slope of the potential E is stated in the following theorem.

Lemma 4.17. *Let $E : \mathcal{X} \to (-\infty, +\infty]$ fulfil Assumptions 4.a and 4.b. Then*

$$|\partial \mathcal{E}|(\mu) = \int_{\mathcal{X}} |\partial E|(x)\, d\mu(x) \tag{4.12}$$

and $|\partial \mathcal{E}|(\mu)$ is a strong upper gradient of $\mathcal{E}$.

https://doi.org/10.1017/S0956792525100120 Published online by Cambridge University Press

Proof. Since $\frac{E(x)-E_\tau(x)}{\tau} \geq 0$ we can use Fatou's lemma to show

$$\begin{aligned}\int_{\mathcal{X}} |\partial E|(x)\,\mathrm{d}\mu(x) &= \int_{\mathcal{X}} \lim_{\tau\to 0} \frac{E(x)-E_\tau(x)}{\tau}\,\mathrm{d}\mu(x)\\ &\leq \liminf_{\tau\to 0} \int_{\mathcal{X}} \frac{E(x)-E_\tau(x)}{\tau}\,\mathrm{d}\mu(x)\\ &\leq \limsup_{\tau\to 0} \int_{\mathcal{X}} \frac{E(x)-E(r_\tau(x))}{\tau}\,\mathrm{d}\mu(x)\\ &= \limsup_{\tau\to 0} \frac{\mathcal{E}(\mu)-\mathcal{E}(\mu_\tau)}{\tau} = |\partial\mathcal{E}|(\mu),\end{aligned}$$

where in the last step, we employ Lemma 2.8 . This implies that when $\int_{\mathcal{X}} |\partial E|(x)\,\mathrm{d}\mu(x) = +\infty$ then $|\partial\mathcal{E}|(\mu) = +\infty$. In the case $\int_{\mathcal{X}} |\partial E|(x)\,\mathrm{d}\mu(x) < +\infty$, we use Lemma 2.8 (for $\mathcal{E}$) and Lemma 3.5 (for E) to calculate

$$\begin{aligned}\int_{\mathcal{X}} |\partial E|(x)\,\mathrm{d}\mu(x) &= \int_{\mathcal{X}} \lim_{\tau\to 0} \frac{E(x)-E_\tau(x)}{\tau}\,\mathrm{d}\mu(x)\\ &= \lim_{\tau\to 0} \frac{\int_{\mathcal{X}} E(x)-E(r_\tau(x))\,\mathrm{d}\mu(x)}{\tau}\\ &= \lim_{\tau\to 0} \frac{\mathcal{E}(\mu)-\mathcal{E}(\mu_\tau)}{\tau} = |\partial\mathcal{E}|(\mu),\end{aligned}$$

where the dominated convergence theorem was used to draw the limit into the integral. For the upper bound, we observe

$$\begin{aligned}|\partial E|(x) &= \limsup_{z\to x} \frac{(E^{\mathrm{c}}(x)-E^{\mathrm{c}}(z)+E^{\mathrm{d}}(x)-E^{\mathrm{d}}(z))^+}{\|x-z\|}\\ &\geq \limsup_{z\to x} \frac{(E^{\mathrm{c}}(x)-E^{\mathrm{c}}(z))^+}{\|x-z\|} - \frac{|E^{\mathrm{d}}(x)-E^{\mathrm{d}}(z)|}{\|x-z\|}\\ &\geq \limsup_{z\to x} \frac{(E^{\mathrm{c}}(x)-E^{\mathrm{c}}(z))^+}{\|x-z\|} - \mathrm{Lip}(E^{\mathrm{d}}) = |\partial E^{\mathrm{c}}|(x) - \mathrm{Lip}(E^{\mathrm{d}})\end{aligned}$$

and by [2, Theorem 2.4.9]

$$\sup_{z\neq x} \frac{(E^{\mathrm{c}}(x)-E^{\mathrm{c}}(z))^+}{\|x-z\|} = |\partial E^{\mathrm{c}}|(x).$$

Then we can give an upper bound by

$$\begin{aligned}\frac{E(x)-E(r_\tau(x))}{\tau} &\leq \frac{E^{\mathrm{c}}(x)-E^{\mathrm{c}}(r_\tau(x))}{\|x-r_\tau(x)\|} + \frac{E^{\mathrm{d}}(x)-E^{\mathrm{d}}(r_\tau(x))}{\|x-r_\tau(x)\|}\\ &\leq \sup_{z\neq x} \frac{(E^{\mathrm{c}}(x)-E^{\mathrm{c}}(z))^+}{\|x-z\|} + \mathrm{Lip}(E^{\mathrm{d}}) = |\partial E^{\mathrm{c}}|(x) + \mathrm{Lip}(E^{\mathrm{d}})\\ &\leq |\partial E|(x) + 2\mathrm{Lip}(E^{\mathrm{d}}).\end{aligned}$$

To prove that $|\partial\mathcal{E}|$ is a strong upper gradient, let μ_t be an absolutely continuous curve in $\mathcal{W}_\infty(\mathcal{X})$. Since $|\partial\mathcal{E}|(\mu) = \int_{\mathcal{X}} |\partial E|(x)\,\mathrm{d}\mu(x)$ and by Item 4 the slope $|\partial E|(x)$ is lower semicontinuous, it follows from [2, Lemma 5.1.7] that $|\partial\mathcal{E}|(\mu)$ is lower semicontinuous w.r.t. narrow convergence and in particular $t \mapsto |\partial\mathcal{E}|(\mu_t)$ is lower semicontinuous and thus Borel. Assume that $\int_s^t \int_{\mathcal{X}} |\partial E|(x)\,\mathrm{d}\mu_r(x)|\mu'|(r)$

https://doi.org/10.1017/S0956792525100120 Published online by Cambridge University Press

$\mathrm{d}r = \int_s^t |\partial\mathcal{E}|\,(\mu_r)|\mu'|(r)\,\mathrm{d}r < +\infty$, otherwise (1.4) holds trivially. We can estimate

$$\begin{aligned}
|\mathcal{E}(\mu_t) - \mathcal{E}(\mu_s)| &= \left| \int_{\mathcal{X}} E(x)\,\mathrm{d}\mu_t(x) - \int_{\mathcal{X}} E(x)\,\mathrm{d}\mu_s(x) \right| \\
&= \left| \int_{\mathcal{X}} E(x)\mathrm{d}e_{t\#}\eta(x) - \int_{\mathcal{X}} E(x)\mathrm{d}e_{s\#}\eta(x) \right| \\
&= \left| \int_{C(0,T;\,\mathcal{X})} E(u(t))\,\mathrm{d}\eta(u) - \int_{C(0,T;\,\mathcal{X})} E(u(s))\mathrm{d}\eta(u) \right| \\
&\leq \int_{C(0,T;\,\mathcal{X})} |E(u(t)) - E(u(s))|\,\mathrm{d}\eta(u) \\
&\overset{(i)}{\leq} \int_{C(0,T;\,\mathcal{X})} \int_s^t |\partial E|(u(r))\,|u'|(r)\,\mathrm{d}r\,\mathrm{d}\eta(u) \\
&\overset{(ii)}{=} \int_s^t \int_{C(0,T;\,\mathcal{X})} |\partial E|(u(r))\,|u'|(r)\,\mathrm{d}\eta(u)\,\mathrm{d}r \\
&\overset{(iii)}{\leq} \int_s^t \int_{C(0,T;\,\mathcal{X})} |\partial E|(u(r))\,\mathrm{d}\eta(u)|\mu'|(r)\,\mathrm{d}r \\
&= \int_s^t \int_{\mathcal{X}} |\partial E|(x)\,\mathrm{d}\mu_r|\mu'|(r)\,\mathrm{d}r < +\infty,
\end{aligned} \tag{4.13}$$

where $(t,u) \mapsto |\partial E|(u(t))$ is $\bar{\eta}$-measurable since it is lower semicontinuous on $[0,T] \times C(0,T,\mathcal{X})$ and measurability of $|u'|$ follows as in the proof of [62, Theorem 7] and Theorem 4.9. For (ii), we use the theorem of Fubini–Tonelli, while for (i), we observe that η from Theorem 4.7 is concentrated on $\mathrm{AC}^\infty(0,T;\mathcal{X})$ and $|\partial E|$ is a strong upper gradient (c.f. Definition 1.4) and for (iii) we use $|\mu'|(t) = \||u'|(t)\|_{L_p(\eta)}$. □

The main result of this section now states that ∞-curves of maximal slope on $W_\infty(\mathcal{X})$ can be equivalently characterized, by the property that η-a.e. curve fulfils the differential inclusion w.r.t. the potential E on the Banach space $\mathcal{X}$.

Theorem 4.18. *Let $\mathcal{E} : W_\infty(\mathcal{X}) \to (-\infty, +\infty]$ be a potential energy with the potential E satisfying Assumptions 4.a and 4.b, $\mu_t \in \mathrm{dom}\,(\mathcal{E})$ for all $t \in [0,T]$ and $\mu \in \mathrm{AC}^\infty(0,T;\mathcal{W}_\infty)$ with η from Theorem 4.7. Let further $\mathcal{E} \circ \mu$ be for a.e. $t \in [0,T]$ equal to a non-increasing map $\psi : [0,T] \to \mathbb{R}$. Then the following statements are equivalent:*

- *(i) $|\mu'|(t) \leq 1$ and $\psi'(t) \leq -|\partial\mathcal{E}|(\mu(t))$ for a.e. $t \in (0,T)$.*
- *(i) For η-a.e. curve $u \in C(0,T;\mathcal{X})$ it holds, that $E \circ u$ is for a.e. $t \in (0,T)$ equal to a non-increasing map ψ_u:$[0,T] \to \mathbb{R}$ and*

$$u'(t) \;\in\; \partial\|\cdot\|_*(-\xi) \quad \forall \xi \in \partial^\circ E(u(t)) \neq \emptyset, \quad \text{for a.e. } t \in (0,T).$$

Proof. Step 1: (i) $\Longrightarrow$ (ii).
Because of Remark 2.3 we know that μ_t satisfies the energy dissipation equality (2.10). Making a similar estimate as in (4.13), we obtain

$$\begin{aligned}
\mathcal{E}(\mu_0) - \mathcal{E}(\mu_T) &= \int_{C(0,T;\,\mathcal{X})} E(u(0)) - E(u(T))\,\mathrm{d}\eta(u) \\
&\leq \int_{C(0,T;\,\mathcal{X})} \int_0^T |\partial E|(u(r))|u'|(t)\,\mathrm{d}r\,\mathrm{d}\eta(u) \\
&\leq \int_{C(0,T;\,\mathcal{X})} \int_0^T |\partial E|(u(r))\,\mathrm{d}r\,\mathrm{d}\eta(u) \\
&= \int_0^T |\partial\mathcal{E}|(\mu(r))\,\mathrm{d}r = \mathcal{E}(\mu_0) - \mathcal{E}(\mu_T).
\end{aligned} \tag{4.14}$$

https://doi.org/10.1017/S0956792525100120 Published online by Cambridge University Press

This implies that $|\partial E|(u(r))|u'|(t) \in L^1(0,T)$ for η-a.e. u. Since $|\partial E|$ is a strong upper gradient $\psi_u : t \mapsto E(u(t))$ has to be absolutely continuous for η-a.e. u and

$$E(u(s)) - E(u(t)) \le \int_s^t |\partial E|(u(r))|u'|(r)\,\mathrm{d}r \le \int_s^t |\partial E|(u(r))\,\mathrm{d}r \quad \text{for } \eta\text{-a.e. } u \text{ for all } 0 \le s < t \le T.$$

Equality in (4.14) can then only hold if for all $0 \le s < t \le T$ we have

$$E(u(s)) - E(u(t)) = \int_s^t |\partial E|(u(r))|u'|(t)\,\mathrm{d}r = \int_s^t |\partial E|(u(r))\,\mathrm{d}r \quad \text{for } \eta\text{-a.e. } u$$

and thus $\psi_u := E \circ u$ is a non-increasing map for η-a.e. u. Lemma 3.7 and Item 3 imply that for every $\xi \in \partial^\circ E(u(t))$ we obtain

$$\langle \xi, u'(t) \rangle = (E \circ u)'(t) = -|\partial E|(u(t)) = -\|\xi\|_* - \chi_{\overline{B_1}}(u'(t)),$$

where we use Lemma E.4 and $\|\,|u'|(t)\|_{L^\infty(\eta)} = |\mu'|(t) \le 1$ for a.e. $t \in (0,T)$ to infer that $|u'|(t) \le 1$ for a.e. $t \in (0,T)$. Using the equivalence of Item 3 and Item 1 yields

$$u'(t) \in \partial\, \|\cdot\|_*(-\xi)$$

for a.e. $t \in (0,T)$ and η-a.e. curve u.

Step 2: (ii) $\Longrightarrow$ (i).

Due to Remark 2.3, $E \circ u$ is for η-a.e. curve $u \in C(0,T;\mathcal{X})$ an absolutely continuous curve, and it satisfies the energy dissipation equality

$$E(u(t)) - E(u(s)) = \int_s^t -|\partial E|(u(r))\,\mathrm{d}r \quad \text{for } 0 \le s \le t \le T.$$

Therefore, we obtain

$$\begin{aligned}\mathcal{E}(\mu_t) - \mathcal{E}(\mu_s) &= \int_{C(0,T;\mathcal{X})} E(u(t)) - E(u(s))\,\mathrm{d}\eta(u) = \int_{C(0,T;\mathcal{X})} \int_s^t -|\partial E|(u(r))\,\mathrm{d}r\,\mathrm{d}\eta(u) \\ &= \int_s^t \int_{C(0,T;\mathcal{X})} -|\partial E|(u(r))\,\mathrm{d}\eta(u)\,\mathrm{d}r = \int_s^t -|\partial \mathcal{E}|(\mu_r)\,\mathrm{d}r \le 0,\end{aligned}$$

where the application of Fubini–Tonelli is justified due to the assumption $\mu_t \in \mathrm{dom}(\mathcal{E})$ for all $t \in [0,T]$, which yields that $|\mathcal{E}(\mu_t) - \mathcal{E}(\mu_s)| < \infty$ for all $s,t \in [0,T]$. By the Lebesgue differentiation theorem, we obtain

$$(\mathcal{E} \circ \mu)'(t) = -|\partial \mathcal{E}|(\mu_t)$$

for almost every $t \in (0,T)$. Furthermore, $\mathcal{E} \circ \mu$ is a non-increasing map and Theorem 4.7 yields that

$$|\mu'|(t) = \eta(u) - \operatorname{ess\,sup} |u'|(t) \le 1$$

for a.e. $t \in (0,T)$ and since $\mu \in \mathrm{AC}^\infty$ this yields that $|\mu'|(t) \le 1$ for a.e. $t \in (0,T)$. □

Remark 4.19. In particular, those curves of maximal slope satisfy the continuity equation for the velocity field

$$v_t(x) := \int_{C(0,T;\mathcal{X})} u'(t)\mathrm{d}\bar{\eta}_{x,t} \quad \text{for } \bar{\mu}\text{-a.e. } (t,x) \in (0,T) \times \mathcal{X}.$$

If $\partial^\circ E(x)$ is unique, i.e., if $\|\cdot\|$ is strictly convex or $E(x) \in C^1(\mathcal{X})$, then for $\bar{\eta}_{x,t}$-a.e. $u \in C(0,T;\mathcal{X})$ the derivatives $u'(t)$ lie in the closed and convex set $\partial\|\cdot\|_*(-\partial^\circ E(x))$. Thus

$$v_t(x) \in \partial\|\cdot\|_*(-\partial^\circ E(x)) \quad \text{for } \bar{\mu}\text{-a.e. } (t,x) \in (0,T) \times \mathcal{X}.$$

As the last result in this section, we give an explicit setting where the existence of curves of maximum slope is ensured. Here, we restrict ourselves to finite dimensions, mimicking Corollary 3.2.

Corollary 4.20 (Existence in finite dimensions). *Let $\mathcal{X} = (\mathbb{R}^d, \|\cdot\|)$ and $E : \mathbb{R}^d \to (-\infty, \infty]$ be a C^1-perturbation of a proper, lower semicontinuous, convex function. For every $\mu^0 \in \mathrm{dom}(\mathcal{E})$, there exists at*

https://doi.org/10.1017/S0956792525100120 Published online by Cambridge University Press

least one curve of maximal slope in the sense of Definition 2.1 with $\mu_0 = \mu^0$. Further, this curve satisfies the energy dissipation equality (2.10).

Proof. We simply check the conditions of Theorem 2.11. Choosing σ to be the narrow topology, Lemma 4.5 guarantees Assumption 1.a. To check Assumption 1.b, we know that for any sequence $(\mu^n)_{n\in\mathbb{N}}$, $W_\infty(\mu^k, \mu^m) < \infty \quad \forall k, m \in \mathbb{N}$ implies that $\cup_n supp(\mu^n)$ is bounded. Since we are in the finite dimensional case, we can now apply Prokhorov's Theorem to obtain relative compactness of the sequence.

We are left to check Assumptions 2.a and 2.b for $\mathcal{E}$:

Assumption 2.a

Let $\mu^n \in \operatorname{dom}(\mathcal{E})$ be a sequence converging in W_∞ to μ. This sequence has to be bounded in W_∞ such that $\overline{\cup_n \operatorname{supp}(\mu^n)}$ is bounded. Since E is lower-semicontinuous $\operatorname{dom}(E)$ is closed and thus $\overline{\cup_n \operatorname{supp}(\mu^n)} \cap \operatorname{dom}(E)$ is compact and we obtain due to lower-semicontinuity

$$\min_{x\in\overline{\cup_n \operatorname{supp}(\mu^n)}\cap\operatorname{dom}(E)} E(x) > -\infty$$

Thus the negative part of $E(x)$ denoted by $E^-(x)$ is uniformly integrable with respect to $\{\mu^n\}_{n\in\mathbb{N}}$ and we can apply [2, Lemma 5.1.7] to obtain

$$\liminf_{n\to\infty} \int E(x)\,\mathrm{d}\mu^n(x) \geq \int E(x)\,\mathrm{d}\mu.$$

Assumption 2.b:

Since μ has bounded domain the differentiable part E^d satisfies a Lischitz condition and thus by Lemma 4.17

$$|\partial\mathcal{E}|(\mu) = \int |\partial E|(x)\,\mathrm{d}\mu$$

and by Proposition 3.1 $|\partial E|(x)$ is lower semicontinuous and non-negative. Thus $|\partial E|(x)$ is uniformly integral, and we can apply [2, Lemma 5.1.7] to obtain

$$\liminf_{n\to\infty} \int |\partial E|(x)\,\mathrm{d}\mu^n \geq \int |\partial E|(x)\,\mathrm{d}\mu$$

for all μ^n converging narrowly to μ. □

5. Relation to adversarial attacks

This section explores the connection of the previous results to our initial motivation, adversarial attacks. As mentioned before, we now consider an energy defined as

$$E(x) := -\ell(h(x), y)$$

for a classifier h and $x \in \mathcal{X}, y \in \mathcal{Y}$. The goal in (AdvAtt) is to maximize this function on the set $\overline{B_\varepsilon}(x^0)$, where $x^0 \in \mathcal{X}$ is the initial input. Roughly following the idea in the original paper proposing (FGSM), we derive the scheme, via linearizing E around x^0 and consider the linearized minimizing movement scheme in Definition 3.10. Assuming that $\ell(h(\,\cdot\,), y)$ is C^1, we consider

$$E^{\mathrm{sl}}(x;z) = -\ell(h(z), y) - \langle \nabla_x \ell(h(z), y), x - z\rangle,$$

where z denotes the point of linearization. Lemma 3.12 yields that the semi-implicit minimizing movement scheme in Definition 3.10 can be expressed as

$$x^{k+1}_{\mathrm{si},\tau} \in x^k_{\mathrm{si},\tau} - \tau\,\partial\,\|\cdot\|_*\,(DE(x^k_{\mathrm{si},\tau})).$$

We note that this scheme can be understood as an explicit Euler discretization [38] of the differential inclusion in Theorem 3.8,

$$u'(t) \in \arg\max_{x\in\overline{B_1}} \langle x, -DE(u(t))\rangle = \partial\,\|\cdot\|_*\,(DE(u(t))), \tag{5.1}$$

https://doi.org/10.1017/S0956792525100120 Published online by Cambridge University Press

which in turn is an equivalent characterization of ∞-curves of maximal slope. In this section, we consider the finite dimensional adversarial setting, i.e., the Banach space $(\mathcal{X}, \|\cdot\|) = (\mathbb{R}^d, \|\cdot\|_p)$.

Corollary 5.1. *Given $x^0 \in \mathbb{R}^d$, the iteration*

$$x^{k+1}_{\mathrm{si},\tau} = x^k_{\mathrm{si},\tau} + \tau \ \operatorname{sign}(\nabla_x E(x^k_{\mathrm{si},\tau})) \cdot \left(\frac{\left|\nabla_x E(x^k_{\mathrm{si},\tau})\right|}{\left\|\nabla_x E(x^k_{\mathrm{si},\tau})\right\|_q} \right)^{q-1}, \qquad x^0_{\mathrm{si},\tau} = x^0$$

fulfils the semi-implicit minimizing movement scheme in Definition 3.10 in the space $(\mathbb{R}^d, \|\cdot\|_p)$ with $1/p + 1/q = 1$. In this sense, (FGSM) is a one-step explicit Euler discretization of the differential inclusion (5.1) with step size ε.

Remark 5.2. We note that for $p \in \{1, \infty\}$, the expression in Corollary 5.1 is to be understood in the sense of subdifferentials, as the following proof shows. However, the elements of the subdifferential we choose can be understood as the limit cases of $p \to 1$ and $p \to \infty$, respectively.

Proof. We choose $\mathcal{X} = \mathbb{R}^d$ with $\|\cdot\| = \|\cdot\|_p$. For $p = \infty$, we have that

$$\operatorname{sign}(\xi) \in \partial \|\cdot\|_1(\xi) = \partial(\|\cdot\|_\infty)_*(\xi),$$

for all $\xi \in \mathbb{R}^d$ and therefore, the following iteration fulfils the semi-implicit minimizing movement scheme,

$$x^{k+1}_{\mathrm{si},\tau} = x^k_{\mathrm{si},\tau} - \tau \operatorname{sign}(\nabla_x E(x^k_{\mathrm{si},\tau})) = x^k_{\mathrm{si},\tau} + \varepsilon \operatorname{sign}(\nabla_x \ell(h(x^k_{\mathrm{si},\tau}), y)),$$

and for $\varepsilon = \tau$ the statement follows. For $p = 1$, we choose the following element of the subdifferential $g(\xi)$, with

$$g(\xi)_i := \#\{j : \left|\xi_j\right| = \|\xi\|_\infty\}^{-1} \cdot \begin{cases} \operatorname{sign}(\xi_i) & \text{if } |\xi_i| = \|\xi\|_\infty, \\ 0 & \text{else,} \end{cases}$$

and proceed as before. If we instead choose a finite $p \in (1, \infty)$, we obtain for $1/p + 1/q = 1$,

$$\partial(\|\cdot\|_p)_*(\xi) = \partial \|\cdot\|_q(\xi) = \|\xi\|_q^{1-q} (\xi_1 |\xi_1|^{q-2}, \dots, \xi_d |\xi_d|^{q-2}) = \operatorname{sign}(\xi) \cdot \left(\frac{|\xi|}{\|\xi\|_q} \right)^{q-1},$$

where the absolute value and the multiplication is to be understood entrywise. As above, this yields the statement also for $p \in (1, \infty)$. □

5.1. Convergence of IFGSM to curves of maximal slope

Our main goal is to derive a convergence result of (IFGSM) for $\tau \to 0$. As mentioned before, Lemma 3.12 yields an iteration, which can be expressed as normalized gradient descent in the finite-dimensional case. The main obstacle that prohibits us from directly applying the convergence result for semi-implicit schemes (see Theorem 3.16) is the budget constraint, $u'(t) \in \overline{B^p_\varepsilon}(x^0)$ for all t. Here and in the following, we now assume that the norm exponent of the underlying space and of the budget constraint norm are the same. In (IFGSM), this is enforced via a projection onto this set in each iteration. An easy way to circumvent this issue is to only consider the iteration up to the step, where it would leave the constraint set. In this case, the projection never has any effect and we essentially consider signed gradient descent. Intuitively, the Lipschitz condition $\|u'(t)\| \le 1$ allows us to control how far $u(t)$ is away from x^0. This mimicked in the discrete scheme, where we know that

$$\left\|x^i_{\mathrm{si},\tau} - x^0\right\| \le \sum_{k=0}^{n-1} \left\|x^{k+1}_{\mathrm{si},\tau} - x^k_{\mathrm{si},\tau}\right\| \le n\tau = T,$$

https://doi.org/10.1017/S0956792525100120 Published online by Cambridge University Press

for every $i = 0, \dots, n$. Therefore, we can choose $T = \varepsilon$ to ensure that $x^i_{\mathrm{si},\tau} \in \overline{B^p_\varepsilon}(x^0)$ for every $i = 0, \dots, n$. This yields the following result.

Corollary 5.3. *We consider the space $(\mathcal{X}, \|\cdot\|) = (\mathbb{R}^d, \|\cdot\|_p)$ for $p \in [1, \infty]$ and $E : \mathbb{R}^d \to \mathbb{R}$, a continuously differentiable energy, with a Lipschitz continuous gradient. Then for $T = \varepsilon$, there exists a ∞-curve of maximal slope $u : [0, T] \to \mathbb{R}^d$, with respect to E, and a subsequence of $\tau_n := T/n$ such that*

$$\left\| u^{\lceil t/\tau_{n_i} \rceil}_{\mathrm{IFGS},\tau_{n_i}} - u(t) \right\| \xrightarrow{i\to\infty} 0 \qquad \text{for all } t \in [0, T].$$

Proof. From Lemma 3.12 and the calculation in the proof of Corollary 5.1, we know that the iterates of (IFGSM) fulfil the linearized minimizing movement scheme in Definition 3.10. Here, we used that for $T = \varepsilon$, the iterates do not leave the set $\overline{B^p_\varepsilon}(x^0)$ and therefore the projection has no effect. Assumption 3.a is stated as an assumption of this corollary and Remark 3.9 yields that Assumption 3.b holds true. Furthermore, using Proposition 3.1, we know that Assumptions 1.a to 2.b are fulfilled, and therefore, we can apply Theorem 3.16 to obtain the desired result. □

Above, we only consider convergence up to a subsequence. While the convergence of the whole sequence for (IFGSM) is left unanswered in this work, we note that at least for $p \in \{1, \infty\}$, this cannot be expected, since in this case ∞-curves of maximal slope lack uniqueness, even in the simple finite dimensional case, as the following example shows.

Example 4 (Non uniqueness for $p \in \{1, \infty\}$). *Let $(\mathcal{X}, \|\cdot\|) = (\mathbb{R}^2, \|\cdot\|_\infty)$ and consider the energy be given by*

$$E : (x_1, x_2) \in \mathbb{R}^2 \mapsto x_1 \in \mathbb{R}$$

then both $u_1(t) = (-t, 0)$ and $u_2(t) = (-t, -t)$ are ∞-curves of maximal slope on $[0, T]$, $T > 0$, with $u_1(0) = u_2(0)$ since

$$u_1'(t) = (-1, 0) \in -\partial \|\cdot\|_1 (1, 0) = -\partial \|\cdot\|_1 (\nabla E(u_1(t)))$$

and

$$u_2'(t) = (-1, -1) \in -\partial \|\cdot\|_1 (1, 0) = -\partial \|\cdot\|_1 (\nabla E(u_2(t))).$$

In two dimensions for $p = 1$, we can simply rotate the above setup to deduce the same non-uniqueness. Namely for $E(x_1, x_2) = x_1 + x_2$, we have that $u_1(t) = (-t, 0)$ fulfils

$$u_1'(t) = (-1, 0) \in -\partial \|\cdot\|_\infty (1, 1) = -\partial \|\cdot\|_\infty (\nabla E(u_1(t)))$$

and also $u_2(t) = \frac{1}{2}(-t, -t)$ fulfils

$$u_2'(t) = \frac{1}{2}(-1, -1) \in -\partial \|\cdot\|_\infty (1, 1) = -\partial \|\cdot\|_\infty (\nabla E(u_2(t))).$$

In Corollary 5.3, we only allow the iteration to run until it hits the boundary. However, in practice, it is more common to also iterate beyond the time ε. In order to incorporate the budget constraint in this case, we modify the energy to

$$E(x) := -\ell(h(x), y) + \chi_{\overline{B^p_\varepsilon}(x^0)}(x),$$

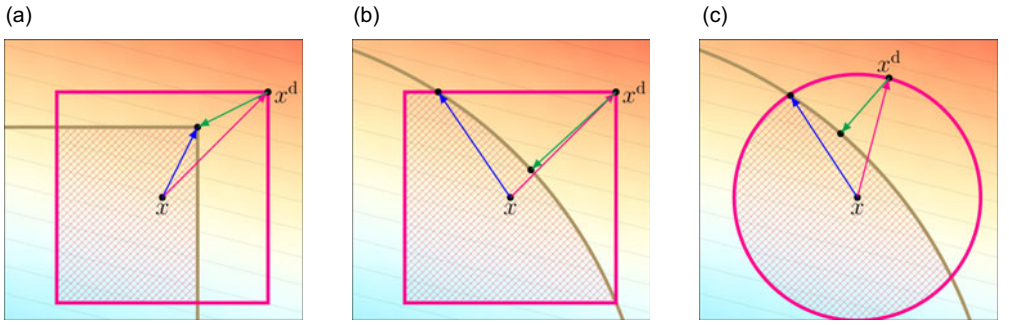
which yields the semi-implicit energy

$$E^{\mathrm{sl}}(x; z) = -\ell(h(z), y) - \langle \nabla_x \ell(h(z), y), x - z \rangle + \chi_{\overline{B^p_\varepsilon}(x^0)}(x).$$

In order to show that (IFGSM) corresponds to the minimizing movement scheme, we need to show that first minimizing on $\overline{B^p_\tau}(x)$ and then projecting to $\overline{B^p_\varepsilon}(x^0)$ is equivalent to directly minimizing on $\overline{B^p_\varepsilon}(x^0) \cap \overline{B^p_\tau}(x)$. Here, we restrict ourselves to the case $p = \infty$, which corresponds to the standard case of (IFGSM) as proposed in [46]. For $p \neq \infty$, a more refined analysis would be required, c.f. Figure 3. In the following lemma, we use the projection defined componentwise as

$$\mathrm{Clip}_{x^0,\varepsilon}(x)_j := \Pi_{\overline{B^\infty_\varepsilon}(x^0)}(x)_j = x^0_j + \max\{\min\{x_j - x^0_j, \varepsilon\}, -\varepsilon\}.$$

https://doi.org/10.1017/S0956792525100120 Published online by Cambridge University Press



Underlying space and budget constraint with $p = \infty$.

Norm mismatch: underlying space with $p = \infty$ and budget constraint with $q = 2$.

Underlying space and budget constraint with $p = 2$.

Figure 3. *Visualization of one (IFGSM) step, employing different norm constraints and underlying norms. The beige line marks the boundary of $B_\varepsilon^p(x^0)$, the pink line the boundary of $B_\tau^q(x)$ and the intersection $\overline{B_\varepsilon^p}(x^0) \cap \overline{B_\tau^q}(x)$ is hatched. For the case $p = q = \infty$ minimizing a linear function on the intersection (blue arrow) is equivalent to first minimizing on $\overline{B_\tau^\infty}(x)$ (pink arrow) and then projecting back to the intersection (green arrow). This is not true for $p = 2$. Therefore, we need to choose the appropriate projection in Lemma 5.4.*

The proof relies on the basic intuition in the original paper [46] that maximizing the linearized energy on a hyper-cube is a linear programme [41, 95] with a solution being attained in a corner. We also note that this does not directly work for other choices of budget constraints, see Figure 3

Lemma 5.4. *For $x \in \overline{B_\varepsilon^\infty}(x^0)$ and $\tau > 0$, it holds that*

$$\mathrm{Clip}_{x^0,\varepsilon}(x + \tau\ \mathrm{sign}\,(\nabla_x \ell(h(x), y))) \in \operatorname*{arg\,min}_{\tilde{x} \in \overline{B_\tau^\infty}(x)} E^{\mathrm{sl}}(\tilde{x}; x).$$

Proof. Without loss of generality, we assume that $x^0 = 0$. Let $\xi := -\nabla_x \ell(h(x), y)$, then we know that $x^{\mathrm{d}} = x - \tau\ \mathrm{sign}\,(\xi)$ is a minimizer of $\tilde{x} \mapsto \langle \xi, \tilde{x} \rangle$ on $\overline{B_\tau^\infty}(x)$. Furthermore, we define $\delta \in \mathbb{R}^n$ as

$$\delta_i := -\mathrm{sign}\,(x_i^{\mathrm{d}}) \max\{\left|x_i^{\mathrm{d}}\right| - \varepsilon, 0\},$$

i.e., we have that $\mathrm{Clip}_{0,\varepsilon}(x^{\mathrm{d}}) = x^{\mathrm{d}} + \delta$. The important fact, where the choice of budget constraint matters, is that $\tilde{x} - \delta \in \overline{B_\tau^\infty}(x)$ for all $\tilde{x} \in \overline{B_\tau^\infty}(x) \cap \overline{B_\varepsilon^\infty}(0)$, since we have

$$\max\{-\varepsilon, x_i - \tau\} \le \tilde{x}_i \le \min\{\varepsilon, x_i + \tau\}$$

$$\Rightarrow \begin{cases} \delta_i = 0 & : \quad |\tilde{x}_i - \delta_i - x_i| = |\tilde{x}_i - x_i| \le \tau \\ \delta_i < 0 & : \quad x_i \le \varepsilon < x_i^{\mathrm{d}} \le x_i + \tau \\ & \quad \Rightarrow |\tilde{x}_i - \delta_i - x_i| \le \left|\varepsilon + x_i^{\mathrm{d}} - \varepsilon - x_i\right| \le \tau \\ \delta_i > 0 & : \quad |\tilde{x}_i - \delta_i - x_i| \le \tau, \quad \text{analogously to the case above.} \end{cases}$$

Now assume that there exists $\tilde{x} \in \overline{B_\varepsilon^\infty}(0) \cap \overline{B_\tau^\infty}(x)$ such that $\langle \xi, \tilde{x} \rangle < \langle \xi, x^{\mathrm{d}} + \delta \rangle$. Then we infer that

$$\langle \xi, \tilde{x} - \delta \rangle < \langle \xi, x^{\mathrm{d}} + \delta \rangle - \langle \xi, \delta \rangle = \langle \xi, x^{\mathrm{d}} \rangle$$

and therefore x^{d} is not a minimizer on $\overline{B_\tau^\infty}(x)$, which is a contradiction. Therefore, we have that

$$x^{\mathrm{d}} + \delta = \mathrm{Clip}_{0,\varepsilon}(x^{\mathrm{d}}) = \mathrm{Clip}_{0,\varepsilon}(x + \tau\ \ \mathrm{sign}\,(\xi)) \in \operatorname*{arg\,min}_{\tilde{x} \in \overline{B_\tau^\infty}(x) \cap \overline{B_\varepsilon^\infty}(0)} \langle \xi, \tilde{x} \rangle = \operatorname*{arg\,min}_{\tilde{x} \in \overline{B_\tau^\infty}(x)} E^{\mathrm{sl}}(\tilde{x}; x).$$

□

https://doi.org/10.1017/S0956792525100120 Published online by Cambridge University Press

This result shows that when we choose $p = \infty$ for the budget constraint (IFGSM) again fulfils the semi-implicit minimizing movement scheme, beyond the time restriction in Corollary 5.3.

Theorem 5.5. *We consider the space* $(\mathcal{X}, \|\cdot\|) = (\mathbb{R}^d, \|\cdot\|_\infty)$*, the energy* $E = E^{\mathrm{d}} + \chi_{\overline{B^\infty_\varepsilon}(x^0)}$*, with a continuously differentiable part* E^{d}*, which has a Lipschitz continuous gradient. Then for* $T > 0$*, there exists a* ∞*-curve of maximal slope* $u : [0, T] \to \mathbb{R}^d$*, with respect to E, and a subsequence of* $\tau_n := T/n$ *such that*

$$\left\| x^{\lceil t/\tau_{n_i} \rceil}_{\mathrm{IFGS},\tau_{n_i}} - u(t) \right\| \xrightarrow{i\to\infty} 0 \qquad \text{for all } t \in [0, T].$$

Proof. Since Lemma 5.4 yields that (IFGSM) fulfils the semi-implicit minimizing movement scheme, we can proceed similarly as in the proof of Corollary 5.3. We note that all the necessary assumptions are fulfilled, since the indicator function $\chi_{\overline{B^\infty_\varepsilon}(x^0)}$ is lower semicontinuous. □

5.2. Adversarial training and distributional adversaries

As before, we assume that the underlying spaces are finite dimensional, i.e., $\mathcal{X} = \mathbb{R}^d$, $\mathcal{Y} = \mathbb{R}^m$ with norms $\|\cdot\|_\mathcal{X}$, $\|\cdot\|_\mathcal{Y}$ and $\mathcal{P}(\mathcal{X} \times \mathcal{Y})$ denotes the space of Borel probability measures. We consider the adversarial training task, as proposed in [46, 58],

$$\inf_{h\in\mathcal{H}} \int \sup_{\tilde{x}\in\overline{B_\varepsilon}(x^0)} \ell(h(\tilde{x}), y)\, \mathrm{d}\mu(x, y), \tag{5.2}$$

where $\mu \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ denotes the data distribution and $\ell(h(\,\cdot\,), y) \in C^1(\mathcal{X} \times \mathcal{Y})$. This interpretation of adversarial learning in the distributional setting has sparked a lot of interest in recent years, see e.g., [18, 23, 66, 81, 82, 96, 97, 107]. In order to rewrite this task as a DRO problem, we equip $\mathcal{P}(\mathcal{X} \times \mathcal{Y})$ with a suitable optimal transport distance

$$D(\mu, \tilde{\mu}) := \inf_{\gamma\in\Gamma(\mu,\tilde{\mu})} \gamma - \operatorname{ess\,sup}\ c(x, y, \tilde{x}, \tilde{y}),$$

where

$$c(x, y, \tilde{x}, \tilde{y}) := \begin{cases} \|x - \tilde{x}\|_\mathcal{X} & \text{if } y = \tilde{y}, \\ +\infty & \text{if } y \neq \tilde{y}, \end{cases} \tag{5.3}$$

and $\Gamma(\mu, \tilde{\mu})$ denotes the set of transport plans between μ and $\tilde{\mu}$. Notably, the extended distance c is not the one naturally generated by the norms of the underlying Banach spaces $\mathcal{X}$ and $\mathcal{Y}$. Nonetheless, c is compatible with respect to $\|\cdot\|_\mathcal{X} + \|\cdot\|_\mathcal{Y}$ in the sense that

$$\liminf_{n\to\infty} c(x_n, y_n, \tilde{x}_n, \tilde{y}_n) \geq c(x, y, \tilde{x}, \tilde{y}),$$
$$\forall (x, y), (\tilde{x}, \tilde{y}) \in (\mathcal{X} \times \mathcal{Y}) : (x_n, y_n) \to (y, x), (\tilde{x}_n, \tilde{y}_n) \to (\tilde{x}, \tilde{y}) \text{ w.r.t. } \|\cdot\|_\mathcal{X} + \|\cdot\|_\mathcal{Y},$$

compare [63, Eq. (1)]. This ensures that, as we equip $\mathcal{P}(\mathcal{X} \times \mathcal{Y})$ with D, it is a well-defined extended distance, see [63, section 2.6]. The cost functional c was similarly employed in [13, 18]; furthermore, a similar setup was considered in [97].

Remark 5.6. Assume that $\gamma \in \Gamma(\mu, \tilde{\mu})$ is a coupling, i.e., $\gamma \in \mathcal{P}(\mathcal{Z} \times \mathcal{Z})$, where $\mathcal{Z} = \mathcal{X} \times \mathcal{Y}$, with $\gamma - \operatorname{ess\,sup} c(x, y, \tilde{x}, \tilde{y}) < \infty$. Then we have that for every measurable set $A \subset \mathcal{Y}$,

$$\gamma(\mathcal{X} \times A \times \mathcal{Z}) = \gamma(\mathcal{X} \times A \times \mathcal{X} \times A) = \gamma(\mathcal{Z} \times \mathcal{X} \times A),$$

which we see by contradiction: assume there exists a measurable set $A \subset \mathcal{Y}$ s.t., for $B := \mathcal{X} \times A \times \mathcal{X} \times (\mathcal{Y} \setminus A)$ we have $\gamma(B) > 0$. Then we know that $c(x, y, \tilde{x}, \tilde{y}) = +\infty$ for all $(x, y, \tilde{x}, \tilde{y}) \in B$ and since $\gamma(B) > 0$ this yields that

$$\gamma - \operatorname{ess\,sup} c(x, y, \tilde{x}, \tilde{y}) \geq \gamma - \operatorname*{ess\,sup}_B c(x, y, \tilde{x}, \tilde{y}) = +\infty.$$

https://doi.org/10.1017/S0956792525100120 Published online by Cambridge University Press

The other identity can be proven analogously. Therefore, if $D(\mu, \tilde{\mu}) < \infty$ we know that there exists a coupling γ fulfilling the above assumption and thus for every measurable set $A \subset \mathcal{Y}$ we obtain

$$\mu(\mathcal{X} \times A) = \int_{\mathcal{X}\times A\times \mathcal{Z}} \mathrm{d}\gamma = \int_{\mathcal{Z}\times \mathcal{X}\times A} \mathrm{d}\gamma = \tilde{\mu}(\mathcal{X} \times A).$$

If we now consider a disintegration of μ and $\tilde{\mu}$ along the $\mathcal{X}$-axis, i.e., we obtain $\mathrm{d}\mu = \mathrm{d}\mu_y \mathrm{d}\nu(y), \mathrm{d}\tilde{\mu} = \mathrm{d}\tilde{\mu}_y \mathrm{d}\tilde{\nu}(y)$, with

$$\nu(A) = \mu((\pi^y)^{-1}(A)) = \mu(\mathcal{X} \times A) = \tilde{\mu}(\mathcal{X} \times A) = \tilde{\mu}((\pi^y)^{-1}(A)) = \tilde{\nu}(A)$$

for every measurable $A \subset \mathcal{Y}$, where $\pi^y(x, y) := y$ is the projection onto the $\mathcal{Y}$-component.

The transport distance D behaves like the ∞-Wasserstein distance in the $\mathcal{X}$-direction (compare section section 4) and penalizes movement of mass into the $\mathcal{Y}$-direction, such that no movement in $\mathcal{Y}$ can occur when $D(\mu, \tilde{\mu})$ is finite (see Remark 5.6). Thus, all calculations done in section 4 apply with minor adaptation to this case. We only state corresponding lemmas and theorems, while adapted proofs can be found in Appendix G. The first property we prove in this section is that the adversarial training problem (5.2) is equivalent to the distributional robust optimization problem, (DRO). Note that now we need to consider a potential defined on the space $\mathcal{X} \times \mathcal{Y}$, namely $E(x, y) := -\ell(h(x), y)$, where the label y is now also a variable argument.

Corollary 5.7. *It holds that*

$$\int \max_{\tilde{x}\in\overline{B_\varepsilon}(\tilde{x})} \ell(h(x), y)\, d\mu(x, y) = \max_{\tilde{\mu}\,:\, D(\tilde{\mu},\mu)\le\varepsilon} \int \ell(h(x), y) d\tilde{\mu}(x, y) \tag{5.4}$$

where the maximizing argument is given by $\mu_{\max} = (r_\varepsilon)_{\#}\mu$, *with* $r_\varepsilon : \mathcal{X} \times \mathcal{Y} \to \mathcal{X}$ *being a* $\mathcal{B}(\mathcal{X} \times \mathcal{Y})$*-measurable selector from Lemma G.1*

Proof. We employ the $\mathcal{B}(\mathcal{X} \times \mathcal{Y})$-measurable selector r_ε, from Lemma G.1 and compute

$$\begin{aligned}
\int \max_{\tilde{x}\in\overline{B_\varepsilon}(x)} \ell(h(\tilde{x}), y)\ \mathrm{d}\mu(x, y) &= \int \max_{(\tilde{x},\tilde{y}):c(x,y,\tilde{x},\tilde{y})\le\varepsilon} -E(\tilde{x}, \tilde{y})\ \mathrm{d}\mu(x, y) = \int -E(r_\varepsilon(x, y))\ \mathrm{d}\mu(x, y) \\
&= -\int E(x, y) d(r_\varepsilon)_{\#}\mu(x, y) \overset{(i)}{=} -\min_{\tilde{\mu}\,:\, D(\mu,\tilde{\mu})\le\varepsilon} \int E(x, y)\, d\tilde{\mu}(x, y) \\
&= \max_{\tilde{\mu}\,:\, D(\tilde{\mu},\mu)\le\varepsilon} \int \ell(h(x), y) d\tilde{\mu}(x, y),
\end{aligned}$$

where in (i) we employ (G.1). □

Remark 5.8. In other works considering distributional adversarial attacks, for example [81, 82], the well-definedness of the expressions in Corollary 5.7 is not always ensured. In [18], this was resolved by considering open balls for the budget constraint. However, due to our assumption that $\ell(h(\,\cdot\,), y) \in C^1(\mathcal{X} \times \mathcal{Y})$, we do not encounter similar measurability issues, as shown in [67].

For the main result in this section, we now consider the energy defined via the potential defined on $\mathcal{X} \times \mathcal{Y}$, i.e.,

$$\mathcal{E}(\mu) := \int E(x, y)\, \mathrm{d}\mu(x, y) = \int -\ell(h(x), y)\, \mathrm{d}\mu(x, y),$$

where the underlying extended metric space is chosen as $\mathcal{D} = (\mathcal{P}_\infty(\mathcal{X} \times \mathcal{Y}), c)$, with $\mathcal{P}_\infty(\mathcal{X} \times \mathcal{Y})$ denoting the subset of Borel probability measures with bounded support in $\mathcal{X}$- and $\mathcal{Y}$-direction.

Remark 5.9. Theorem 4.7 also holds for extended distances, i.e., distances which take values in $[0, +\infty]$, compare [63, Theorem 3.1]. The distance $c(\cdot, \cdot)$ introduced in (5.3) is such an extended distance. For this particular choice of extended distance, the measure

$$\eta \in \mathcal{P}\left(C(0, T; ((\mathcal{X} \times \mathcal{Y}), \|\cdot\|_{\mathcal{X}} + \|\cdot\|_{\mathcal{Y}}))\right.$$

https://doi.org/10.1017/S0956792525100120 Published online by Cambridge University Press

is concentrated on $AC^\infty(0,T;((\mathcal{X}\times\mathcal{Y}),c)) = AC^\infty(0,T;(\mathcal{X},\|\cdot\|_\mathcal{X})\times\mathcal{Y})$. Notice that the continuous curves are continuous w.r.t. $\|\cdot\|_\mathcal{X}+\|\cdot\|_\mathcal{Y}$, while absolute continuity is w.r.t. $c(\cdot,\cdot)$ (compare [63, section 2.3.]).

The theorem below is a variant of Theorem 4.18 for the adversarial setting. Namely, we show that ∞-curves of maximal slope that are used to solve (DRO) can be characterized by employing a representing measure η on $C(0,T;\mathcal{X}\times\mathcal{Y})$, where η-a.e. curve fulfils the differential inclusion w.r.t. the potential E. Here, we enforce the condition $D(\mu,\tilde{\mu})\le\varepsilon$, by only considering the evolution until time $T=\epsilon$.

Theorem 5.10. *For $T=\varepsilon$, let $\mu\in \mathrm{AC}^\infty(0,T;\mathcal{D})$ with η from Theorem 4.7. Let further $\mathcal{E}\circ\mu$ be for a.e. $t\in[0,T]$ equal to a non-increasing map $\psi:[0,T]\to\mathbb{R}$.*

Then the following statements are equivalent:

(i) $|\mu'|(t)\le 1$ *and* $\psi'(t)\le -|\partial\mathcal{E}|(u(t))$ *for a.e.* $t\in(0,T)$.

(ii) For η-a.e. curve $u\in C(0,T;\mathcal{X}\times\mathcal{Y})$ it holds, that $E\circ u$ is for a.e. $t\in(0,T)$ equal to a non-increasing map ψ_u:$[0,T]\to\mathbb{R}$ and

$$u'(t)\ \in\ (\partial\|\cdot\|_{\mathcal{X}^*}(-\nabla_x E(u(t))),0),\quad \text{for a.e. } t\in(0,T).$$

Proof. Step 1: (i) $\Longrightarrow$ (ii).

By Lemma G.2, we know that $|\partial\mathcal{E}|$ is a strong upper gradient such that by Remark 2.3 μ_t satisfies the energy dissipation equality (2.10). Similar to Theorem 4.18, we estimate

$$\begin{aligned}\mathcal{E}(\mu_0)-\mathcal{E}(\mu_T) &= \int_{C(0,T;\mathcal{X})} E(u(0))-E(u(T))\,\mathrm{d}\eta(u)\\ &\le \int_{C(0,T;\mathcal{X})}\int_0^T \|\nabla_x E(u(r))\|_{\mathcal{X}^*}|u'|(t)\,\mathrm{d}r\,\mathrm{d}\eta(u)\\ &\le \int_{C(0,T;\mathcal{X})}\int_0^T \|\nabla_x E(u(r))\|_{\mathcal{X}^*}\,\mathrm{d}r\,\mathrm{d}\eta(u)\\ &= \int_0^T |\partial\mathcal{E}|(\mu(r))\,\mathrm{d}r = \mathcal{E}(\mu_0)-\mathcal{E}(\mu_T).\end{aligned}\tag{5.5}$$

and observe that this equality can only hold if for η.a.e. u

$$E(u(s))-E(u(t)) = \int_s^t \|\nabla_x E(u(r))\|_{\mathcal{X}^*}|u'|(t)\,\mathrm{d}r = \int_s^t \|\nabla_x E(u(r))\|_{\mathcal{X}^*}\,\mathrm{d}r \quad \text{for all } 0\le s<t\le T.$$

and thus $\psi_u := E\circ u$ is a non-increasing absolutely continuous map for η-a.e. u.

We use Lemma E.4 and $\||u'|(t)\|_{L^\infty(\eta)} = |\mu'|(t)\le 1$ for a.e. $t\in(0,T)$ to infer that for η-a.e. curve $u\in C(0,T;\mathcal{X}\times\mathcal{Y})$ it holds,

$$\left|u'\right|(t)\le 1\quad \text{for a.e. } t\in(0,T).$$

Denoting by $(u'(t))_x$ and $(u'(t))_y$ the $\mathcal{X}$ and $\mathcal{Y}$ corresponding parts of the derivative $u'(t)$ and keeping Lemma 3.7 in mind we obtain for η-a.e. curve $u\in C(0,T;\mathcal{X}\times\mathcal{Y})$

$$\begin{aligned}\langle\nabla_x E(x,y),(u'(t))_x\rangle &= \langle\nabla E(x,y),u'(t)\rangle\\ &= (E\circ u)'(t)\\ &= -\|\nabla_x E(u(r))\|_{\mathcal{X}^*} = -\,\|\nabla_x E(u(r))\|_{\mathcal{X}^*} - \chi_{\overline{B_1}}((u'(t))_x)\end{aligned}$$

for a.e. $t\in(0,T)$. Using the equivalence of Item 3 and Item 1 we obtain

$$u'(t)\ \in\ (\partial\|\cdot\|_{\mathcal{X}^*}(-\nabla_x E(x,y)),0)\quad \text{for a.e. } t\in(0,T)$$

for η-a.e. curve u.

https://doi.org/10.1017/S0956792525100120 Published online by Cambridge University Press

Step 2: (ii) $\Longrightarrow$ (i).
For η-a.e. $u \in C(0,T;\mathcal{X}\times\mathcal{Y})$ we know by Remark 2.3 that the energy dissipation equality

$$E(u(s)) - E(u(t)) = \int_s^t \|\nabla_x E(u(r))\|_{\mathcal{X}^*}\,\mathrm{d}r \quad \text{for all } 0 \le s < t \le T.$$

holds. In particular, it is absolutely continuous such that Remark 1.3 applies. We calculate,

$$\begin{aligned}\mathcal{E}(\mu_t) - \mathcal{E}(\mu_s) &= \int_{C(0,T;\mathcal{X})} E(u(t)) - E(u(s))\,\mathrm{d}\eta(u) = \int_{C(0,T;\mathcal{X})}\int_s^t (E\circ u)'(r)\,\mathrm{d}r\,\mathrm{d}\eta(u)\\ &= \int_{C(0,T;\mathcal{X})}\int_s^t \langle \nabla_x E(u(r)), (u'(r))_x\rangle\,\mathrm{d}r\,\mathrm{d}\eta(u) \overset{(i)}{=} \int_{C(0,T;\mathcal{X})}\int_s^t -\|\nabla_x E(u(r))\|_{\mathcal{X}^*}\,\mathrm{d}r\,\mathrm{d}\eta(u)\\ &= \int_{C(0,T;\mathcal{X})}\int_s^t -\|\nabla_x E(u(r))\|_{\mathcal{X}^*}\,\mathrm{d}\eta(u)\,\mathrm{d}r \overset{(ii)}{=} \int_s^t -|\partial\mathcal{E}|(\mu_r)\,\mathrm{d}r,\end{aligned}$$

Where for (i) we use the equivalence of Item 3 and Item 1, while for (ii) we use Lemma G.2. This implies $\mathcal{E}\circ\mu_t$ is monotone non-increasing and $(\mathcal{E}\circ\mu)'(t) \le -|\partial\mathcal{E}|(\mu_t)$ for a.e. $t\in(0,T)$. Further, by Theorem 4.7, we have

$$|\mu'|(t) = \eta(u) - \operatorname{ess\,sup}|u'|(t) = \eta(u) - \operatorname{ess\,sup}\|u'(t)\|_{\mathcal{X}} \le 1,$$

since all elements in $\partial\|\cdot\|_{\mathcal{X}^*}(-\nabla_x E(x,y))$ have norm smaller than 1. □

6. Conclusion and outlook

In this work, we considered the limit case $p\to\infty$ of the well-known p-curves of maximum slope, which yield a versatile gradient flow framework in metric spaces, [2]. In the abstract setting, we proved existence by employing the minimizing movement scheme, adapted to the case $p=\infty$. Assuming that the underlying space is Banach, we were able to characterize ∞-curves of maximum slope via differential inclusions. Furthermore, we also demonstrated the convergence of a semi-implicit scheme to the continuum flow. This insight constitutes the interface to the field of adversarial attacks. Namely, we showed that the well-known FGSM, and its iterative variant, correspond to the semi-implicit scheme and therefore converge to the flow, when sending the step size to zero. More generally, this result holds true for a whole class of normalized gradient descent algorithms. Furthermore, we also considered Wasserstein gradient flows, where we first used the theory developed in [63] to derive an alternative characterization of absolutely continuous curves via the continuity equation. As our main result in this section, we prove that being an ∞-curve of maximal slope is equivalent to the existence of a representing measure on the space of continuous curves, where almost every curve, fulfils a differential inclusion on the underlying Banach space. This finally allowed us to generate distributional adversaries, in an adapted ∞-Wasserstein distance, via curves of maximum slope. Similar to section 5, we could also consider the energy

$$\mathcal{E}(\mu) := \int_{\mathcal{X}\times\mathcal{Y}} E(x,y)\,\mathrm{d}\mu + \chi_{B^D_\epsilon(\mu_0)}(\mu)$$

to generate distributional adversarial attacks. We strongly suspect that corresponding ∞-curves of maximal slope in $\mathcal{D}$ would take the following form: Let $\mu\in AC^\infty(0,T;\mathcal{X})$ be a ∞-curve of maximal slope and η its corresponding probability measure over the space $C(0,T;\mathcal{X}\times\mathcal{Y})$, then for η-a.e. $u\in C(0,T;\mathcal{X}\times\mathcal{Y})$

$$u'(t) \;\in\; (\partial\|\cdot\|_{\mathcal{X}^*}(-\nabla_x E_{u_0}(u(t))), 0), \quad \text{for a.e. } t\in(0,T),$$

where

$$E_{u_0}(x,y) = E(x,y) + \chi_{B_\epsilon((u_0)_x)}(x).$$

https://doi.org/10.1017/S0956792525100120 Published online by Cambridge University Press

In [105], the authors suggested to combine *FGSM* with stochastic elements. They proposed to use a single step

$$\sigma \sim \text{Uniform}\left(\overline{B_\epsilon^\infty}(x_0)\right),$$
$$x_{\frac{1}{2}} = x_0 + \sigma,$$
$$x_1 = \text{Clip}_{0,\varepsilon}\left(x_{\frac{1}{2}} + \text{sign}\left(\nabla \ell(h(x_{\frac{1}{2}}), y)\right)\right).$$

This is reminiscent of the classical Langevin algorithm, therefore, it would be interesting if this stochasticity could be incorporated into our framework.

Acknowledgements. MB, TR and LW acknowledge support from DESY (Hamburg, Germany), a member of the Helmholtz Association HGF. This research was supported in part through the Maxwell computational resources operated at Deutsches Elektronen-Synchrotron DESY, Hamburg, Germany. TR further wants to thank Samira Kabri for many insightful discussions. Parts of this study were carried out while LW and TR were affiliated with the Friedrich-Alexander-Universität Erlangen-Nürnberg.

Financial support. MB and TR acknowledge funding by the German Ministry of Science and Technology (BMBF) under grant agreement No. 01IS24072A (COMFORT). MB and LW acknowledge funding from the German Research Foundation, project BU 2327/20-1.

Competing interests. The authors declare none.

References

[1] Ambrosio, L. (1990) Metric space valued functions of bounded variation. *Ann. Scuola Norm.-Sci.* **17**(3), 439–478.
[2] Ambrosio, L., Gigli, N. & Savaré, G. (2005) *Gradient Flows: In Metric Spaces and in the Space of Probability Measures*, Springer Science & Business Media.
[3] Armstrong, S. N. & Smart, C. K. (2010) An easy proof of Jensen's theorem on the uniqueness of infinity harmonic functions. *Calc. Var. Part. Differ.* **37**, 381–384.
[4] Bailo, R., Barbaro, A., Gomes, S. N., et al. (2024) CBX: Python and Julia packages for consensus-based interacting particle methods. arXiv: 2403.14470 [math.OC].
[5] Balles, L., Pedregosa, F. & Roux, N. L. (2020) The geometry of sign gradient descent. arXiv: 2002.08056.
[6] Barbu, V. & Precupanu, T. (2012) *Convexity and Optimization in Banach Spaces*, Springer Science & Business Media.
[7] Billingsley, P. (2013) *Convergence of Probability Measures*, John Wiley & Sons.
[8] Blanchard, P. & Brüning, E. (2015) *Mathematical Methods in Physics: Distributions, Hilbert Space Operators, Variational Methods, and Applications in Quantum Physics*, Vol. **69**, Birkhäuser..
[9] Boltzmann, L. (1868) Studien über das gleichgewicht der lebenden kraft. *Wissensch. Abhand.* **1**, 49–96
[10] Brasco, L. & Santambrogio, F. (2011) An equivalent path functional formulation of branched transportation problems". *Discrete Contin. Dyn. Syst.* **29**, 845–871.
[11] Brendel, W., Rauber, J. & Bethge, M. (2017) Decision-based adversarial attacks: Reliable attacks against black-box machine learning models. arXiv preprint arXiv: 1712.04248.
[12] Brendel, W., Rauber, J., Kümmerer, M., Ustyuzhaninov, I. & Bethge, M. (2019) Accurate, reliable and fast robustness evaluation. *Adv. Neural Inf. Process. Syst.* **32**.
[13] Bui, T. A., Le, T., Tran, Q., Zhao, H. & Phung, D. (2022) A unified wasserstein distributional robustness framework for adversarial training. arXiv preprint arXiv: 2202.13437.
[14] Bungert, L. & Burger, M. (2020) Asymptotic profiles of nonlinear homogeneous evolution equations of gradient flow type. *J. Evol. Equ.* **20**(3), 1061–1092.
[15] Bungert, L., Burger, M., Chambolle, A. & Novaga, M. (2021) Nonlinear spectral decompositions by gradient flows of one-homogeneous functionals. *Anal. PDE* **14**(3), 823–860.
[16] Bungert, L., Calder, J. & Roith, T. (2023) Uniform convergence rates for Lipschitz learning on graphs. *IMA J. Numer. Anal.* **43**(4), 2445–2495.
[17] Bungert, L., Calder, J. & Roith, T. (2024) Ratio convergence rates for Euclidean first-passage percolation: Applications to the graph infinity Laplacian. *Ann. Appl. Probab.* arXiv: 2210.09023 (math.PR).
[18] Bungert, L., Trillos, N. García & Murray, R. (2023) The geometry of adversarial training in binary classification. *Inf. Inference: J, IMA* **12**(2), 921–968.
[19] Bungert, L., Hoffmann, F., Kim, D. Y. & Roith, T.(2025) MirrorCBO: A consensus-based optimization method in the spirit of mirror descent. arXiv preprint arXiv: 2501.12189.
[20] Bungert, L., Laux, T. & Stinson, K. (2024) A mean curvature flow arising in adversarial training. arXiv: 2404.14402.
[21] Bungert, L., Bungert, R., Roith, T., Schwinn, L., Tenbrinck, D. (2021) CLIP Cheap Lipschitz training of neural networks. In: *Scale Space and Variational Methods in Computer Vision: 8th International Conference, SSVM 2021, Proceedings*. Springer, pp. 307–319.
[22] Bungert, L. & Stinson, K. (2024) Gamma-convergence of a nonlocal perimeter arising in adversarial machine learning. *Calc. Var. Part. Differ.* **63**(5), 114.

https://doi.org/10.1017/S0956792525100120 Published online by Cambridge University Press

[23] Bungert, L., Trillos, N. G., Jacobs, M., D. McKenzie & Wang, Q. (2023) It begins with a boundary: A geometric view on probabilistically robust learning. arXiv: 2305.18779.
[24] Charalambos, K. C. B., Aliprantis, D. (2006) *Infinite Dimensional Analysis. A Hitchhiker's Guide*, Springer.
[25] Chen, F. & Ren, W. (2020) Sign projected gradient flow: A continuous-time approach to convex optimization with linear equality constraints. *Automatica* **120**, 109156. ISSN: 0005-1098.
[26] Chzhen, E. & Schechtman, S. (2023) SignSVRG: Fixing SignSGD via variance reduction. arXiv: 2305.13187 [math.OC].
[27] Cortés, J. (2006) Finite-time convergent gradient flows with applications to network consensus. *Automatica* **42**(11), 1993–2000.
[28] Cutkosky, A. & Mehta, H. (2020) Momentum improves normalized sgd. In: *International Conference on Machine Learning*, PMLR, pp. 2260–2268.
[29] Cybenko, G., O'Leary, D. P. & Rissanen, J. (1998) *The Mathematics of Information Coding, Extraction and Distribution*, Vol. **107**, Springer Science & Business Media.
[30] Dacorogna, B. (2007) *Direct Methods in the Calculus of Variations*, vol. **78**, Springer Science & Business Media.
[31] De Giorgi, E., Marino, A. & Tosques, M. (1980) Problemi di evoluzione in spazi metrici e curve di massima pendenza". *Atti Accad. Naz. Lincei Classe Sci. Fisiche Mat. Nat. Rend.* **68**, 180–187.
[32] Degiovanni, M., Marino, A. & Tosques, M. (1985) Evolution equations with lack of convexity. *Nonlinear Anal.: Theory, Methods Appl.* **9**(12), 1401–1443.
[33] Dellacherie, C. & Meyer, P.-A. (1978) *Probabilities and potential*, Vol. **29**, North-Holland Mathematics Studies.
[34] Diestel, J. & Uhl, J.JrVector measures (American mathematical society, Providence, RI. 1977). *With Foreword BJ Pettis, Math. Surv.* **15**, 27.
[35] Dong, Y., Liao, F., Pang, T., Su, H., Zhu, J., Hu, X. (2018) Boosting adversarial attacks with momentum. *Proc. IEEE Confer. Comput. Vis. Pattern Recogn.* 9185–9193.
[36] Duchi, J., Shalev-Shwartz, S., Singer, Y. & Chandra, T. (2008) Efficient projections onto the l 1-ball for learning in high dimensions. (*Proceedings of the 25th International Conference on Machine Learning*, pp. 272–279.
[37] Edgar, G. (1977) Measurability in a Banach space. *Indiana Univ. Math. J.* **26**(4), 663–677. ISSN: 0022-2518.
[38] Euler, L. (1794) *Institutiones calculi integralis*, Vol. **4**, Academia Imperialis Scientiarum.
[39] Fenchel, W. & Blackett, D. W. (1953) *Convex Cones, Sets, and Functions*, Department of Mathematics, Logistics Research Project, Princeton University.
[40] Fleißner, F. (2019) Γ-convergence and relaxations for gradient flows in metric spaces: A minimizing movement approach. *ESAIM: Control, Optim. Calc. Var.* **25**, 28.
[41] Fourier, J. (1824) Histoire de l'Académie, partie mathématique, 1824. *Mém. l'Acad. Sci. l'Inst. France* **7**, 38.
[42] Fukushima, K. (1980) Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biol. Cybern.* **36**(4), 193–202
[43] Giorgi, E. D. (1993) New problems on minimizing movements. *Bound. Value Probl. PDEs Appl*. Masson, pp. 81–89.
[44] Givens, C. R. & Shortt, R. M. (1984) A class of Wasserstein metrics for probability distributions. *Mich Math J* **31**(2), 231–240.
[45] Good, I. J. (1952) Rational decisions. *J. R. Stat. Soc.: Ser. B (Methodol.)* **14**(1), 107–114.
[46] Goodfellow, I. J., Shlens, J. & Szegedy, C. (2014) Explaining and harnessing adversarial examples. arXiv preprint arXiv: 1412.6572.
[47] Gouk, H., Frank, E., Pfahringer, B. & Cree, M. J. (2020) Regularisation of neural networks by enforcing Lipschitz continuity. *Machine Learning* **110**, 1–24..
[48] Gruntkowska, K., Li, H., Rane, A. & Richtárik, P. (2025) The ball-proximal (="broximal") point method: A new algorithm, convergence theory, and applications. arXiv preprint arXiv: 2502.02002.
[49] Hausdorff, F. (1919) Über halbstetige funktionen und deren verallgemeinerung. *Math Z.* **5**(3), 292–309.
[50] Hazan, E., Levy, K. & Shalev-Shwartz, S. (2015) Beyond convexity: Stochastic quasi-convex optimization". *Adv. Neural Inf. Process. Syst.* **28**.
[51] Hendrycks, D. & Gimpel, K. (2023) *Gaussian error linear units (GELUs)*. eprint: 1606.08415 (cs.LG).
[52] Howard, S. T. (2022) A 'Hello world' for pyTorch. https://seanhoward.me/blog/2022/hello_world_pytorch/. Accessed 05 Jun 2024.
[53] Ilyas, A., Engstrom, L., Athalye, A. & Lin, J. (2018) Black-box adversarial attacks with limited queries and information". In: *International conference on machine learning*, PMLR, pp. 2137–2146.
[54] Ioffe, S. & Szegedy, C. (2015) Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: *International Conference on Machine Learning*, PMLR, pp. 448–456.
[55] Jang, U., Wu, X. & Jha, S. (2017) Objective metrics and gradient descent algorithms for adversarial examples in machine learning". In: *Proceedings of the 33rd Annual Computer Security Applications Conference*, 262–277.
[56] Kingma, D. P. & Ba, J. (2017) Adam: A method for stochastic optimization. arXiv: 1412.6980.
[57] Krishnan, V., Makdah, A., AlRahman, A. & Pasqualetti, F. (2020) Lipschitz bounds and provably robust training by laplacian smoothing". *Adv. Neural Inf. Process. Syst.* **33**, 10924–10935.
[58] Kurakin, A., Goodfellow, I. & Bengio, S.(2016) Adversarial machine learning at scale. arXiv preprint arXiv: 1611.01236.
[59] Kurakin, A., Goodfellow, I. J. & Bengio, S. (2018) Adversarial examples in the physical world. In: *Artificial Intelligence Safety and Security*. Chapman and Hall/CRC, pp. 99–112.
[60] Levy, K. Y.(2016) The power of normalization: Faster evasion of saddle points. arXiv preprint arXiv: 1611.04831.
[61] Li, X., Lin, K.-Y., Li, L., Hong, Y. & Chen, J.(2023) On faster convergence of scaled sign gradient descent. *IEEE Trans. Industr. Informat*, pp. 1732–1741.

https://doi.org/10.1017/S0956792525100120 Published online by Cambridge University Press

[62] Lisini, S. (2007) Characterization of absolutely continuous curves in wasserstein spaces. *Calc. Var. Part. Differ. Equ.* **28**, 85–120.

[63] Lisini, S. (2014) Absolutely continuous curves in extended wasserstein-orlicz spaces. arXiv: 1402.7328.

[64] Madry, A., Makelov, A., Schmidt, L., Tsipras, D. & Vladu, A.(2018) Towards deep learning models resistant to adversarial attacks. In: *International Conference on Learning Representations*.

[65] Marino, A., Saccon, C. & Tosques, M. (1989) Curves of maximal slope and parabolic variational inequalities on non-convex constraints. *Ann. Scuola Norm.-Sci.* **16**(2), 281–330.

[66] Mehrabi, M., Javanmard, A., Rossi, R. A., Rao, A. & Mai, T. (2021) Fundamental tradeoffs in distributionally adversarial training". In: *International Conference on Machine Learning*, PMLR, pp. 7544–7554.

[67] Meunier, L., Scetbon, M., Pinot, R. B., Atif, J. & Chevaleyre, Y. (2021) Mixed nash equilibria in the adversarial examples game". In: *International Conference on Machine Learning*, PMLR, pp. 7677–7687.

[68] Mielke, A., Rossi, R. & Savaré, G. (2012) Variational convergence of gradient flows and rate-independent evolutions in metric spaces. *Milan J. Math.* **80**, 381–410.

[69] Mielke, A., Rossi, R. & Savaré, G. (2013) Nonsmooth analysis of doubly nonlinear evolution equations. *Calc. Vari. Part. Differ. Equ.* **46**, 253–310.

[70] Mohammadi, A. & Janaideh, M. Al (2023) Sign gradient descent algorithms for kinetostatic protein folding. In: *International Conference on Manipulation, Automation and Robotics at Small Scales (MARSS)*, IEEE, pp. 1–6.

[71] Moosavi-Dezfooli, S.-M., Fawzi, A. & Frossard, P. (2016) Deepfool: a simple and accurate method to fool deep neural networks". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2574–2582.

[72] Moreau, J.-J. (1965) Proximité et dualité dans un espace hilbertien. *Bull. Soc. Math. France* **93**, 273–299.

[73] Morrison, T. J. (2011) *Functional Analysis: An Introduction to Banach Space Theory*, John Wiley & Sons.

[74] Moulay, E., Léchappé, V. & Plestan, F. (2019) Properties of the sign gradient descent algorithms. *Inf. Sci.* **492**, 29–39. ISSN:0020-0255.

[75] Murray, R., Swenson, B. & Kar, S. (2019) Revisiting normalized gradient descent: Fast evasion of saddle points. *IEEE Trans. Autom. Control* **64**(11), 4818–4824.

[76] Paszke, A., Gross, S., Chintala, S., et al. (2017) Automatic differentiation in PyTorch. Conference on Neural Information Processing Systems (NeurIPS 2017), Long Beach, CA, USA.

[77] Pauli, P., Koch, A., Berberich, J., Kohler, P. & Allgöwer, F. (2021) Training robust neural networks using Lipschitz bounds. *IEEE Control Syst. Lett.* **6**, 121–126.

[78] Pedregosa, F., et al. (2011) Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830.

[79] Pinnau, R., Totzeck, C., Tse, O. & Martin, S. (2017) A consensus-based model for global optimization and its mean-field limit. *Math. Models Methods Appl. Sci.* **27**(01), 183–204.

[80] Pintor, M., Roli, F., Brendel, W. & Biggio, B. (2021) Fast minimum-norm adversarial attacks through adaptive norm constraints. *Adv. Neural Inf. Process. Syst.* **34**, 20052–20062.

[81] Pydi, M. S. & Jog, V. (2020) Adversarial risk via optimal transport and optimal couplings. In: *International Conference on Machine Learning*, pp.7814–7823.

[82] Pydi, M. S. & Jog, V. (2021) The many faces of adversarial risk. *Adv. Neural Inf. Process. Syst.* **34**, 10000–10012.

[83] Riedmiller, M. & Braun, H. (1993) A direct adaptive method for faster backpropagation learning: the RPROP algorithm. In: *IEEE International Conference on Neural Networks*, vol.**1**, pp. 586–591.

[84] Rockafellar, R. (1970) On the maximal monotonicity of subdifferential mappings. *Pac. J. Math.* **33**(1), 209–216.

[85] Roith, T. & Bungert, L. (2023) Continuum limit of Lipschitz learning on graphs. *Found. Comput. Math.* **23**(2), 393–431.

[86] Rosenblatt, F. (1958) The perceptron: A probabilistic model for information storage and organization in the brain. *Psychol. Rev.* **65**(6), 386.

[87] Rossi, R., Mielke, A. & Savaré, G. (2008) A metric approach to a class of doubly nonlinear evolution equations and applications". *Ann. Della Scuola Norm. Super. Pisa-Classe Sci.* **7**, 197–169.

[88] Roth, K., Kilcher, Y. & Hofmann, T.(2019) Adversarial training is a form of data-dependent operator norm regularization. In: NeurIPS.

[89] Ryan, Raymond A. (2002) *Introduction to Tensor Products of Banach Spaces*, Vol. **73**, Springer.

[90] Saks, S. (1937) Theory of the integral, Second revised edition, English translation by L. C. Young, With two additional notes by Stefan Banach, Monografie Mattemaztyczen Tom. 7, Hafner Publishing Company, New York.

[91] Santambrogio, F. (2015) *Optimal Transport for Applied Mathematicians*, Birkhäuser, Cham.

[92] Schaefer, H. H. & Wolff, M. P. (1999) *Topological Vector Spaces*, Springer, New York. ISBN: 9781461214687.

[93] Schuster, T., Kaltenbacher, B., Hofmann, B. & Kazimierski, K. S. (2012) *Regularization methods in banach spaces*, De Gruyter, Berlin, Boston. ISBN:9783110255720.

[94] Shafahi, A., Najibi, M., Ghiasi, M. A., et al.(2019) Adversarial training for free!. *Adv. Neural Inf. Process. Syst.* **32**.

[95] Sierksma, G. & Zwols, Y. (2015) *Linear and Integer Optimization: Theory and Practice*, CRC Press.

[96] Sinha, A., Namkoong, H., Volpi, R. & Duchi, J. (2017) Certifying some distributional robustness with principled adversarial training. arXiv preprint arXiv: 1710.10571.

[97] Staib, M. & Jegelka, S. (2017) Distributionally robust deep learning as a generalization of adversarial training". *NIPS Workshop Mach. Learn. Comput. Secur.* **3**, 4.

[98] Stefanelli, U. (2022) A new minimizing-movements scheme for curves of maximal slope". *ESAIM: Control, Optim. Calc. Var.* **28**, 59.

[99] Stepanov, E. & Trevisan, D. (2017) Three superposition principles: Currents, continuity equations and curves of measures. *J. Funct. Anal.* **272**(3), 1044–1103.

https://doi.org/10.1017/S0956792525100120 Published online by Cambridge University Press

[100] Suzuki, Y., Yano, H., Raymond, R. & Yamamoto, N. (2021) Normalized gradient descent for variational quantum algorithms". In: *IEEE International Conference on Quantum Computing and Engineering (QCE)*, IEEE, pp. 1–9.
[101] Szegedy, C., Zaremba, W., Sutskever, I., et al. (2013) Intriguing properties of neural networks. arXiv preprint arXiv: 1312.6199.
[102] Tao, T. (2011) *An Introduction to Measure Theory*, Vol. **126**, American Mathematical Society.
[103] Turan, B., Uribe, C. A., Wai, H.-T. & Alizadeh, M. (2021) On robustness of the normalized subgradient method with randomly corrupted subgradients. In: *American Control Conference (ACC)*, IEEE, pp. 965–971.
[104] Villani, C., et al. (2009) *Optimal Transport: Old and New*, Vol. **338**, Springer.
[105] Wong, E., Rice, L. & Kolter, J. Z. (2020) Fast is better than free: Revisiting adversarial training. arXiv: 2001.03994.
[106] Zhang, H., Hui, Q., Moulay, E. & Coirault, P. (2020) Sign gradient descent method based bat searching algorithm with application to the economic load dispatch problem". In: *59th IEEE Conference on Decision and Control (CDC)*, IEEE, pp. 1140–1145.
[107] Zheng, T., Chen, C. & Ren, K. (2019) Distributionally adversarial attack. *Proc. AAAI Confer. Artif. Intell.* **33**(01), 2253–2260,

Appendix

A Convex analysis

This section gives an overview over well-known definitions and statements in convex analysis. In the following, $\mathcal{X}$ denotes a Banach space and $\mathcal{X}^*$ its dual.

Definition A.1 (Subdifferential). *For a convex function $f : \mathcal{X} \to (-\infty, \infty]$, we denote by*

$$\partial f(x) := \{\xi \in \mathcal{X}^* : f(z) - f(x) \geq \langle \xi, z - x\rangle \quad \forall z \in \mathcal{X}\} \subset \mathcal{X}^*$$

the subdifferential of f at $x \in \mathcal{X}$.

If $f(\cdot) = \|\cdot\|$, then the subdifferential is given by

$$\partial \|\cdot\|(x) = \{\xi \in \mathcal{X}^* | \langle \xi, x\rangle = \|x\|, \|\xi\|_* \leq 1\} \tag{A.1}$$

Definition A.2 (Fenchel conjugate). *For a function $f : \mathcal{X} \to [-\infty, +\infty]$, we denote by $f^*: \mathcal{X}^* \to [-\infty, +\infty]$,*

$$f^*(\xi) := \sup_{x \in \mathcal{X}} \langle \xi, x\rangle - f(x) \quad \text{for } \xi \in \mathcal{X}^*$$

the Fenchel conjugate of f.

A direct consequence of this definition is the so called Fenchel–Young inequality

$$\langle \xi, x\rangle \leq f(x) + f^*(\xi). \tag{A.2}$$

The next proposition yields the conditions under which the equality in (A.2) is obtained.

Proposition A.3 [6, Proposition 2.33]. *Let $f : \mathcal{X} \to]-\infty, +\infty]$ be a proper convex function. Then for $x \in \mathcal{X}$, the following three properties are equivalent:*

(i) $\xi \in \partial f(x)$.
(ii) $f(x) + f^*(\xi) \leq \langle \xi, x\rangle$.
(iii) $f(x) + f^*(\xi) = \langle \xi, x\rangle$.

If, in addition, f is lower-semicontinuous, then all of these properties are equivalent to the following one.

(i) $x \in \partial f^*(\xi)$.

Remark A.4. In Item 1, we use the canonical embedding to obtain the subspace relation $\mathcal{X} \subset \mathcal{X}^{**}$. Following [6, Remark 2.35], if $\mathcal{X}$ is reflexive, i.e. $\mathcal{X}^{**} = \mathcal{X}$, then it follows from Proposition A.3 that

$$x \in \partial f^*(\xi) \Longleftrightarrow \xi \in \partial f(x),$$

https://doi.org/10.1017/S0956792525100120 Published online by Cambridge University Press

which yields

$$(\partial f)^{-1}(\xi) = \{x \in \mathcal{X} : \xi \in \partial f(x)\} = \{x \in \mathcal{X} : x \in \partial f^*(\xi)\} = \partial f^*(\xi)$$

In the non-reflexive case, one can not argue as above, and we do not obtain the simple relation between ∂f^* and ∂f, see, e.g., [84].

An important corollary of Proposition A.3 is its application to the indicator function of the closed unit ball $f = \chi_{\overline{B_1}}$, where its convex conjugate for $\xi \in \mathcal{X}^*$ is given by

$$\chi^*_{\overline{B_1}}(\xi) = \sup_{x \in \mathcal{X}} \langle \xi, x \rangle - \chi_{\overline{B_1}}(x) = \sup_{x \in \overline{B_1}} \langle \xi, x \rangle = \|\xi\|_*.$$

Corollary A.5. *For a Banach space $\mathcal{X}$, and $\xi \in \mathcal{X}^*$ we have that*

$$\partial \, \|\cdot\|_* (\xi) \cap \mathcal{X} = \arg\max_{x \in \overline{B_1}} \langle \xi, x \rangle. \tag{A.3}$$

Proof. Since $\chi_{\overline{B_1}}$ is lower semicontinuous, we can use the equivalence of Item 3 and Item 1, to infer

$$x \in \partial \, \|\cdot\|_* (\xi) \quad \Leftrightarrow \quad \|\xi\|_* = \langle \xi, x \rangle - \chi_{\overline{B_1}}(x).$$

In the second statement, using the definition of $\|\xi\|_*$ as in (A.3), therefore yields that each x above realizes the supremum, which concludes the proof. □

B Refined version of Ascoli–Arzelà

Proposition B.1 [2, Proposition 3.3.1]. *Let $u^n : [0, T] \to \mathcal{S}$ be a sequence of curves, that fulfils the following conditions:*

(AA-i) There is a σ-sequentially compact set $K \subset \mathcal{S}$, such that

$$u^n(t) \in K \quad \text{for every} \quad t \in [0, T] \quad \text{and every} \quad n \in \mathbb{N}.$$

(AA-ii) There is a symmetric function $\omega : [0, T] \times [0, T] \to [0, +\infty)$ with $\lim_{(s,t)\to(r,r)} \omega(s, t) = 0$ for all $r \in [0, T] \setminus C$, where C is an at most countable set, such that

$$\limsup_{n \to \infty} d(u^n(s), u^n(t)) \leq \omega(s, t) \quad \text{for all} \quad s, t \in [0, T].$$

Then there exists a subsequence u^{n_k} and a limit curve $u : [0, T] \to \mathcal{S}$, which is d-continuous in $[0, T] \setminus C$, such that

$$u^{n_k}(t) \xrightarrow{\sigma} u(t) \quad \text{for all} \quad t \in [0, T].$$

C Taylor's formula in Banach spaces

Theorem C.1. *Suppose E, F are real Banach spaces, $U \subset E$ an open and nonempty subset, and $f \in C^n(U, F)$. Given $x_0 \in U$ choose $r > 0$ such that $x_0 + B_r \subset U$, where B_r is the open ball in E with centre 0 and radius r. Then for all $h \in B_r$ we have, using the abbreviation $h^k = (h, \ldots, h)$, k terms,*

$$f(x_0 + h) = \sum_{k=0}^{n} \frac{1}{k!} f(k)(x_0)(h)^k + R_n(x_0, h),$$

where the remainder R_n is of form

$$R_n(x_0, h) = \frac{1}{(n-1)!} \int_0^1 (1-t)^{n-1} [f^{(n)}(x_0 + th) - f^{(n)}(x_0)](h)^n \, dt.$$

Proof. A proof for this statement can be found, e.g., in [8, Theorem 30.1.3]. □

https://doi.org/10.1017/S0956792525100120 Published online by Cambridge University Press

D Prokhorov's theorem

Theorem D.1 (Prokhorov [7, Theorems 5.1–5.2]). *If a set $\mathcal{K} \subset \mathcal{P}(\mathcal{X})$ is tight, i.e.,*

$$\forall \epsilon > 0 \quad \exists K_\epsilon \text{ compact in } \mathcal{X} \text{ such that } \mu(\mathcal{X} \setminus K_\epsilon) \leq \epsilon \quad \forall \mu \in \mathcal{K}, \tag{D.1}$$

then $\mathcal{K}$ is relatively compact in $\mathcal{P}(\mathcal{X})$. Conversely, if $\mathcal{X}$ is a Polish space, every relatively compact subset of $\mathcal{P}(\mathcal{X})$ is tight.

E Helpful lemmas and supplementary proofs

In the following, we provide the proof of Lemma 2.8, which is a particular case of [2, Lemma 3.1.5]. For completeness, we provide a version of the proof that is specifically adapted to the case $p = \infty$.

Proof of Lemma 2.8. Let us suppose that for all $\tau > 0$, $\mathcal{E}_\tau(x) < \mathcal{E}(x)$ else $|\partial \mathcal{E}|(x) = 0$ and equality (2.4) holds trivially. We calculate

$$\begin{aligned}
\limsup_{\tau \to 0^+} \frac{\mathcal{E}(x) - \mathcal{E}_\tau(x)}{\tau} &= \limsup_{\tau \to 0^+} \sup_{z : 0 < d(x,z) \leq \tau} \frac{\mathcal{E}(x) - \mathcal{E}(z)}{\tau} \\
&= \inf_{\epsilon > 0} \sup_{0 < \tau \leq \epsilon} \sup_{z : 0 < d(x,z) \leq \tau} \frac{\mathcal{E}(x) - \mathcal{E}(z)}{\tau} \\
&= \inf_{\epsilon > 0} \sup_{z, \tau : 0 < d(x,z) \leq \tau \leq \epsilon} \frac{\mathcal{E}(x) - \mathcal{E}(z)}{\tau} \\
&= \inf_{\epsilon > 0} \sup_{z : 0 < d(x,z) \leq \epsilon} \sup_{\tau : d(x,z) \leq \tau \leq \epsilon} \frac{\mathcal{E}(x) - \mathcal{E}(z)}{\tau} \\
&= \inf_{\epsilon > 0} \sup_{z : 0 < d(x,z) \leq \epsilon} \frac{(\mathcal{E}(x) - \mathcal{E}(z))^+}{d(x,z)} - \frac{(\mathcal{E}(x) - \mathcal{E}(z))^-}{\epsilon} \\
&\overset{(*)}{=} \inf_{\epsilon > 0} \sup_{z : 0 < d(x,z) \leq \epsilon} \frac{\mathcal{E}(x) - \mathcal{E}(z)}{d(x,z)} \\
&= \limsup_{z \to x} \frac{\mathcal{E}(x) - \mathcal{E}(z)}{d(x,z)} \\
&= |\partial \mathcal{E}|(x).
\end{aligned}$$

Equality $(*)$ can be verified by the observation that $\mathcal{E}_\tau(x) < \mathcal{E}(x)$ for all $\tau > 0$ ensures the existence of at least one z with $d(x,z) \leq \epsilon$ such that $\mathcal{E}(x) - \mathcal{E}(z) \geq 0$.

□

Lemma E.1. *Let $\phi : [0,T] \to \mathbb{R}$ be continuous and $\psi : [0,T] \to \mathbb{R}$ be non-increasing. If $\phi(t) = \psi(t)$ for a.e. $t \in [0,T]$, then $\phi(t) = \psi(t)$ for all $t \in (0,T)$.*

Proof. Assume there is a $t \in (0,T)$ such that $\phi(t) \neq \psi(t)$. Without loss of generality $\phi(t) > \psi(t)$. Then we can take a sequence t_n with $t_n \to t$ and $\phi(t_n) = \psi(t_n)$ and $t_n > t$. The continuity of ϕ implies that for any $\epsilon > 0$, we can choose t_n small enough, such that $|\phi(t) - \phi(t_n)| < \epsilon$. This contradicts the monotonicity of ψ, since if we choose $\epsilon < \phi(t) - \psi(t)$, we obtain a $t_n > t$ with $\psi(t) < \psi(t_n) = \phi(t_n)$. In the case $\phi(t) < \psi(t)$, we can make the same argument with sequences $t_n < t$. □

In the following, we show that the arguments of [62, Theorem 7] can indeed be adapted to the case $p = \infty$. We closely follow the arguments in [62, Theorem 7], where it was proven for $p \in (1, \infty)$. For convenience, we copy the relevant steps and show how to adapt them to the case $p = \infty$.

Proof of Theorem 4.9. Let $\mathcal{L}^1_{(0,T)}$ denote the Lebesgue measure on $(0,T)$, then for η from Theorem 4.7, we define $\bar{\eta} := \frac{1}{T}\mathcal{L}^1_{(0,T)} \otimes \eta$ and the evaluation map $e{:}[0,T] \times C(0,T;\mathcal{X}) \to [0,T] \times \mathcal{X}$ by

$$e(t,u) = (t, e_t(u)) = (t, u(t)).$$

https://doi.org/10.1017/S0956792525100120 Published online by Cambridge University Press

We observe that $e_{\#}\eta = \bar{\mu}$ and denote by $\bar{\eta}_{x,t}$ the Borel family of probability measures on $C(0,T;\mathcal{X})$ obtained by disintegration of $\bar{\eta}$ with respect to e, such that $d\bar{\eta}_{x,t}(u)d\bar{\mu}(t,x) = d\bar{\eta}(t,u)$. Notably $\bar{\eta}_{x,t}$ is concentrated on $\{u{:}e_t(u) = x\} \subset C(0,T;\mathcal{X})$. Since $\mathcal{X}$ is assumed to satisfy the Radon–Nikodým property the pointwise derivative $u'(t) = \lim_{h\to 0} \frac{u(t+h)-u(t)}{h}$ is defined a.e. for an absolutely continuous curve u. We now show that

$$A := \{(t,u) \in [0,T] \times C(0,T;\mathcal{X}){:}u'(t) \text{ exists}\}$$

is a Borel set and $\bar{\eta}(A^c) = 0$. For every $h \neq 0$, we define the continuous function $g_h{:}[0,T] \times C(0,T;\mathcal{X}) \to \mathcal{X}$ by $g_h(t,u) = \frac{u(t+h)-u(t)}{h}$, where we extend the function u outside of $[0,T]$ by $u(s) = u(0)$ for $s < 0$ and $u(s) = u(T)$ for $s > T$. By completeness of $\mathcal{X}$

$$A^c := \{(t,u) : \limsup_{(h,k)\to(0,0)} \|g_h(t,u) - g_k(t,u)\| > 0\}$$

and because of the continuity of the function $(t,u) \mapsto \|g_h(t,u) - g_k(t,u)\|$, A^c and A are Borel sets. Since $\bar{\eta}$ is concentrated on $[0,T] \times \mathrm{AC}^\infty(0,T;\mathcal{X})$ and $u'(t)$ exists a.e. for an absolutely continuous curve u, by Fubini's theorem $\bar{\eta}(A^c) = 0$. Thus, for $\bar{\eta}$-a.e. (t,u) the map

$$\psi(t,u) = u'(t)$$

is well-defined. For every $x^* \in \mathcal{X}^*$, we define $\psi_{x^*}(t,h) := \langle x^*, u'(t)\rangle$ on $(t,u) \in A$. As a limit of continuous functions ψ_{x^*} is a Borel function on A and thus $\bar{\eta}$ measurable. Since $\mathcal{X}$ is separable, Pettis theorem ensures that ψ is a $\bar{\eta}$-measurable function. Now we can define the vector field

$$v_t(x) := \int_{C(0,T;\mathcal{X})} u'(t)\mathrm{d}\bar{\eta}_{x,t} \quad \text{for } \bar{\mu}\text{-a.e. } (t,x) \in (0,T) \times \mathcal{X}.$$

For clarity, we now indicate the varibles over which the *ess sup* is taken in brackets after the respective measure. Using this notation we estimate

$$\begin{aligned}
\bar{\mu} - \operatorname{ess\,sup} \|v\| &= \bar{\mu}(x,t) - \operatorname{ess\,sup} \left\| \int_{C(0,T;\mathcal{X})} u'(t) d\bar{\eta}_{x,t}(u) \right\| \\
&\leq \bar{\mu}(x,t) - \operatorname{ess\,sup} \int_{C(0,T;\mathcal{X})} \left\|u'(t)\right\| d\bar{\eta}_{x,t}(u) \\
&\leq \bar{\mu}(x,t) - \operatorname{ess\,sup} \left(\bar{\eta}_{x,t}(u) - \operatorname{ess\,sup} \left\|u'(t)\right\|\right) \\
&\leq \bar{\eta}(u,t) - \operatorname{ess\,sup} \left\|u'(t)\right\| < +\infty,
\end{aligned}$$

and thus $v \in L^\infty(\bar{\mu};\mathcal{X})$, where the last inequality follows from Lemma E.3. By Jensen's inequality we have for every $[a,b] \subset [0,T]$,

$$\begin{aligned}
\int_a^b \|v_t\|_{L^\infty(\mu_t;\mathcal{X})}\,\mathrm{d}t &= \int_a^b \mu_t(x) - \operatorname{ess\,sup} \|v_t(x)\|\,\mathrm{d}t \\
&= \int_a^b \mu_t(x) - \operatorname{ess\,sup} \left\| \int_{C(0,T;\mathcal{X})} u'(t) d\bar{\eta}_{x,t} \right\| \mathrm{d}t \\
&\leq \int_a^b \mu_t(x) - \operatorname{ess\,sup} \int_{C(0,T;\mathcal{X})} \left\|u'(t)\right\| d\bar{\eta}_{x,t}\,\mathrm{d}t \qquad \text{(E.1)} \\
&\leq \int_a^b \mu_t(x) - \operatorname{ess\,sup}\ \bar{\eta}_{x,t}(u) - \operatorname{ess\,sup} \left\|u'(t)\right\|\ \mathrm{d}t \\
&= \int_a^b \eta(u) - \operatorname{ess\,sup} \|u'(t)\|\,\mathrm{d}t = \int_a^b |\mu'|(t)\,\mathrm{d}t.
\end{aligned}$$

such that $\|v_t\|_{L^p(\mu_t;\mathcal{X})} \leq |\mu|'(t)$ for a.e. $t \in (0,T)$. In the last inequality we used the fact, that $\mathrm{d}\eta = d\bar{\eta}_{x,t}\,\mathrm{d}\mu_t$ holds for a.e. $t \in (0,T)$ together with Lemma E.3. For more rigorous justifications regarding measurably and integrability of all involved quantities, we refer to [62, Theorem 7]. To show that $(\mu, v) \in \mathrm{EC}^\infty(\mathcal{X})$ we take $\varphi \in C_b^1(\mathcal{X})$ and observe that $t \to \int_{\mathcal{X}} \varphi(x)\,\mathrm{d}\mu_t(x)$ is absolutely continuous, since for $\gamma \in \Gamma_0(\mu_t, \mu_s)$

https://doi.org/10.1017/S0956792525100120 Published online by Cambridge University Press

$$\left|\int_{\mathcal{X}} \varphi \,\mathrm{d}\mu_t - \int_{\mathcal{X}} \varphi \,\mathrm{d}\mu_s\right| \leq \int_{\mathcal{X}\times\mathcal{X}} |\varphi(x) - \varphi(\tilde{x})| d\gamma \leq$$
$$\sup_{x\in\mathcal{X}} \|D\varphi(x)\| \int_{\mathcal{X}\times\mathcal{X}} \|x - \tilde{x}\| d\gamma \leq \sup_{x\in\mathcal{X}} \|D\varphi(x)\| W_\infty(\mu_t, \mu_s).$$

Further,

$$\begin{aligned}\int_{\mathcal{X}} \varphi \,\mathrm{d}\mu_t - \int_{\mathcal{X}} \varphi \,\mathrm{d}\mu_s &= \int_{C(0,T;\mathcal{X})} \varphi(u(t)) - \varphi(u(s)) \,\mathrm{d}\eta(u)\\ &= \int_{C(0,T;\mathcal{X})} \langle D\varphi(u(s)), u(t) - u(s)\rangle \,\mathrm{d}\eta(u) + \int_{C(0,T;\mathcal{X})} \|u(t) - u(s)\| \omega_{u(s)}(u(t)) \,\mathrm{d}\eta(u)\\ &= \int_{C(0,T;\mathcal{X})} \langle D\varphi(u(s)), \int_s^t u'(r)\,\mathrm{d}r\rangle \,\mathrm{d}\eta(u) + \int_{C(0,T;\mathcal{X})} \|u(t) - u(s)\| \omega_{u(s)}(u(t)) \,\mathrm{d}\eta(u)\end{aligned}$$

where

$$\omega_x(y) = \frac{\varphi(y) - \varphi(x) - \langle D\varphi(u(x)), y - x\rangle}{\|y - x\|}.$$

We observe

$$\frac{1}{t-s}\langle D\varphi(u(s)), \int_s^t u'(r)\,\mathrm{d}r\rangle \to \langle D\varphi(u(s)), u'(s)\rangle \quad \text{for } \eta\text{-a.e. } u$$

and

$$\frac{\|u(t) - u(s)\|}{t-s}\omega_{u(s)}(u(t)) \to 0 \quad \text{for } \eta\text{-a.e. } u$$

and have for η-a.e. u the upper bounds

$$\begin{aligned}\frac{1}{|t-s|}|\langle D\varphi(u(s)), \int_s^t u'(r)\,\mathrm{d}r\rangle| &\leq \sup_{x\in\mathcal{X}} \|D\varphi(x)\|_* \frac{\left\|\int_s^t u'(r)\,\mathrm{d}r\right\|}{|s-t|}\\ &\leq \sup_{x\in\mathcal{X}} \|D\varphi(x)\|_* \operatorname*{ess\,sup}_{r\in[0,T]} |\mu'|(r) < +\infty\end{aligned}$$

and

$$\begin{aligned}\frac{\|u(t) - u(s)\|}{|t-s|}|\omega_{u(s)}(u(t))| &\leq \operatorname*{ess\,sup}_{r\in[0,T]} |\mu'|(r) \left(\frac{|\varphi(u(t)) - \varphi(u(s))|}{\|u(t) - u(s)\|} + \frac{|\langle D\varphi(u(s)), u(t) - u(s)\rangle|}{\|u(t) - u(s)\|}\right)\\ &\leq \operatorname*{ess\,sup}_{r\in[0,T]} |\mu'|(r)\, 2\mathrm{Lip}(\varphi) < +\infty.\end{aligned}$$

Dividing by $t - s$ and passing to the limit $t \to s$ by using Lebesgue theorem, we obtain

$$\frac{\mathrm{d}}{ds}\int_{\mathcal{X}} \varphi \,\mathrm{d}\mu_s = \int_{C(0,T;\mathcal{X})} \langle D\varphi(u(s)), u'(s)\rangle \,\mathrm{d}\eta(u) = \int \langle D\varphi, v_t\rangle \,\mathrm{d}\mu_t \quad \text{for a.e. } s \in (0,T).$$

This pointwise derivative corresponds to the distributional derivative and we obtain $(\mu, v) \in \mathrm{EC}^\infty(\mathcal{X})$. □

Similarly, we can adapt [62, Theorem 8] to the case $p = \infty$, which we again show by reusing most of the arguments from the corresponding proof in [62].

Proof of Theorem 4.11. This theorem was proven in [62, Theorem 8] for $p \in (1, +\infty)$ and can easily be extended to the case $p = +\infty$. Let $(\mu_t)_{t\in[0,T]}$ be a family of measures in $\mathcal{P}_\infty(\mathcal{X})$ and for each t we have a velocity field $v_t \in L^\infty(\mu_t;\mathbb{R}^d)$ with *ess sup*$\|v_t\|_{L^\infty(\mu_t)} < \infty$, solving the continuity equation in the sense of distributions. Since

$$\|v\|_{L^p(\tilde{\mu};\mathcal{X})} \leq T^{1/p} \operatorname*{ess\,sup} \|v_t\|_{L^\infty(\mu_t)} < \infty$$

we can apply [62, Theorem 8] (i.e., the statement of Theorem 4.11) for all $p \in (1, \infty)$ and get

$$|\mu'|_{(p)}(t) \leq \|v_t\|_{L^p(\mu_t;\mathcal{X})} \text{ for a.e. } t \in (0,T) \text{ and all } p \in (1,\infty).$$

https://doi.org/10.1017/S0956792525100120 Published online by Cambridge University Press

Therefore,

$$W_p(\mu_t, \mu_s) \le \int_t^s |\mu'|_{(p)}(\tilde{t})d\tilde{t} \le \int_t^s \|v_t\|_{L^p(\mu_{\tilde{t}};\mathcal{X})}d\tilde{t} \le \int_t^s \|v_t\|_{L^\infty(\mu_{\tilde{t}};\mathcal{X})}d\tilde{t}$$

for all $t, s \in [0, T]$ with $t \le s$ and $p \in (1, \infty)$, where $|\mu'|_{(p)}$ denotes the metric derivative of μ in W_p. Taking the limit $p \to \infty$, we get

$$W_\infty(\mu_t, \mu_s) = \lim_{p\to\infty} W_p(\mu_t, \mu_s) \le \int_t^s \|v_t\|_{L^\infty(\mu_{\tilde{t}};\mathcal{X})}d\tilde{t}$$

for all $t, s \in [0, T]$ with $t \le s$ and thus by the minimality of the metric derivative, see Remark 1.2,

$$|\mu'|_{(\infty)}(t) \le \|v_t\|_{L^\infty(\mu_t;\mathcal{X})} \text{ for a.e. } t \in (0, T).$$

□

Lemma E.2. *Let μ be a Borel probability measure on $\mathcal{X}$ and $v{:}\mathcal{X} \to \mathcal{X}$, $\tilde{v} : \mathcal{X} \to \mathcal{X}$ be two μ-measurable functions with*

$$\int \langle D\varphi(x), v(x)\rangle \, d\mu(x) = \int \langle D\varphi(x), \tilde{v}(x)\rangle \, d\mu(x) \quad \forall \varphi \in C_b^1(\mathcal{X})$$

then

$$\int \langle \xi, v(x)\rangle \, d\mu(x) = \int \langle \xi, \tilde{v}(x)\rangle \, d\mu(x) \quad \forall \xi \in \mathcal{X}^*. \tag{E.2}$$

Proof. Let $g_n : \mathbb{R} \to \mathbb{R}$ be the function with $g(0) = 0$ and

$$g'(x) = \begin{cases} 0 & \text{for } |x| > n+1, \\ 1 & \text{for } |x| < n, \\ n+1-x & \text{for } x \in [n, n+1], \\ n+1+x & \text{for } x \in [-(n+1), -n]. \end{cases}$$

Then for each $\xi \in \mathcal{X}^*$ we get $G_n{:}x \mapsto g_n(\langle \xi, x\rangle) \in C_b^1(\mathcal{X})$ with $DG_n(x) = g_n'(\langle \xi, x\rangle)\, \xi$ and

$$\begin{aligned} \int g_n'(\langle \xi, x\rangle)\langle \xi, v(x)\rangle \, \mathrm{d}\mu &= \int \langle DG_n(x), v(x)\rangle \, \mathrm{d}\mu = \int \langle DG_n(x), \tilde{v}(x)\rangle \, \mathrm{d}\mu \\ &= \int g_n'(\langle \xi, x\rangle)\langle \xi, \tilde{v}(x)\rangle \, \mathrm{d}\mu \end{aligned}$$

Since for $n \to \infty$ we have $g'(\langle \xi, x\rangle) \to 1$ pointwise we can apply the Lebesgue dominated convergence theorem (with the functions $|\langle \xi, v(x)\rangle|$ and $|\langle \xi, \tilde{v}(x)\rangle|$ as bound) to obtain (E.2). □

The following lemma shows that the disintegration property can be transferred to an inequality for essential suprema. For more details on disintegration, we refer to [2, Ch. 5.3] and [33, Ch. III-70]. The proof strategy is taken from [85, Lemma 2] and amounts to controlling the null sets of the measures involved.

Lemma E.3. *Given $\mathcal{X}$, $\mathcal{Z}$ Radon separable metric spaces, a measure $\mu \in \mathcal{P}(\mathcal{X})$, a Borel map $\pi : \mathcal{X} \to \mathcal{Z}$ and a disintegration $d\mu = d\mu_z d\nu$, with $\nu = \pi_\# \mu$ and $\{\mu_z\}_{z\in\mathcal{Z}} \subset \mathcal{P}(\mathcal{X})$ being a family of probability measures, then we have that*

$$\mu(x) - \operatorname{ess\,sup} f(x) \ge \nu(z) - \operatorname{ess\,sup}\ \mu_z(x) - \operatorname{ess\,sup} f(x)$$

for every Borel map $f : \mathcal{X} \to [0, \infty]$.

Proof. Using the disintegration property for every Borel set A, we obtain

$$\mu(A) = 0 \quad \Leftrightarrow \quad \mu_z(A) \text{ for } \nu - \text{a.e. } z \in \mathcal{Z}.$$

https://doi.org/10.1017/S0956792525100120 Published online by Cambridge University Press

Now assume that $\mu(A) = 0$, then we know that there exists a Borel set $B \subset \mathcal{Z}$ with $\nu(B) = 0$ and $\mu_z(A) = 0$ for all $z \in \mathcal{Z} \setminus B$. Therefore,

$$\begin{aligned}
&\sup_{x\in\mathcal{X}\setminus A} f(x) \geq \inf_{\tilde{A}\,:\,\mu_z(\tilde{A})=0} \sup_{x\in\mathcal{X}\setminus\tilde{A}} f(x) = \mu_z(x) - \operatorname{ess\,sup} f(x) \qquad \text{for all } z \in \mathcal{Z} \setminus B \\
\Rightarrow\ &\sup_{x\in\mathcal{X}\setminus A} f(x) \geq \sup_{z\in\mathcal{Z}\setminus B} \mu_z(x) - \operatorname{ess\,sup} f(x) \\
&\qquad\qquad \geq \inf_{\tilde{B}\,:\,\nu(\tilde{B})=0} \sup_{z\in\mathcal{Z}\setminus\tilde{B}} \mu_z(x) - \operatorname{ess\,sup} f(x) \\
&\qquad\qquad = \nu(z) - \operatorname{ess\,sup}\ \mu_z(x) - \operatorname{ess\,sup} f(x)
\end{aligned}$$

and since this holds for every μ-null set A, we can take the infimum to obtain

$$\mu(x) - \operatorname{ess\,sup} f(x) = \inf_{A:\mu(A)=0} \sup_{x\in\mathcal{X}\setminus A} f(x) \geq \nu(z) - \operatorname{ess\,sup}\ \mu_z(x) - \operatorname{ess\,sup} f(x).$$

□

Lemma E.4. *Let $\eta \in \mathcal{P}(C(0,T;\mathcal{X}))$, then we have that*

$$\eta(u) - \operatorname{ess\,sup} \left|u'\right|(t) \leq 1 \quad \textit{for a.e. } t \in (0,T) \Longleftrightarrow \operatorname*{ess\,sup}_{t\in(0,T)} \left|u'\right|(t) \leq 1 \qquad \textit{for } \eta \textit{ a.e. } u \in C(0,T;\mathcal{X}).$$

Proof. Choosing $\psi = \chi_{[0,1]}$, and observing that $\psi(|u'|(t))$ is $\bar{\eta}$-measurable (see [63, Eq. (55)]) implies

$$\begin{aligned}
&\eta(u) - \operatorname{ess\,sup} \left|u'\right|(t) \leq 1 \quad \text{for a.e. } t \in (0,T) \\
&\qquad\qquad \Longleftrightarrow \int_{C(0,T,\mathcal{X})} \psi(\left|u'\right|(t))\,\mathrm{d}\eta(u) = 0 \quad \text{for a.e. } t \in (0,T) \\
\Longleftrightarrow \int_0^T \int_{C(0,T,\mathcal{X})} \psi(\left|u'\right|(t))\,\mathrm{d}\eta(u)\,\mathrm{d}t = 0 &\Longleftrightarrow \int_{C(0,T,\mathcal{X})} \int_0^T \psi(\left|u'\right|(t))\,\mathrm{d}\eta(u)\,\mathrm{d}t = 0 \\
\Longleftrightarrow \int_{C(0,T,\mathcal{X})} \int_0^T \psi(\left|u'\right|(t))\,\mathrm{d}t\,\mathrm{d}\eta(u) = 0 &\Longleftrightarrow \int_0^T \psi(\left|u'\right|(t))\,\mathrm{d}t = 0 \quad \text{for } \eta \text{ a.e. } u \in C(0,T;\mathcal{X}) \\
&\Longleftrightarrow \operatorname*{ess\,sup}_{t\in(0,T)} \left|u'\right|(t) \leq 1 \quad \text{for } \eta \text{ a.e. } u \in C(0,T;\mathcal{X}),
\end{aligned}$$

where we use Fubini–Tonelli theorem to change the order of integration. □

F Multivalued correspondences

For multivalued correspondences, generalizations of continuity and measurability can be defined. We use the definitions from [24]. In the following, we write $\varphi : \mathcal{X} \rightrightarrows \mathcal{Z}$ to denote a mapping $\varphi : \mathcal{X} \to 2^{\mathcal{Z}}$.

Definition F.1 (Weak measurability). *Let (S, Σ) be a measurable space and $\mathcal{X}$ be a topological space. We say that a correspondence φ:$S \rightrightarrows \mathcal{X}$ is weakly measurable, if*

$$\varphi^l(G) \in \Sigma \textit{ for all open sets } G \textit{ of } \mathcal{X},$$

where

$$\varphi^l(G) := \{s \in S | \varphi(s) \cap G \neq \emptyset\} \tag{F.1}$$

is the so-called lower inverse.

Definition F.2 (Measurability). *Let (S, Σ) be a measurable space and $\mathcal{X}$ a topological space. We say that a correspondence φ:$S \rightrightarrows \mathcal{X}$ is measurable, if*

$$\varphi^l(F) \in \Sigma \textit{ for all closed sets } F \textit{ of } \mathcal{X}.$$

https://doi.org/10.1017/S0956792525100120 Published online by Cambridge University Press

The next theorem is known as the measurable maximum theorem, where we refer to [24, Theorem 18.19] for the proof of this statement.

Theorem F.3 (Measurable maximum theorem). *Let $\mathcal{X}$ be a separable metrizable space and (S, Σ) a measurable space. Let $\varphi: S \rightrightarrows \mathcal{X}$ be a weakly measurable correspondence with nonempty compact values, and suppose $f: S \times \mathcal{X} \to \mathbb{R}$ is a Carathéodory function. Define the value function $m: S \to \mathbb{R}$ by*

$$m(s) = \max_{x \in \varphi(s)} f(s, x),$$

and the correspondence $\mu: S \rightrightarrows \mathcal{X}$ of maximizers by

$$\mu(s) = \{x \in \varphi(s) : f(s, x) = m(s)\}.$$

Then

- *The value function m is measurable.*
- *The "argmax" correspondence μ has nonempty and compact values.*
- *The "argmax" correspondence μ is measurable and admits a measurable selector.*

G Calculations for distributional adversaries

For completeness, we state all lemmas used in section 5.2 here. Those lemmas correspond to a lemma proven in section section 4.3 and are only adapted to the setting of the transport distance D.

Lemma G.1. *Let $\mathcal{X} \times \mathcal{Y} = \mathbb{R}^d \times \mathbb{R}^m$, then the correspondence*

$$r_\varepsilon(x, y) = \operatorname*{arg\,min}_{(\tilde{x}, \tilde{y}) : c(x, y, \tilde{x}, \tilde{y}) \le \varepsilon} E(\tilde{x}, \tilde{y}) = \left(\operatorname*{arg\,min}_{\tilde{x} \in \overline{B_\varepsilon}(x)} E(\tilde{x}, y), y \right)$$

is measurable and admits a $\mathcal{B}(\mathcal{X} \times \mathcal{Y})$-measurable selector. Further, for each measurable selector $r_\varepsilon : \mathcal{X} \times \mathcal{Y} \to \mathcal{X}$ we have the following,

$$(r_\varepsilon)_\#(\mu) \in \operatorname*{arg\,min}_{D(\mu, \tilde{\mu}) \le \varepsilon} \int E(x, y)\, \mathrm{d}\mu(x, y). \tag{G.1}$$

Proof. We consider the correspondence $\varphi : \mathcal{X} \times \mathcal{Y} \rightrightarrows \mathcal{X} \times \mathcal{Y}$ given by

$$(x, y) \mapsto \left(\overline{B_\varepsilon}(x), y \right)$$

where on the input space we use the topology induced by $\|\cdot\|_{\mathcal{X}} + \|\cdot\|_{\mathcal{Y}}$ and the output space is interpreted as the standard Euclidean space. Then we have that for every open set $G \in \mathcal{X}$ that

$$\varphi^l(G) = \{(x, y) \in \mathcal{X} \times \mathcal{Y} : \left(\overline{B_\varepsilon}(x), y \right) \cap G \neq \emptyset\} \subset \mathcal{B}(\mathcal{X} \times \mathcal{Y}),$$

is open, which implies weak measurability, according to Definition F.1. Furthermore, we define the map $f((x, y), (\tilde{x}, \tilde{y})) := -E(\tilde{x}, \tilde{y})$ which is a Carathéodory function, since E is continuous, with

$$\max_{(\tilde{x}, \tilde{y}) \in \varphi(x, y)} f((x, y), (\tilde{x}, \tilde{y})) = \min_{(\tilde{x}, \tilde{y}) : c(x, y, \tilde{x}, \tilde{y}) \le \epsilon} -E(\tilde{x}, \tilde{y})$$

Then Theorem F.3 ensures the existence of a measurable selector. To prove (G.1) we observe that if $D(\mu, \tilde{\mu}) \le \varepsilon$, then for an optimal transport plan $\gamma \in \Gamma_0(\mu, \tilde{\mu})$, we know that $y = \tilde{y}$ and $\|x - \tilde{x}\| \le \varepsilon$, γ-a.e. Thus, using the disintegration $d\gamma(x, y, \tilde{x}, \tilde{y}) = d\psi_{x,y}(\tilde{x}, \tilde{y})\, \mathrm{d}\mu(x, y)$, for every $\tilde{\mu}$ with $D(\tilde{\mu}, \mu) \le \varepsilon$, we calculate

$$\int E(\tilde{x}, \tilde{y}) d\tilde{\mu}(\tilde{x}, \tilde{y}) = \int E(\tilde{x}, \tilde{y}) d\gamma(x, y, \tilde{x}, \tilde{y}) = \int \int E(\tilde{x}, \tilde{y}) d\psi_{x,y}(\tilde{x}, \tilde{y})\, \mathrm{d}\mu(x, y)$$

$$= \int \int E(\tilde{x}, y) d\psi_{x,y}(\tilde{x}, \tilde{y})\, \mathrm{d}\mu(x, y) \ge \int \int E(r_\varepsilon(x, y)) d\psi_{x,y}(\tilde{x}, \tilde{y})\, \mathrm{d}\mu(x, y) = \int E(r_\varepsilon(x, y))\, \mathrm{d}\mu(x, y),$$

and (G.1) follows. □

https://doi.org/10.1017/S0956792525100120 Published online by Cambridge University Press

Lemma G.2. *Let $\mathcal{X} \times \mathcal{Y} = \mathbb{R}^d \times \mathbb{R}^m$, $E{:}\mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ be in $C^1(\mathcal{X} \times \mathcal{Y})$ and $\mu \in \mathcal{P}_\infty(\mathcal{X} \times \mathcal{Y})$, then the metric slope with respect to D is given by*

$$|\partial \mathcal{E}|(\mu) = \int \|\nabla_x E(x, y)\|_{\mathcal{X}^*} \, d\mu.$$

and $|\partial \mathcal{E}|$ is a strong upper gardient.

Proof. We follow the arguments of Lemma 4.17. By Lemma G.1 we obtain

$$\begin{aligned} |\partial \mathcal{E}|(\mu) &= \limsup_{\tau \to 0} \frac{\mathcal{E}(\mu) - \mathcal{E}_\tau(\mu)}{\tau} = \limsup_{\tau \to 0} \int \frac{E(x, y) - E(r_\tau(x, y))}{\tau} \, \mathrm{d}\mu \\ &= \int \lim_{\tau \to 0} \frac{E(x, y) - E(r_\tau(x, y))}{\tau} \, \mathrm{d}\mu = \int \|\nabla_x E(x, y)\|_{\mathcal{X}^*} \, \mathrm{d}\mu, \end{aligned}$$

where dominated convergence together with Lemma 3.5 and Item 3 was used to draw the limit into the integral.

To prove that $|\partial \mathcal{E}|$ is a strong upper gradient we observe that $\|\nabla_x E(x, y)\|_{\mathcal{X}^*}$ is continuous and in particular lower semicontinuous such that we can use [2, Lemma 5.1.7] to prove that the map $t \mapsto |\partial \mathcal{E}|(\mu_t)$ is lower semicontinuous and thus Borel for every absolutely continuous curve μ_t. Assume that $\int_s^t \int_{\mathcal{X} \times \mathcal{Y}} \|\nabla_x E(x, y)\|_{\mathcal{X}^*} \, \mathrm{d}\mu_r(x, y) |\mu'|(r) \, \mathrm{d}r = \int_s^t |\partial \mathcal{E}|(\mu_r) |\mu'|(r) \, \mathrm{d}r < +\infty$, otherwise (1.4) holds trivially. By Theorem 4.7 we can estimate

$$\begin{aligned} |\mathcal{E}(\mu_t) - \mathcal{E}(\mu_s)| &= \left| \int_{\mathcal{X} \times \mathcal{Y}} E(x, y) \, \mathrm{d}\mu_t(x, y) - \int_{\mathcal{X} \times \mathcal{Y}} E(x, y) \, \mathrm{d}\mu_s(x, y) \right| \\ &= \left| \int_{C(0,T;\, \mathcal{X} \times \mathcal{Y})} E(u(t)) \, \mathrm{d}\eta(u) - \int_{C(0,T;\, \mathcal{X})} E(u(s)) d\eta(u) \right| \\ &\leq \int_{C(0,T;\, \mathcal{X} \times \mathcal{Y})} |E(u(t)) - E(u(s))| \, d\eta(u) \\ &\leq \int_{C(0,T;\, \mathcal{X} \times \mathcal{Y})} \int_s^t \|\nabla_x E(u(r))\|_{\mathcal{X}^*} \, |u'|(r) \, \mathrm{d}r d\eta(u) \\ &\leq \int_s^t \int_{\mathcal{X} \times \mathcal{Y}} \|\nabla_x E(x, y)\|_{\mathcal{X}^*} \, \mathrm{d}\mu_r(x, y) |\mu'|(r) \, \mathrm{d}r < +\infty. \end{aligned}$$

Here we use that η is concentrated on $AC^\infty(0, T; \mathcal{X} \times \mathcal{Y})$ and by the definition of the the extended distance $c(x, \tilde{x}, y, \tilde{y})$ on $\mathcal{X} \times \mathcal{Y}$ a curve $u(t) \in AC^\infty(0, T; \mathcal{X} \times \mathcal{Y})$ only moves in $\mathcal{X}$-direction and for those curves $\|\nabla_x E(u(r))\|_{\mathcal{X}^*}$ acts like a strong upper gradient.

□

H Details on numerical examples

Here, we give some details on the experiment that produces Figure 1, the source code is provided at github.com/TimRoith/AdversarialFlows.

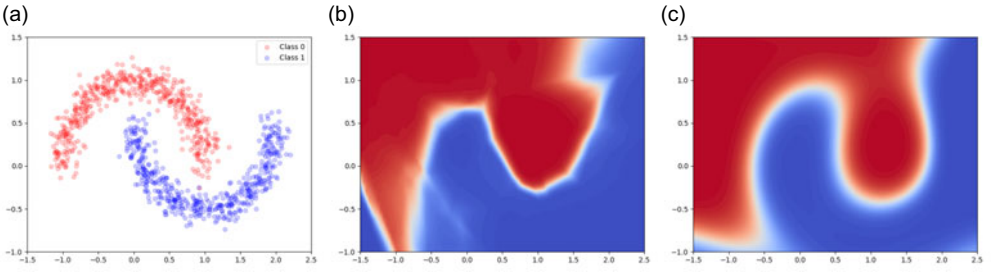
H.1 Training the neural network

We sample $K = 1000$ labelled data points $((x_1, y_1), \dots, (x_K, y_K))$, with $x_k \in \mathbb{R}^2$, $y_k \in \{0, 1\}$, from the two moons data set using the `sci-kit` package [78], see Figure H1a.

Using `PyTorch` [76], we then train a neural network using the architecture displayed in Figure H2 as proposed in [52], to obtain a mapping $h_\theta : \mathbb{R}^2 \to [0, 1]$, parametrized by θ. Here "Linear $d^l \to d^{l+1}$" in the lth layer, denotes an *affine* linear mapping [86] given by

$$z \mapsto Wz + b, \qquad \text{with learnable parameters} \quad W \in \mathbb{R}^{d^{l+1} \times d^l}, b \in \mathbb{R}^{d^{l+1}}$$

https://doi.org/10.1017/S0956792525100120 Published online by Cambridge University Press



The dataset used to train a binary classifier.

The trained classifier, with ReLU activation, evaluated on the domain $[-1.5, 2.5] \times [-1., 1.5]$.

The trained classifier, with GeLU activation, evaluated on the domain $[-1.5, 2.5] \times [-1., 1.5]$.

Figure H1. *Visualization of the dataset and trained classifiers used in the experiments.*

the activation functions "ReLU" [42], "GeLU" [51] and "Sigmoid" are defined entry-wise for $i = 1, \dots, n$, as

$$\mathrm{ReLU}(z_i) := \max\{0, z_i\}, \qquad \mathrm{GeLU}(z_i) := z_i \cdot \Phi(z_i), \qquad \mathrm{Sigmoid}(z_i) := \frac{1}{1 + \exp(-z_i)},$$

where Φ denotes the cumulative distribution function of the standard normal distribution. Here, we included both ReLU and GeLU (as a smooth approximation) to have an activation function, typically used in practice and a differentiable approximation fitting into the framework of section 3.3. During training, we process batches of inputs $\mathbf{z} = (z^1, \dots, z^B)$, with $z_i \in \mathbb{R}^{d^l}$, where "Batch Norm (B)", as proposed in [54], uses the entry-wise mean $\mu(\mathbf{z})_i := \frac{1}{B}\sum_{b=1}^{B} z_i^b$ and variance $\sigma(\mathbf{z})_i := \frac{1}{B}\sum_{b=1}^{B}(z_i^b - \mu(\vec{z}))^2$ and is defined as

$$z_i^b \mapsto \frac{z_i^b - \mu(\mathbf{z})_i}{\sqrt{\sigma(\mathbf{z})_i^2 + \epsilon}} \cdot \gamma_i + \beta_k, \qquad \text{with learnable parameters} \quad \gamma, \beta \in \mathbb{R}^{d^l},$$

where $\epsilon = 10^{-5}$ is a small constant, added for numerical stability. During inference, the mean and variance are replaced by an estimate over the whole dataset, such that the output does not depend on the batch it is given. In total, θ denotes the collection of weights W, biases b and batch norm parameters γ, β of all layers. For training, we consider the loss function

$$\mathcal{L}(\theta) = \frac{1}{2K}\sum_{k=1}^{K} |h_\theta(x_k) - y_k|^2, \tag{H.1}$$

where we employ the Adam optimizer [56], with standard learning rates, to approximate a minimizer. In each step, we employ a batched version of the function in (H.1), i.e., instead of using all data points at once, in each so-called *epoch*, we randomly sample disjoint subsets of $\{1, \dots, K\}$, of size $B = 100$ and only sum over these points. We run this training process for a total of 100 epochs, to obtain a set of parameters θ^*, with a train loss of approximately $\mathcal{L}(\theta^*) \approx 0.002$ for ReLU and $\mathcal{L}(\theta^*) \approx 0.009$ for GeLU. The trained mappings h_θ are visualized in Figure H1 and b.

H.2 Computing IGFSM and the minimizing movement scheme

We now detail the iteration as displayed in Figure 1, first for ReLU. Here, we choose the initial value $x^0 = (0.1, 0.55)$, as it is close to the decision boundary, with $h_{\theta^*}(x^0) \approx 0.97$, an adversarial budget of $\varepsilon = 0.2$ and the energy

$$E(x) := |h_{\theta^*}(x) - 1|^2 + \chi_{B^\infty_{\tilde{\varepsilon}}(x^0)}.$$

https://doi.org/10.1017/S0956792525100120 Published online by Cambridge University Press

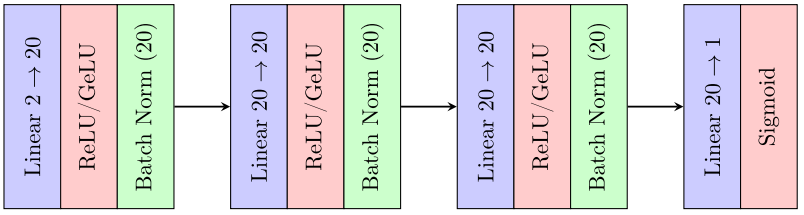


Figure H2. *The network architecture used in the examples.*

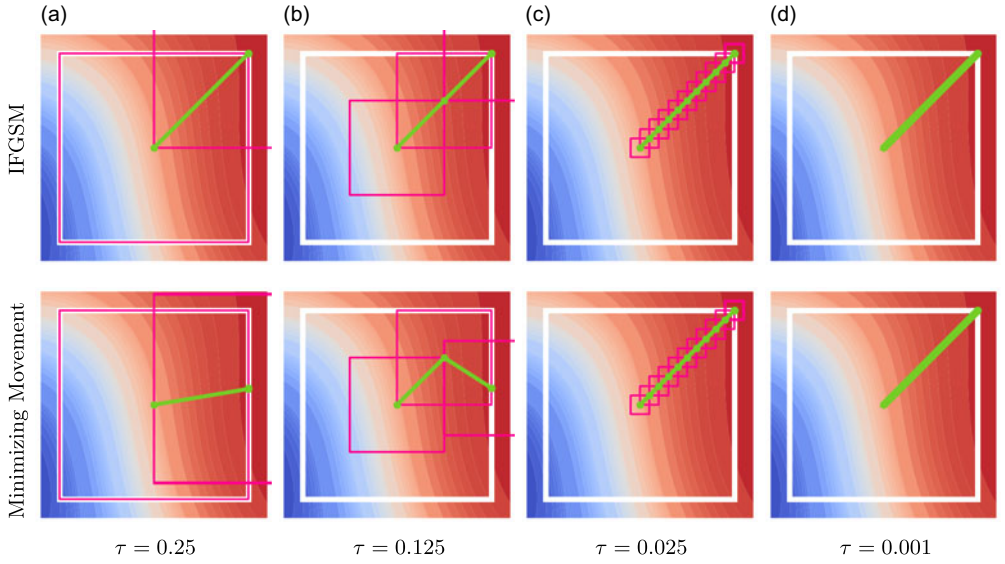


Figure H3. *The same experiment as in Figure 1, but using a net employing the GeLU activation function.*

IFGSM is an explicit iteration and can therefore be implemented directly, where the gradient is computed with the automatic differentiation tools of `PyTorch`. For the minimizing movement scheme MinMove, we need to solve the problem

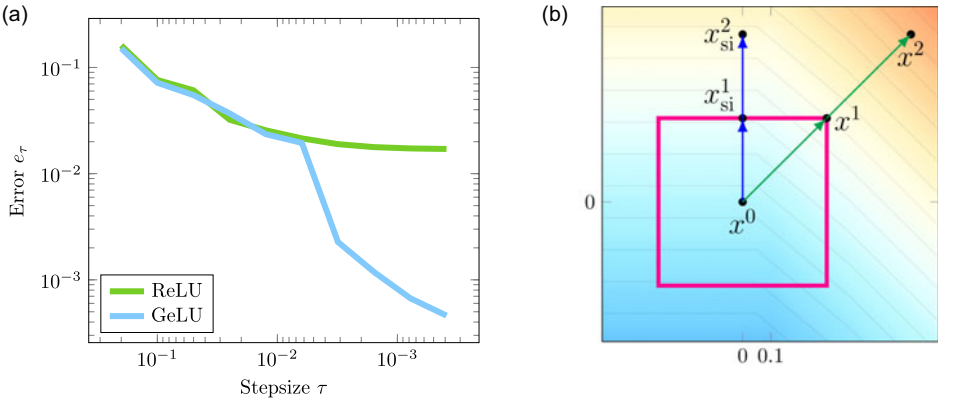
$$x_\tau^{k+1} \in \operatorname*{arg\,min}_{x \in \overline{B_\tau^\infty}(x_\tau^k) \cap \overline{B_\varepsilon^\infty}(x^0)} E(x),$$

in each step. In order to avoid local minima, we do not employ a gradient based method here, but rather a particle based method, which allows exploring the full rectangle $\overline{B_\tau^\infty}(x_\tau^k)$. We use consensus based optimization (CBO) as proposed in [79], using the `CBXPy` package [4]. Concerning the hyperparameters, we choose $N = 30$ particles, a noise scaling of $\sigma = 2$, with standard isotropic noise, a time discretization parameter $\mathrm{d}t = 0.01$, $\alpha = 10^8$ and perform 30 update steps in each inner iteration. In order to ensure the budget constraint and the local restriction given by the step size τ, we project the ensemble of the CBO iteration to the set

$$\overline{B_\varepsilon^\infty}(x^0) \cap \overline{B_\tau^\infty}(x_\tau^k)$$

using the ℓ^∞ projection, i.e., a clipping operation. We refer to [19] for a more detailed numerical study considering projections in CBO schemes, which also suggests the validity of our method here. We repeat the experiment for GeLU with a different initial value $x^0 = (0.45, 0.3)$, $h_{\theta^*}(x^0) \approx 0.74$ and budget $\varepsilon = 0.25$, which is displayed in Figure H3.

https://doi.org/10.1017/S0956792525100120 Published online by Cambridge University Press



Difference between IFGSM and the minimizing movement scheme, as defined in Equation (H.2), for different values of τ, using ReLU and GeLU activation functions in the network architecture.

Minimizing movement scheme (blue arrows) and semi-implicit version (green arrows) for the function $E(x) := -x_2 - \max\{x_1, 0.1\}$ starting from $x^0 = (0, 0)$.

Figure H4. *Difference between IFGSM and the minimizing movement scheme.*

H.3 Convergence of the standard and semi-implicit scheme

In this section, we consider the error between the standard and the semi-implicit minimizing movement, which serves as a very basic validation of the numerical schemes. Our theoretical framework shows that both iterations converge to a ∞-curve of maximum slope, which however is not available numerically. Instead, for $n \in \mathbb{N}$ and $k \leq n$, we can consider

$$\left\| x^k_{\text{si},\tau_n} - x^k_{\tau_n} \right\|_\infty \leq \left\| x^k_{\text{si},\tau_n} - u(k \cdot \tau_n) \right\|_\infty + \left\| x^k_{\tau_n} - u(k \cdot \tau_n) \right\|_\infty ,$$

where $x^k_{\tau_n}$ fulfils the standard minimizing movement scheme and $x^k_{\text{si},\tau_n} = x^k_{\text{IFGS},\tau}$ is given by (IFGSM), i.e., fulfils the semi-implicit scheme. Although our theory does not provide concrete estimates or rates of the error between IFGSM and the minimizing movement scheme, we perform a small numerical experiment using the setup from above. For each choice of τ we sample $S = 50$ different initial values $x^{0,s}$ and compute the iterates $x^{k,s}_{\text{IFGS}}$ and $x^{k,s}_\tau$ for $k \in \{1, \dots, \lfloor 1/\tau \rfloor\}$ and compute the averaged maximal distance

$$e_\tau := \frac{1}{S} \sum_{s=1}^{S} \max_k \left\| x^{k,s}_{\text{IFGS}} - x^{k,s}_\tau \right\|_\infty . \tag{H.2}$$

The errors are plotted in Figure H4a. In both cases, the errors converge to zero; however, we observe that the order of convergence is higher for the GeLU function. We note that our theoretical results only provide a convergence statement for the differentiable case, therefore these results are in line with the analysis. In particular Lemma 3.12 requires a Lipschitz differentiable gradient. However, we hypothesize that the slower convergence in the ReLU case, actually comes from the non-implicit error as visualized in Figure H4b. There we mimic a situation enforced by the ReLU activation function. For $\tau > 0.1$, the minimizing movement scheme always "jumps" across the non-differentiable line $x_1 = 0.1$, to the corner where the minimum on $\overline{B^\infty_\tau}(x^0)$ is attained, which leads the following iterates to follow the gradient into the direction $(1, 1)$. However, in this case the actual flow is given as $u(t) := (t, 0)$, which, in this case, is more accurately prescribed by (FGSM). In this regard, a more exhaustive study, both empirically and theoretically is required, which is left for future work.

Cite this article: Weigand L., Roith T. and Burger M. (2026). Adversarial flows: A gradient flow characterization of adversarial attacks. *European Journal of Applied Mathematics*, **37**, 123–179. https://doi.org/10.1017/S0956792525100120

https://doi.org/10.1017/S0956792525100120 Published online by Cambridge University Press