The Kernel-Density-Estimator Minimizing Movement Scheme

Florentine Catharina Fleißner

The Kernel-Density-Estimator Minimizing Movement Scheme

Florentine Catharina Fleißner *

Abstract

[1] The mathematical theory of a novel variational approximation scheme for general partial differential equations

$$\partial_t u - \nabla \cdot \left(u \nabla \frac{\delta\phi}{\delta u}(u) \left| \nabla \frac{\delta\phi}{\delta u}(u) \right|^{q-2} \right) \;=\; 0, \qquad u \geq 0, \tag{0.1}$$

$q \in (1, +\infty)$, is developed; the Kernel-Density-Estimator Minimizing Movement Scheme (KDE-MM-Scheme) preserves the structure of (0.1) as a steepest descent with regard to an energy functional ϕ and a Wasserstein distance in the space of probability measures, at the same time imitating the corresponding motion of a finite number of particles / data points on a discrete timescale. Roughly speaking, the KDE-MM-Scheme constitutes a simplification of the classical Minimizing Movement scheme for (0.1) (often referred to as 'JKO scheme'), in which the corresponding minimum problems are relaxed and restricted to the values of Kernel Density Estimators each associated with a finite dataset. Rigorous mathematical proofs show that the KDE-MM-Scheme yields solutions to (0.1) if we let the time step sizes and the dataset sizes (particle numbers) simultaneously go to zero and infinity respectively. Uniting abstract analysis in metric spaces with application-orientated concepts from statistics and machine learning, our examinations will form the mathematical foundation for a novel computationally tractable algorithm approximating solutions to (0.1).

A particular ingredient for our analysis is a general and robust stability theory for discrete-time steepest descents under the occurrence of Γ-perturbations of the energy functional in the Minimizing Movement scheme and relaxations of the corresponding minimum problems.

*Technische Universität München, email: fleissne@ma.tum.de.

[1]This paper is a shortened form of the author's arXiv preprint [27] from October 2023.

1 Introduction

The partial differential equation

$$\partial_t u - \nabla \cdot \left(u \nabla \frac{\delta \phi}{\delta u}(u) \left| \nabla \frac{\delta \phi}{\delta u}(u) \right|^{q-2} \right) \ = \ 0, \qquad u \geq 0, \tag{1.1}$$

$q \in (1, +\infty)$, represents a common model for describing the evolution in time of the density u of some quantity in a physical, chemical, biological, ecological or economic process, where total mass is conserved and diffusion is governed by the variational derivative $\frac{\delta \phi}{\delta u}$ of an energy functional ϕ. Let $p \in (1, +\infty)$ be the conjugate exponent of q and $\mathcal{P}_p(\mathbb{R}^d)$ be the space of probability measures with finite moments of order p endowed with the p-Wasserstein distance $\mathcal{W}_p$. In the 90's, Felix Otto originated, together with Jordan and Kinderlehrer [53, 35, 36, 51, 52, 54], the interpretation of dynamics governed by (1.1) as a steepest descent with regard to $\phi : \mathcal{P}_p(\mathbb{R}^d) \to (-\infty, +\infty]$ and $\mathcal{W}_p$, building on a Riemannian formalism ('Otto calculus') and an approximation by discrete-time steepest descents.

Definition 1.1 (Discrete-time steepest descent in $(\mathcal{P}_p(\mathbb{R}^d), \mathcal{W}_p)$)**.** *When a sequence $(\mu_\tau^m)_{m \in \mathbb{N}}$ in $\mathcal{P}_p(\mathbb{R}^d)$ solves the Minimizing Movement (MM) scheme* [2]

$$\mu_\tau^m \quad \textit{is a minimizer for} \quad \Phi(\tau, \mu_\tau^{m-1}, \cdot) \qquad\qquad (m \in \mathbb{N}) \tag{1.2}$$

associated with

$$\Phi(\tau, \mu, \nu) := \phi(\nu) + \frac{1}{p\tau^{p-1}} \mathcal{W}_p(\nu, \mu)^p \tag{1.3}$$

(often referred to as 'JKO scheme') for given time step size $\tau > 0$ and initial datum $\mu_\tau^0 \in \{\phi < +\infty\}$, the corresponding piecewise constant interpolation $\mu_\tau : [0, +\infty) \to \mathcal{P}_p(\mathbb{R}^d)$ defined as

$$\mu_\tau(0) = \mu_\tau^0, \qquad \mu_\tau(t) \equiv \mu_\tau^m \quad \textit{if } t \in ((m-1)\tau, m\tau], \ m \in \mathbb{N}, \tag{1.4}$$

is called discrete solution to (1.2), (1.3) *or discrete-time steepest descent w.r.t. $\phi : \mathcal{P}_p(\mathbb{R}^d) \to (-\infty, +\infty]$.*

---

[2] Introduced by Ennio De Giorgi [19] as "natural meeting point" of many evolution problems from different research fields, the abstract concept of *Minimizing Movements* has proved extremely useful with a wide range of applications in analysis, geometry, physics and numerical analysis, see e.g. [19], the paper [2] by Almgren, Taylor and Wang from which De Giorgi drew his inspiration, [3, 45, 4, 25, 26, 28], [12] and [47, 48].

This paper focuses on a novel approach to transforming (1.2), (1.3) into a computationally tractable MM scheme which does not only offer a simplification of every single step in the scheme but reproduces the dynamics of the whole MM scheme. A general stability theory for the MM scheme (1.2), (1.3) under the occurrence of Γ-perturbations of the energy functional and relaxations of the corresponding minimum problems is established. Building thereon, a rigorous mathematical theory of a novel application-orientated variational approximation scheme for (1.1) is developed: the *Kernel-Density-Estimator Minimizing Movement Scheme* or *KDE-MM-Scheme* for short uses a particular class of Γ-perturbations whose effective domains $\{\phi_n < +\infty\}$ are concentrated in the ranges of Kernel Density Estimation.

1.1 A Robust Stability Theory

Let $\phi_n : \mathcal{P}_p(\mathbb{R}^d) \to (-\infty, +\infty]$ be Γ-perturbations of ϕ in $(\mathcal{P}_p(\mathbb{R}^d), \mathcal{W}_p)$. There exists an a priori correlation $\tau \mapsto n(\tau) \in \mathbb{N}$ between time step sizes and parameters such that discrete-time steepest descents w.r.t. ϕ_n approximate solutions to the diffusion equation (1.1) governed by ϕ whenever the time step sizes $\tau \downarrow 0$ and the corresponding parameters $n = n(\tau) \uparrow +\infty$ simultaneously:

If $(\phi_n)_{n\in\mathbb{N}}$ satisfies mild coercivity conditions, the thorough analysis from the author's paper "Γ-Convergence and Relaxations for Gradient Flows in Metric Spaces: a Minimizing Movement Approach" [25], which is recapitulated in Section 2.2, yields an *a priori choice* $n = n(\tau)$ so that the metric characterization of limit curves of discrete-time steepest descents as solutions to the energy inequality (1.5) (standing below) is *stable* under the occurrence of the Γ-perturbations ϕ_n in the Minimizing Movement scheme (1.2), (1.3). A careful study of the connection between the *metric* and the *differential* characterization of steepest descents in $(\mathcal{P}_p(\mathbb{R}^d), \mathcal{W}_p)$, which is carried out in Section 2.2, then provides a general regularity hypothesis ("chain rule") under which solutions $\mu : [0, +\infty) \to \mathcal{P}_p(\mathbb{R}^d)$ to the energy inequality

$$\phi(\mu(0)) - \phi(\mu(t)) \ \geq \ \frac{1}{q}\int_0^t \big(|\partial^-\phi|(\mu(r))\big)^q \mathrm{d}r \ + \ \frac{1}{p}\int_0^t \big(|\mu'|(r)\big)^p \mathrm{d}r \qquad (1.5)$$

are weak solutions to the partial differential equation (1.1) (with q being the conjugate exponent of p).

Our stability theory still holds true for *approximate* discrete-time steepest descents, i.e. under relaxations of the corresponding minimum problems; we do not need the existence of exact solutions thereto.

We refer the reader to Section 2.1 for the relevant definitions (slope $|\partial^- \phi|$, locally absolutely continuous curve, metric derivative $|\mu'|$) and further information on the energy inequality (1.5). Now, we sketch our theory:

If $(\phi_n)_n$ satisfies mild coercivity conditions, then there exists a correlation $\tau \mapsto n_\tau$ between time step sizes $\tau > 0$ and parameters $n_\tau \in \mathbb{N}$, with $n_\tau \uparrow +\infty$ as $\tau \downarrow 0$, such that the following is true (Thms. 3.4 and 6.1 in [25]):

- Assign a sequence of parameters $n(\tau_k)$ to a given sequence of time step sizes $\tau_k \downarrow 0$ in such a way that $n(\tau_k) \geq n_{\tau_k}$ and set

$$\Phi(\tau_k, \mu, \nu) \ := \ \phi_{n(\tau_k)}(\nu) + \frac{1}{p\tau_k^{p-1}} \mathcal{W}_p(\nu, \mu)^p. \tag{1.6}$$

  Let $\bar{\mu}_{\tau_k}$ be discrete solutions (1.4) to the relaxed MM scheme

$$\Phi(\tau_k, \mu_{\tau_k}^{m-1}, \mu_{\tau_k}^m) \ \leq \ \inf_{\nu \in \mathcal{P}_p(\mathbb{R}^d)} \Phi(\tau_k, \mu_{\tau_k}^{m-1}, \nu) \ + \ \gamma_{\tau_k} \qquad (m \in \mathbb{N}) \tag{1.7}$$

  associated with (1.6) and error terms $\gamma_{\tau_k} > 0$ that are of order $o(\tau_k)$. If the corresponding sequence of initial data $\mu_{\tau_k}^0$ is a 'recovery sequence' for $\mu^0 \in \{\phi < +\infty\}$, then there exists a subsequence of time step sizes $\tau_{k_l} \downarrow 0$ and a locally absolutely continuous curve $\mu : [0, +\infty) \to \mathcal{P}_p(\mathbb{R}^d)$ such that $\mu(0) = \mu^0$,

$$\lim_{l \to +\infty} \sup_{t \in [0,T]} \mathcal{W}_p(\bar{\mu}_{\tau_{k_l}}(t), \mu(t)) \ = \ 0 \qquad \text{for all } T \geq 0$$

  and μ satisfies the energy inequality (1.5) for all $t \geq 0$.

  (see Theorem 2.8(i) and Theorem 2.6)

We fix $n = n(\tau_k) \geq n_{\tau_k}$ a priori. The crucial point is the sequence $(n_\tau)_{\tau > 0}$ is completely independent of initial data and of discrete solutions; it stems from a link between the slope of ϕ and the (p, τ)-Moreau-Yosida difference quotients of ϕ_n that can *always* be established if $(\phi_n)_n$ satisfies mild coercivity conditions and that solely depends on the velocity of Γ-convergence $\phi_n \xrightarrow{\Gamma} \phi$:

- The right a priori choices $n = n(\tau)$ can be precisely determined through relation (2.9) from Theorem 2.6. (see also Remark 2.7, Section 3.2.2)

  We refer the reader to Definition 2.4 for the definition of (p, τ)-Moreau-Yosida difference quotients.

The next step is to link the energy inequality (1.5) to the partial differential equation (1.1). This linkage is well-examined for energy functionals that satisfy a convexity criterion along constant speed geodesics in $(\mathcal{P}_p(\mathbb{R}^d), \mathcal{W}_p)$, cf. Cor. 2.4.10 and Thm. 11.1.3 in [4]. The careful study of the associated underlying structure enables us to introduce a suitable chain rule so that we can establish the linkage between (1.5) and (1.1) for a wider class of energy functionals including functionals that are completely lacking in convexity:

- Let $\mu : [0, +\infty) \to \mathcal{P}_p(\mathbb{R}^d)$ be a solution to the energy inequality (1.5) and let $v : [0, +\infty) \times \mathbb{R}^d \to \mathbb{R}^d$ be its tangent vector field. If $\phi \circ \mu$ belongs to $\mathrm{W}^{1,1}_{\mathrm{loc}}([0, +\infty))$ and satisfies the natural chain rule (2.12) involving v and the limiting subdifferential $\{D_l\phi(\cdot)\}$ of ϕ, then μ is a p-curve of maximal slope w.r.t. ϕ and $|\partial^-\phi|$ in $(\mathcal{P}_p(\mathbb{R}^d), \mathcal{W}_p)$ and solves the differential equation
$$v_t = -|D_l\phi(\mu(t))|^{q-2} D_l\phi(\mu(t)) \qquad \mu_t\text{-a.e.} \tag{1.8}$$
for $\mathcal{L}^1$-a.e. $t > 0$, which is a weak reformulation of (1.1) in the space of probability measures. (see Theorem 2.8(ii) and Remark 2.9)

The definitions of tangent vector field, limiting subdifferential and p-curves of maximal slope are given in Section 2.1. In the author's papers [27, 23], the chain rule is validated for a broad class of energy functionals associated with second order diffusion equations.

Remark 1.2 (A priori choice vs a posteriori choice)**.** Applying the Fundamental Theorem of Γ-convergence and a diagonal argument, we immediately obtain the following statement under suitable coercivity conditions. Suppose that for every time step size $\tau > 0$ and every parameter $n \in \mathbb{N}$, a discrete-time steepest descent $\mu_{\tau,n}$ w.r.t. ϕ_n is given corresponding to discrete values $\mu^m_{\tau,n}$, $m \in \mathbb{N}$, and well-prepared initial data as per Definition 1.1. Then there exist $n = n(\tau) \uparrow +\infty$ and discrete-time steepest descents μ_τ w.r.t. ϕ such that
$$\lim_{\tau \downarrow 0} \mathcal{W}_p(\mu_{\tau,n(\tau)}(t), \mu_\tau(t)) \;=\; 0 \quad \text{for} \quad t \geq 0$$
(Thm. 8.1 in the book [12] by Braides). Though encouraging, this statement is not suited for applications because the proof only yields an *a posteriori* choice $n = n(\tau)$ which heavily depends on the initial data and on the choice of discrete solutions $\mu^m_{\tau,n}$ for every $m \in \mathbb{N}$, $\tau > 0$ and $n \in \mathbb{N}$.

As is explicated above, our theory yields an *a priori choice* $n = n(\tau)$ by contrast, see also Section 2.2 and [25].

Remarks 2.10, 4.1 and 2.11 provide information on extensions of our stability theory concerning a "partial Γ-convergence", the topology and corresponding coercivity conditions, perturbations of the distance term, the chain rule, non-uniform time discretizations and a non-uniform distribution of the error terms.

Please note that the correlation between time step sizes and parameters is crucial because of the lack of control over the slopes of Γ-perturbations, cf. e.g. Sect. 1 and Ex. 1.1 in [25]. In [59, 62, 50], special cases of functionals $\phi_n \xrightarrow{\Gamma} \phi$ are treated in which the corresponding slopes satisfy a Γ-liminf inequality; in these cases *every* choice $n = n(\tau)$ with $n(\tau) \uparrow +\infty$ as $\tau \downarrow 0$ is appropriate, see Sect. 5 in [25] and the end of Section 4.

Our stability theory will form a rigorous mathematical basis for computationally tractable versions of the MM scheme (1.2), (1.3) in general and for our KDE-MM-Scheme in particular.

1.2 Kernel Density Estimation

A function estimator is defined as a random variable with values in some function space $\mathfrak{F}$, emerging from a mapping of independent and identically distributed (i.i.d.) data points $X_1, ..., X_n$ that are drawn from the same but generally unknown probability distribution (cf. e.g. Sect. 5.4 in [32]).

Supposing that $X_1, ..., X_n$ represent an i.i.d. sample from a Borel probability distribution on $\mathbb{R}^d$ having a Lebesgue density function ρ and setting $\mathfrak{F} := \{u \in \mathrm{L}^1(\mathbb{R}^d) \mid u \geq 0, \int_{\mathbb{R}^d} u(x)\mathrm{d}x = 1\}$ (= set of probability density functions), we can try to capture ρ by finding suitable estimators

$$\hat{\rho}_n := f_n(X_1, ..., X_n), \qquad f_n : \left(\mathbb{R}^d\right)^n \to \mathfrak{F}, \qquad n \in \mathbb{N}. \tag{1.9}$$

Therein lies the purpose of general probability density estimation going back to [58, 70, 55]. Rosenblatt [58] and Parzen [55] originated the particular concept of *Kernel Density Estimation*.

Definition 1.3 (Kernel function)**.** *A kernel function is defined as a nonnegative function $\mathcal{K} : \mathbb{R}^d \to [0, +\infty)$ with $\int_{\mathbb{R}^d} \mathcal{K}(x)\mathrm{d}x = 1$.*

Every kernel function $\mathcal{K}$ is associated with a family of functions

$$\mathcal{K}_h : \mathbb{R}^d \to [0, +\infty), \quad \mathcal{K}_h(x) := \frac{1}{h^d}\mathcal{K}\Big(\frac{x}{h}\Big), \quad h > 0. \tag{1.10}$$

The convolution between such scaled kernel function (1.10) and the empirical measure $\frac{1}{n}\sum_{i=1}^{n}\delta_{X_i}$ yields a Kernel Density Estimator for the corresponding probability distribution (which itself is not necessarily absolutely continuous w.r.t. the Lebesgue measure):

Definition 1.4 (Kernel Density Estimator)**.** *Assuming that $X_1, ..., X_n$ is an i.i.d. sample from a probability distribution μ and $\mathcal{K}$ is a kernel function, the function estimator*

$$\hat{\rho}_{n,h} := \frac{1}{n}\sum_{i=1}^{n}\mathcal{K}_h(\cdot - X_i) \tag{1.11}$$

is called Kernel Density Estimator (KDE) for μ associated with the sample size $n \in \mathbb{N}$ and bandwidth $h > 0$.

Kernel Density Estimators are used for example in the wide areas of clustering and topological data analysis with various applications in computer vision, text analysis, biology, chemistry and astronomy including image processing, anomaly detection, unsupervised and semi-supervised classification, genetic profiling and protein analysis, to name but a few (cf. [58, 55, 1, 42, 57, 6, 69, 34, 22, 39] and the references therein).

The asymptotic behaviour of Kernel Density Estimators and their derivatives as the sample sizes $n \uparrow +\infty$ and the bandwidths $h \downarrow 0$ simultaneously is well-documented, including strong consistency results and concentration inequalities that are extremely useful for our purposes; we refer the reader to Section 3.2, Proposition 3.6, Proposition 3.8 and Remarks 3.9 and 3.11 and to [39, 29] and the references therein for further information thereon.

1.3 The KDE-MM-Scheme

The Kernel-Density-Estimator Minimizing Movement Scheme (KDE-MM-Scheme) is developed and introduced as a novel application-orientated variational approximation scheme for (1.1). We fix a kernel function $\mathcal{K}$ according to Definition 1.3 with finite moment of order p (where $p \in (1, +\infty)$ is the conjugate exponent of q), i.e.

$$\mathcal{M}_{\mathcal{K},p} := \int_{\mathbb{R}^d} |x|^p \mathcal{K}(x)\mathrm{d}x < +\infty, \tag{1.12}$$

and we perform the relaxed Minimizing Movement scheme (1.7), (1.6) along a Γ-KDE-Approximation of the energy functional.

Definition 1.5 (Γ-KDE-Approximation). *We say that a sequence of energy functionals $\phi_n : \mathcal{P}_p(\mathbb{R}^d) \to (-\infty, +\infty]$ is a Γ-KDE-Approximation of $\phi : \mathcal{P}_p(\mathbb{R}^d) \to (-\infty, +\infty]$ associated with the kernel function $\mathcal{K}$ and the correlation $n \mapsto h(n)$ between sample sizes $n \in \mathbb{N}$ and bandwidths $h(n) > 0$ if $h(n) \downarrow 0$ as $n \uparrow +\infty$,*

($\mathcal{H}$1) the effective domains $\{\phi_n < +\infty\}$ are concentrated in the KDE ranges corresponding to $\mathcal{K}$, n, $h(n)$, i.e.

$$\phi_n(\mu) < +\infty \quad \Rightarrow \quad \exists y_1, .., y_n \in \mathbb{R}^d : \ \mu = \Big(\frac{1}{n}\sum_{i=1}^{n} \mathcal{K}_{h(n)}(\cdot - y_i)\Big)\mathcal{L}^d,$$

($\mathcal{H}$2) $\phi_n \xrightarrow{\Gamma} \phi$ in $(\mathcal{P}_p(\mathbb{R}^d), \mathcal{W}_p)$, and

($\mathcal{H}$3) whenever $\mu \in \{\phi < +\infty\}$ and $\hat{\rho}_{n,h(n)}$, $n \in \mathbb{N}$, is a sequence of Kernel Density Estimators (1.11) *for μ, the corresponding sequence of measure-valued random variables $\hat{\mu}_{n,h(n)} := \hat{\rho}_{n,h(n)}\mathcal{L}^d$, $n \in \mathbb{N}$, almost surely constitutes a recovery sequence for μ, i.e.*

$$\lim_{n\to+\infty} \mathcal{W}_p(\hat{\mu}_{n,h(n)}, \mu) \ = \ 0 \quad \textit{and} \quad \lim_{n\to+\infty} \phi_n(\hat{\mu}_{n,h(n)}) \ = \ \phi(\mu) \qquad (1.13)$$

with probability 1.

It is reasonable to assume that *the exact form of the initial probability distribution is generally unknown* in practical applications of our theory so that we need to *capture it by an i.i.d. sample.* This is our first motivation for Definition 1.5 and Hypothesis ($\mathcal{H}$3):

- The KDE-MM-Scheme is performed with Kernel Density Estimators (1.11) as initial data corresponding to some initial probability distribution.

Secondly, under ($\mathcal{H}$1), the process of minimization (1.7), (1.6) is restricted to a clearly structured subset of probability densities emerging from $\mathcal{K}$ through basic function operations, which should prove advantageous with regard to a practical implementation of the scheme. Furthermore, a special feature of the KDE-MM-Scheme, which is due to ($\mathcal{H}$1) and the relaxation (1.7) of the minimum problems, is the comparatively easy computation of the distance term:

- The Wasserstein distance term in (1.6) can be replaced by a simple 'particle distance' corresponding to the optimal transport between discrete measures $\frac{1}{n(\tau)} \sum_{i=1}^{n(\tau)} \delta_{y_i}$.

We refer the reader to Definition 3.1 and the instructions from Theorem 3.3 for the resultant KDE-MM-Scheme and to Proposition 3.2 for a consideration of how the differences in the distance terms affect (1.7). Correspondingly adapting the right a priori choices of the parameters $n = n(\tau)$ and $h = h(n(\tau))$, we can directly apply the theory from [25] and Section 2.2, which is outlined in Section 1.1; please recall $(\mathcal{H}3)$ with regard to the initial data. We obtain the strong consistency of the KDE-MM-Scheme as an approximation scheme for continuous-time steepest descents w.r.t. ϕ in $(\mathcal{P}_p(\mathbb{R}^d), \mathcal{W}_p)$ characterized by the partial differential equation (1.1):

- If the Γ-KDE-Approximation satisfies mild coercivity properties, the KDE-MM-Scheme approximates solutions to the energy inequality (1.5) corresponding to any initial probability distribution $\mu^0 \in \{\phi < +\infty\}$ with probability 1. If the chain rule (2.12) holds true, the solutions to (1.5) are p-curves of maximal slope w.r.t. ϕ and $|\partial^- \phi|$ in $(\mathcal{P}_p(\mathbb{R}^d), \mathcal{W}_p)$ and solve the differential equation (1.8), which means that they are weak solutions to (1.1).

As is stated in Section 1.1 and Remark 2.9, the chain rule linking the energy inequality (1.5) to the partial differential equation (1.1) is true whenever ϕ satisfies a convexity criterion along constant speed geodesics in $(\mathcal{P}_p(\mathbb{R}^d), \mathcal{W}_p)$ and moreover, it can be validated for a wider class of energy functionals including functionals that are completely lacking in convexity, cf. [27, 23].

We refer the reader to Theorem 3.3 and Remark 3.4 for the precise a priori selection of the parameters for the KDE-MM-Scheme and the strong consistency / convergence statement.

Please note that the KDE-MM-Scheme both preserves the steepest descent character of the original MM scheme (1.2), (1.3) in the space of probability measures and mimics the corresponding motion of particles / data points.

Variations (partial and weak Γ-KDE-Approximation) on Definition 1.5 and corresponding extensions of the theory are introduced in Remark 3.5 and Remark 4.1(i). Further extensions of the theory concern the chain rule and non-uniform time discretizations and are discussed in Remark 4.1(ii)-(iv).

- The application of the KDE-MM-Scheme to general second order diffusion equations of the form (1.1) is explicated in [27, 23] and sketched in Section 3.2.2. The KDE-MM-Scheme approximates weak solutions to such an equation under natural and quite general hypotheses and can model diverse processes from physics, biology, chemistry etc., see Section 3.2.2, Example 3.7, Remark 3.9 and Remark 3.10.

- The concept of the KDE-MM-Scheme is well suited for approaching fourth order examples of (1.1), too, see Section 3.2.3.

Emphasizing the novelty of both our KDE-MM-Scheme and our stability theory as a general foundation for computationally tractable versions of the MM scheme (1.2), (1.3), we refer the reader to [24] for an investigation of the rich literature on computational approaches to (1.1).

Plan of the paper. Section 2.1 provides preliminary definitions and remarks regarding the analysis of steepest descents. Section 2.2 yields a general and robust stability theory for discrete-time steepest descents in $(\mathcal{P}_p(\mathbb{R}^d), \mathcal{W}_p)$ under the occurrence of Γ-perturbations of the energy functional in the MM scheme (1.2), (1.3) and relaxations of the corresponding minimum problems. In Sections 3.1 and 3.3, the concept of the KDE-MM-Scheme is explicated and its convergence to solutions of (1.1) is proved. Section 3.2 deals with the practical application of the abstract theory of the KDE-MM-Scheme to general second (Section 3.2.2) and fourth (Section 3.2.3) order diffusion equations of the form (1.1); one ingredient of the examinations is the strong consistency of KDEs in $(\mathcal{P}_p(\mathbb{R}^d), \mathcal{W}_p)$ (Section 3.2.1). In Section 4, some aspects and extensions of the stability theory and the theory of the KDE-MM-Scheme are discussed.

2 Stability of Steepest Descents in $(\mathcal{P}_p(\mathbb{R}^d), \mathcal{W}_p)$

2.1 Preliminaries

Let $p \in (1, +\infty)$ and $\mathcal{P}_p(\mathbb{R}^d)$ be the space of Borel probability measures with finite moments of order p (i.e. $\int_{\mathbb{R}^d} |x|^p \mathrm{d}\mu < +\infty$), endowed with the p-Wasserstein distance $\mathcal{W}_p$,

$$\mathcal{W}_p(\mu_1, \mu_2)^p := \min_{\gamma \in \Gamma(\mu_1, \mu_2)} \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^p \mathrm{d}\gamma, \quad \mu_i \in \mathcal{P}_p(\mathbb{R}^d),$$

with $\Gamma(\mu_1, \mu_2)$ being the set of Borel probability measures on $\mathbb{R}^d \times \mathbb{R}^d$ whose first and second marginals coincide with μ_1 and μ_2 respectively (see e.g. [66, 67]). Let $q \in (1, +\infty)$ be the conjugate exponent of p.

The *metric characterization* of steepest descents in $(\mathcal{P}_p(\mathbb{R}^d), \mathcal{W}_p)$ w.r.t. a general energy functional $\phi : \mathcal{P}_p(\mathbb{R}^d) \to (-\infty, +\infty]$ involves an abstraction of the modulus of the gradient to the metric and nonsmooth setting.

Definition 2.1 (Local and relaxed slope)**.** *The local slope $|\partial\phi|$ of ϕ at a probability measure $\mu \in \{\phi < +\infty\}$ is defined as*

$$|\partial\phi|(\mu) := \limsup_{\mathcal{W}_p(\nu,\mu)\to 0} \frac{(\phi(\mu) - \phi(\nu))^+}{\mathcal{W}_p(\mu,\nu)}.$$

The relaxed slope $|\partial^-\phi|$ at $\mu \in \{\phi < +\infty\}$ is a slight modification of the lower semicontinuous envelope of the local slope, i.e.

$$|\partial^-\phi|(\mu) := \inf\{\liminf_{n\to+\infty} |\partial\phi|(\mu_n) \ : \ \lim_{n\to+\infty} \mathcal{W}_p(\mu_n,\mu) = 0,\ \sup_n \phi(\mu_n) < +\infty\}.$$

If ϕ satisfies suitable coercivity conditions and $\mu_{\tau_k} : [0, +\infty) \to \mathcal{P}_p(\mathbb{R}^d)$ is a sequence of discrete-time steepest descents w.r.t. ϕ, initial data $\mu^0_{\tau_k} = \mu^0 \in \{\phi < +\infty\}$ and time step sizes $\tau_k \downarrow 0$ as per Definition 1.1, then there exists a locally uniformly converging subsequence $\mu_{\tau_{k_l}}$ with locally absolutely continuous limit curve $\mu : [0, +\infty) \to \mathcal{P}_p(\mathbb{R}^d)$ solving the energy inequality

$$\phi(\mu(0)) - \phi(\mu(t)) \ \geq \ \frac{1}{q}\int_0^t \big(|\partial^-\phi|(\mu(r))\big)^q \mathrm{d}r \ + \ \frac{1}{p}\int_0^t \big(|\mu'|(r)\big)^p \mathrm{d}r \qquad (2.1)$$

for all $t \geq 0$, see Sect. 2 in [3] and Chaps. 2 and 3 in [4].

Assuming a metric chain rule or a 'continuity condition', the inequality (2.1) can be proved not only for $s = 0$ but for $\mathcal{L}^1$-a.e. $s \in (0, t)$ ("energy dissipation inequality") so that μ is a p-curve of maximal slope w.r.t. ϕ and $|\partial^-\phi|$ in $(\mathcal{P}_p(\mathbb{R}^d), \mathcal{W}_p)$, see Chaps. 1 and 2 in [4], Sects. 2 and 4 in [43] and Sect. 2 in [3]; this abstract concept of a *continuous-time steepest descent in a metric space* originates from De Giorgi, Marino and Tosques [20] and is further developed in [21, 43, 4].

Definition 2.2 (Locally absolutely continuous curve)**.** *We say that a curve $\mu : [0, +\infty) \to \mathcal{P}_p(\mathbb{R}^d)$ is locally absolutely continuous in $(\mathcal{P}_p(\mathbb{R}^d), \mathcal{W}_p)$ if there exists $m \in \mathrm{L}^1_{\mathrm{loc}}(0, +\infty)$ such that*

$$\mathcal{W}_p(\mu(s), \mu(t)) \leq \int_s^t m(r)\mathrm{d}r \quad \textit{for all } 0 \leq s \leq t < +\infty.$$

If $(\mu_t)_{t\geq 0}$ is a locally absolutely continuous curve in $(\mathcal{P}_p(\mathbb{R}^d), \mathcal{W}_p)$, the limit

$$|\mu'|(t) := \lim_{s\to t} \frac{\mathcal{W}_p(\mu(s), \mu(t))}{|s-t|}$$

exists for $\mathcal{L}^1$-a.e. $t \in (0, +\infty)$, the *metric derivative* $t \mapsto |\mu'|(t)$ belongs to $\mathrm{L}^1_{\mathrm{loc}}(0, +\infty)$ and is $\mathcal{L}^1$-a.e. the smallest admissible function m in the definition above (cf. Thm. 1.1.2 in [4]). Moreover, there exists an essentially unique Borel vector field $w : [0, +\infty) \times \mathbb{R}^d \to \mathbb{R}^d$ satisfying both

$$w_t \in \mathrm{L}^p(\mu_t; \mathbb{R}^d), \qquad \|w_t\|_{\mathrm{L}^p(\mu_t;\mathbb{R}^d)} \;=\; |\mu'|(t) \quad \text{for } \mathcal{L}^1\text{-a.e. } t > 0, \tag{2.2}$$

and the *continuity equation*

$$\partial_t \mu_t + \nabla \cdot (w_t \mu_t) \;=\; 0, \tag{2.3}$$

in the distributional sense, i.e.

$$\int_0^{+\infty} \int_{\mathbb{R}^d} (\partial_t \xi(t, x) + \langle \nabla_x \xi(t, x), w(t, x)\rangle) \mathrm{d}\mu_t(x) \mathrm{d}t \;=\; 0$$

for all $\xi \in \mathrm{C}_c^\infty((0, +\infty) \times \mathbb{R}^d)$ (cf. Thm. 8.3.1 and Prop. 8.4.5 in [4]). We refer to w satisfying (2.2) and (2.3) as "tangent" vector field associated with the curve $(\mu_t)_{t\geq 0}$.

The *differential characterization* of limit curves of discrete-time steepest descents as weak solutions to the partial differential equation (1.1) entails a certain regularity of ϕ so that the first variation calculus from [36] can be performed, cf. e.g. [53, 35, 36, 51, 52], [44, 31], Chaps. 10 and 11 in [4], Chap. 8 in [60], Chap. 4 in [61]. In [4], the gradient flow approach to (1.1) is formalized by means of a subdifferential calculus which translates the concept of the Fréchet subdifferential and its weak*-closure from a Banach space into a suitable concept in $(\mathcal{P}_p(\mathbb{R}^d), \mathcal{W}_p)$.

Definition 2.3 (Strong and limiting subdifferential). *If ϕ satisfies*

(ϕ1) $\phi(\mu) < +\infty$ implies $\mu \ll \mathcal{L}^d$,

the strong subdifferential $\partial_s\phi(\mu)$ of ϕ at $\mu \in \{\phi < +\infty\}$ is defined as the set of vector fields $\zeta \in \mathrm{L}^q(\mu; \mathbb{R}^d)$ satisfying

$$\phi(T_\# \mu) - \phi(\mu) \;\geq\; \int_{\mathbb{R}^d} \langle \zeta(x), T(x) - x\rangle \mathrm{d}\mu(x) \;+\; o(\|T - \mathrm{id}\|_{\mathrm{L}^p(\mu;\mathbb{R}^d)}).$$

The limiting subdifferential $\partial_l\phi(\mu)$ of ϕ at $\mu \in \{\phi < +\infty\}$ is defined as the set of vector fields $\zeta \in \mathrm{L}^q(\mu;\mathbb{R}^d)$ for which there exist $\mu_n \xrightarrow{\mathcal{W}_p} \mu$ and $\zeta_n \in \partial_s\phi(\mu_n)$ $(n \in \mathbb{N})$ such that $\sup_n \left\{\phi(\mu_n), \int_{\mathbb{R}^d} |\zeta_n(x)|^q \mathrm{d}\mu_n(x)\right\} < +\infty$ and $\zeta_n\mu_n$ converges to $\zeta\mu$ in the distributional sense.

Assuming $\phi(\mu) := \int \mathcal{F}(x,u)\mathrm{d}x$ or $\phi(\mu) := \int \mathcal{F}(x,u,\nabla u)\mathrm{d}x$ (for $\mu = u\mathcal{L}^d$), it is not difficult to see that Definition 2.3 of $\partial_s\phi(\mu)$ and $\partial_l\phi(\mu)$ generalizes the expression $\nabla\frac{\delta\phi}{\delta u}(u)$ from standard variational calculus for integral functionals to the nonsmooth setting in $(\mathcal{P}_p(\mathbb{R}^d), \mathcal{W}_p)$, cf. Lem. 10.4.1 in [4]. The heuristic principle that

$$\partial_l\phi(\mu) = \left\{\nabla\frac{\delta\phi}{\delta u}(u)\right\}$$

is further substantiated through the concrete computation of limiting subdifferentials, see Example 11.1.9 in [4], which is an example of a general second order diffusion equation, and Sect. 5.3 in [31] and Sect. 2.4 in [44] both dealing with fourth order examples of (1.1).

We refer the reader to Remark 2.9 for further information on steepest descents in $(\mathcal{P}_p(\mathbb{R}^d), \mathcal{W}_p)$ and end this preliminary section by introducing the notion of a difference quotient of $\phi : \mathcal{P}_p(\mathbb{R}^d) \to (-\infty, +\infty]$.

Let $\mathcal{Y}_\tau^{(p)}\phi$ denote the (p,τ)-*Moreau-Yosida approximation* of ϕ defined as

$$\mathcal{Y}_\tau^{(p)}\phi(\mu) := \inf_{\nu\in\mathcal{P}_p(\mathbb{R}^d)} \left\{\phi(\nu) + \frac{1}{p\tau^{p-1}}\mathcal{W}_p(\nu,\mu)^p\right\}.$$

Definition 2.4 ((p,τ)-Moreau-Yosida difference quotient)**.** *Let a probability measure $\mu \in \{\phi < +\infty\}$ and a step size $\tau > 0$ be given. We define*

$$\mathbb{D}^{\mathcal{Y}}_{p,\tau}\phi(\mu) \ := \ \frac{\phi(\mu) - \mathcal{Y}_\tau^{(p)}\phi(\mu)}{\tau} \tag{2.4}$$

and call $\mathbb{D}^{\mathcal{Y}}_{p,\tau}\phi(\mu)$ the (p,τ)-Moreau-Yosida difference quotient of ϕ at μ.

Remark 2.5 (Equivalent definition of $|\partial\phi|$)**.** By Lem. 3.1.5 and Rem. 3.1.7 in [4], the local slope of ϕ at $\mu \in \{\phi < +\infty\}$ satisfies

$$\frac{1}{q}\big(|\partial\phi|(\mu)\big)^q = \limsup_{\tau\downarrow 0} \mathbb{D}^{\mathcal{Y}}_{p,\tau}\phi(\mu)$$

whenever there exist $\tau_\star > 0$ and $\mu_\star \in \mathcal{P}_p(\mathbb{R}^d)$ such that $\mathcal{Y}_{\tau_\star}^{(p)}\phi(\mu_\star) > -\infty$.

2.2 Γ-Convergence for Gradient Flows in $(\mathcal{P}_p(\mathbb{R}^d), \mathcal{W}_p)$: A Minimizing Movement Approach

Let $p,\ q \in (1, +\infty)$ be conjugate exponents. Let $\phi,\ \phi_n : \mathcal{P}_p(\mathbb{R}^d) \to (-\infty, +\infty]$ be energy functionals and suppose that $\phi_n \xrightarrow{\Gamma} \phi$ in $(\mathcal{P}_p(\mathbb{R}^d), \mathcal{W}_p)$, i.e.

$$\phi(\mu) \ \leq \ \liminf_{n\to+\infty} \phi_n(\mu_n) \quad \text{ whenever } \quad \lim_{n\to+\infty} \mathcal{W}_p(\mu_n, \mu) = 0 \tag{2.5}$$

and for all $\mu \in \mathcal{P}_p(\mathbb{R}^d)$ there exists a 'recovery sequence'

$$\exists \ \text{ a sequence } \bar{\mu}_n,\ \lim_{n\to+\infty} \mathcal{W}_p(\bar{\mu}_n, \mu) = 0 : \ \phi(\mu) = \lim_{n\to+\infty} \phi_n(\bar{\mu}_n). \tag{2.6}$$

We *precisely determine* appropriate *a priori* correlations $\tau \mapsto n(\tau)$ so that discrete-time steepest descents w.r.t. ϕ_n approximate continuous-time steepest descents (1.1) w.r.t. ϕ as the time step sizes $\tau \downarrow 0$ and the corresponding parameters $n = n(\tau) \uparrow +\infty$ simultaneously. The statement is even proved for discrete solutions to the *relaxed MM scheme*

$$\Phi(\tau, \mu_\tau^{m-1}, \mu_\tau^m) \ \leq \ \inf_{\nu \in \mathcal{P}_p(\mathbb{R}^d)} \Phi(\tau, \mu_\tau^{m-1}, \nu) \ + \ \gamma_\tau^{(m)} \qquad (m \in \mathbb{N}) \tag{2.7}$$

associated with

$$\Phi(\tau, \mu, \nu) \ := \ \phi_{n(\tau)}(\nu) + \frac{1}{p\tau^{p-1}} \mathcal{W}_p(\nu, \mu)^p, \qquad \gamma_\tau^{(m)} > 0. \tag{2.8}$$

'A priori' means that the choices $n = n(\tau)$ are completely independent of initial data and of discrete solutions to the scheme; they solely depend on the velocity of Γ-convergence $\phi_n \xrightarrow{\Gamma} \phi$. As is stated in Theorems 2.8 and 2.6, right a priori choices $n = n(\tau)$ stem from a connection between the relaxed slope $|\partial^- \phi|$ of ϕ and the (p,τ)-Moreau-Yosida difference quotients $\mathbb{D}^{\mathcal{Y}}_{p,\tau}\phi_n$ of ϕ_n that can *always* be established if $(\phi_n)_{n\in\mathbb{N}}$ satisfies mild coercivity conditions (see Definitions 2.1 and 2.4 for $|\partial^-\phi|$ and $\mathbb{D}^{\mathcal{Y}}_{p,\tau}\phi_n$ respectively):

We assume

(ϕ_n1) there exist $A, B > 0,\ \mu_\star \in \mathcal{P}_p(\mathbb{R}^d)$ s.t. $\phi_n(\cdot) \ \geq \ -A - B\mathcal{W}_p(\cdot, \mu_\star)^p$ for all $n \in \mathbb{N}$,

(ϕ_n2) the combined compactness property

$$\sup_n \{\phi_n(\mu_n), \mathcal{W}_p(\mu_n, \mu_\star)\} < +\infty \quad \Rightarrow \quad \exists n_k \uparrow +\infty, \mu : \ \mathcal{W}_p(\mu_{n_k}, \mu) \to 0.$$

Theorem 2.6 (Existence of right choices $n = n(\tau)$)**.** *Let the sequence of functionals $\phi_n : \mathcal{P}_p(\mathbb{R}^d) \to (-\infty, +\infty]$, $n \in \mathbb{N}$, satisfy (ϕ_n1), (ϕ_n2) and Γ-converge to $\phi : \mathcal{P}_p(\mathbb{R}^d) \to (-\infty, +\infty]$ in $(\mathcal{P}_p(\mathbb{R}^d), \mathcal{W}_p)$.*

Then there exists a sequence $(n_\tau)_{\tau>0}$ in $\mathbb{N}$ with $n_\tau \uparrow +\infty$ as $\tau \downarrow 0$ such that the following holds good whenever the correlation $\tau \mapsto n(\tau)$ between step sizes $\tau > 0$ and parameters $n(\tau) \in \mathbb{N}$ is selected in such a way that $n(\tau) \geq n_\tau$:

$$\frac{1}{q}\big(|\partial^- \phi|(\nu)\big)^q \;\leq\; \Big[\mathfrak{G} - \liminf_{\tau\downarrow 0} \mathbb{D}^{\mathcal{Y}}_{p,\tau}\phi_{n(\tau)}\Big](\nu) \quad \text{for all} \;\; \nu \in \{\phi < +\infty\}, \tag{2.9}$$

where $\Big[\mathfrak{G} - \liminf_{\tau\downarrow 0} \mathbb{D}^{\mathcal{Y}}_{p,\tau}\phi_{n(\tau)}\Big](\nu)$ *is defined as*

$$\inf\Big\{\liminf_{\tau\downarrow 0} \mathbb{D}^{\mathcal{Y}}_{p,\tau}\phi_{n(\tau)}(\nu_\tau) : \; \lim_{\tau\downarrow 0} \mathcal{W}_p(\nu_\tau, \nu) = 0, \; \sup_\tau \phi_{n(\tau)}(\nu_\tau) < +\infty\Big\}. \tag{2.10}$$

Proof. As $(\mathcal{P}_p(\mathbb{R}^d), \mathcal{W}_p)$ is separable (see e.g. Prop. 7.1.5 in [4]), we get Theorem 2.6 by applying Thm. 6.1 from [25], which can be easily extended to the case $p \in (1, +\infty)$. □

Remark 2.7 (Exact computation of right choices $n = n(\tau)$)**.** Whilst Theorem 2.6 shows the existence of right choices $n = n(\tau)$ with regard to (2.9), the detailed examinations in [27, 23] and [25] demonstrate the practical and exact computation thereof.

The mild coercivity conditions (ϕ_n1) and (ϕ_n2) naturally arise from the study of the relaxed MM scheme (2.7), (2.8); they guarantee the existence of discrete solutions to (2.7), (2.8) for $\tau < \big(\frac{1}{pB}\big)^{1/(p-1)}$ and of converging subsequences thereof (cf. Sects. 3.1 and 3.2 in [25]). Since we allow approximate minimizers in (2.7), we do not need to impose any lower semicontinuity or compactness condition on the single functionals ϕ_n.

Theorem 2.8 (Γ-convergence for discrete-time steepest descents)**.** *Let the sequence of functionals $\phi_n : \mathcal{P}_p(\mathbb{R}^d) \to (-\infty, +\infty]$, $n \in \mathbb{N}$, satisfy (ϕ_n1), (ϕ_n2) and Γ-converge to $\phi : \mathcal{P}_p(\mathbb{R}^d) \to (-\infty, +\infty]$ in $(\mathcal{P}_p(\mathbb{R}^d), \mathcal{W}_p)$.*

A given sequence $(\tau_k)_{k\in\mathbb{N}}$ of time step sizes $\tau_k \downarrow 0$ is assigned a sequence of parameters $n(\tau_k) \uparrow +\infty$ in accordance with (2.9)*.*

Let $\bar{\mu}_{\tau_k}$ be discrete solutions (1.4) *to the relaxed MM scheme* (2.7), (2.8), *associated with τ_k, $n(\tau_k)$ and error terms $\gamma^{(m)}_{\tau_k} = \gamma_{\tau_k} > 0$ such that*

$$\lim_{k\to+\infty} \frac{\gamma_{\tau_k}}{\tau_k} \;=\; 0. \tag{2.11}$$

If $\bar{\mu}_{\tau_k}(0)$ is a recovery sequence (2.6) for $(\phi_{n(\tau_k)})_k$, ϕ and some initial datum $\mu^0 \in \{\phi < +\infty\}$, then the following is true:

(i) There exist a subsequence of time step sizes $(\tau_{k_l})_{l\in\mathbb{N}}$, $\tau_{k_l} \downarrow 0$, and a limit curve $\mu : [0, +\infty) \to \mathcal{P}_p(\mathbb{R}^d)$ such that $\mu(0) = \mu^0$ and

$$\lim_{l\to+\infty} \sup_{t\in[0,T]} \mathcal{W}_p(\bar{\mu}_{\tau_{k_l}}(t), \mu(t)) \;=\; 0 \qquad \text{for all } T \geq 0.$$

The curve μ is locally absolutely continuous in $(\mathcal{P}_p(\mathbb{R}^d), \mathcal{W}_p)$ and satisfies the energy inequality (2.1) for all $t \geq 0$.

(ii) Let v denote the tangent vector field of μ. If ϕ satisfies (ϕ1) (from Definition 2.3) and the function $\phi \circ \mu$ belongs to $\mathrm{W}^{1,1}_{\mathrm{loc}}([0, +\infty))$ with $\partial_l\phi(\mu(t)) = \{D_l\phi(\mu(t))\}$ and

$$(\phi \circ \mu)'(t) \;=\; \int_{\mathbb{R}^d} \langle D_l\phi(\mu(t))(x), v_t(x)\rangle \mathrm{d}\mu_t(x) \tag{2.12}$$

for $\mathcal{L}^1$-a.e. $t > 0$, then the curve μ solves both the differential equation

$$v_t \;=\; -|D_l\phi(\mu(t))|^{q-2} D_l\phi(\mu(t)) \qquad \mu_t\text{-a.e.} \tag{2.13}$$

and the energy dissipation equality

$$\phi(\mu(0)) - \phi(\mu(t)) = \frac{1}{q}\int_0^t \left(|\partial^-\phi|(\mu(r))\right)^q \mathrm{d}r + \frac{1}{p}\int_0^t \left(|\mu'|(r)\right)^p \mathrm{d}r \tag{2.14}$$

for $\mathcal{L}^1$-a.e. $t \geq 0$. In this case,

$$\lim_{l\to+\infty} \phi_{n(\tau_{k_l})}(\bar{\mu}_{\tau_{k_l}}(t)) \;=\; \phi(\mu(t)) \tag{2.15}$$

and

$$|\partial^-\phi|(\mu(t)) = \|D_l\phi(\mu(t))\|_{\mathrm{L}^q(\mu_t;\mathbb{R}^d)} \tag{2.16}$$

for $\mathcal{L}^1$-a.e. $t > 0$.

Remark 2.9 (Chain rule)**.** Theorem 2.8 unites three different approaches to the concept of steepest descent in $(\mathcal{P}_p(\mathbb{R}^d), \mathcal{W}_p)$. First, the metric characterization of limit curves of *discrete-time steepest descents* by the energy inequality (2.1) proves stable under the occurrence of Γ-perturbations and relaxations in the corresponding JKO scheme / Minimizing Movement scheme.

The energy dissipation equality (2.14) qualifies the curve μ as a *p-curve of maximal slope* in $(\mathcal{P}_p(\mathbb{R}^d), \mathcal{W}_p)$. Finally, the limiting subdifferential can be identified with $\nabla\frac{\delta\phi}{\delta u}(u)$ (as is outlined in Section 2.1) so that the gradient-flow-type *differential equation* (2.13), together with the continuity equation (2.3), represents a weak reformulation of the diffusion equation (1.1) in the space of probability measures.

The hypothesis that $\phi \circ \mu$ belongs to $\mathrm{W}^{1,1}_{\mathrm{loc}}([0, +\infty))$ satisfying the chain rule (2.12) allows the linkage between all three characterizations of steepest descents in $(\mathcal{P}_p(\mathbb{R}^d), \mathcal{W}_p)$. It arises quite naturally out of the theory of gradient flows; bridging the gap between a metric and a differential characterization of a gradient flow typically involves a chain rule, cf. [[21], Prop. 2.2] and [[43], Thm. 1.11] for the Hilbertian case and [[4], Prop. 1.4.1] for the Banach case. In the case of $(\mathcal{P}_p(\mathbb{R}^d), \mathcal{W}_p)$, the linkage between the differential equation (2.13) and p-curves of maximal slope w.r.t. a displacement convex energy functional is well examined, see Thm. 11.1.3 in [4]; the underlying structure fits into the hypothesis about $\phi \circ \mu$ in part (ii) of Theorem 2.8.

If ϕ is displacement convex (cf. [46], often referred to as convex along constant speed geodesics [4]), then for every solution μ to the energy inequality (2.1), the function $\phi \circ \mu$ belongs to $\mathrm{C}([0, +\infty)) \cap \mathrm{W}^{1,1}_{\mathrm{loc}}([0, +\infty))$ and the chain rule (2.12) holds good for $\mathcal{L}^1$-a.e. $t > 0$, see Cor. 2.4.10, Lem. 10.1.3 and "chain rule" in Sect. 10.1.2 in [4]. It is not difficult to see that the same is true if ϕ satisfies a wider convexity criterion along constant speed geodesics in $(\mathcal{P}_p(\mathbb{R}^d), \mathcal{W}_p)$ as in [[4], Def. 9.1.1] or [[17], Def. 2.4].

Our hypothesis about $\phi \circ \mu$ translates the associated underlying structure into a general concept which also applies to energy functionals ϕ that are lacking in convexity; Theorem 2.8(ii) thus offers a general approach to the equivalence between the characterization (2.1) of limit curves of discrete-time steepest descents, the abstract idea of p-curves of maximal slope and the differential characterization (2.13) of a gradient flow in $(\mathcal{P}_p(\mathbb{R}^d), \mathcal{W}_p)$.

We refer the reader to [27, 23] for the validation of the chain rule (2.12) for a wide class of energy functionals related to second order diffusion equations.

The application of Theorem 2.8(ii) to examples in which the energy functional does not satisfy (ϕ1) or its limiting subdifferentials possibly contain more than one element is discussed in Remark 4.1(ii) and (iii).

We prove Theorem 2.8:

Proof. We refer the reader to Thm. 3.4 in [25] for statement (i); the estimates in the first and second part of the proof therein, which can be easily extended to the case $p \in (1, +\infty)$, show the locally uniform convergence of a subsequence of discrete solutions to a locally absolutely continuous curve satisfying the energy inequality (2.1). The roles of the relation (2.9) between $|\partial^- \phi|$ and $\left(\mathbb{D}^{\mathcal{Y}}_{p,\tau}\phi_{n(\tau)}\right)_{\tau>0}$ and of condition (2.11) on the error terms manifest themselves in that proof.

Let us suppose that ϕ satisfies (ϕ1) and prove (ii). It follows from (ϕ_n1), (ϕ_n2) and the Γ-convergence of ϕ_n to ϕ w.r.t. $\mathcal{W}_p$ that for small $\tau > 0$ and all $\nu \in \mathcal{P}_p(\mathbb{R}^d)$, there exists a solution to the minimum problem

$$\min_{\bar{\nu}\in\mathcal{P}_p(\mathbb{R}^d)} \left\{ \phi(\bar{\nu}) + \frac{1}{p\tau^{p-1}} \mathcal{W}_p(\bar{\nu}, \nu)^p \right\}. \tag{2.17}$$

A simple adaptation of the proof of Lem. 4.6 in [5], an application of Lem. 10.3.4 and Rem. 3.1.7 in [4] then show that

$$|\partial^- \phi|(\nu) < +\infty \qquad \Rightarrow \quad \partial_l \phi(\nu) \neq \emptyset \tag{2.18}$$

and

$$\|D_l\phi(\nu)\|_{\mathrm{L}^q(\nu;\mathbb{R}^d)} \ \leq \ |\partial^- \phi|(\nu) \tag{2.19}$$

if $D_l\phi(\nu)$ is the only element of $\partial_l\phi(\nu)$. By (2.1) and (2.18), the limiting subdifferential $\partial_l\phi(\mu(t))$ is nonempty for $\mathcal{L}^1$-a.e. $t > 0$.

If $\phi \circ \mu$ belongs to $\mathrm{W}^{1,1}_{\mathrm{loc}}([0, +\infty))$ satisfying $\partial_l\phi(\mu(t)) = \{D_l\phi(\mu(t))\}$ and (2.12) for $\mathcal{L}^1$-a.e. $t > 0$, then

$$\phi(\mu(t)) - \phi(\mu(0)) \ \geq \ \int_0^t \int_{\mathbb{R}^d} \langle D_l\phi(\mu(r))(x), v_r(x)\rangle \mathrm{d}\mu_r(x)\mathrm{d}r \tag{2.20}$$

for $\mathcal{L}^1$-a.e. $t > 0$ since $\phi \circ \mu$ is lower semicontinuous at $s = 0$; we infer (2.13), (2.14) and (2.16) from (2.1), (2.20), (2.19), (2.2) and an application of Cauchy-Schwarz inequality and Young's inequality.

Finally, (2.15) follows from the facts that $\lim_{k\to+\infty} \phi_{n(\tau_k)}(\bar{\mu}_{\tau_k}(0)) = \phi(\mu(0))$ and

$$\begin{aligned}
\phi(\mu(0)) - \phi(\mu(t)) \ &\geq \ \limsup_{l\to+\infty} \left[\phi_{n(\tau_{k_l})}(\bar{\mu}_{\tau_{k_l}}(0)) - \phi_{n(\tau_{k_l})}(\bar{\mu}_{\tau_{k_l}}(t))\right] \\
&\geq \ \liminf_{l\to+\infty} \left[\phi_{n(\tau_{k_l})}(\bar{\mu}_{\tau_{k_l}}(0)) - \phi_{n(\tau_{k_l})}(\bar{\mu}_{\tau_{k_l}}(t))\right] \\
&\geq \ \frac{1}{q}\int_0^t \left(|\partial^- \phi|(\mu(r))\right)^q \mathrm{d}r \ + \ \frac{1}{p}\int_0^t \left(|\mu'|(r)\right)^p \mathrm{d}r \\
&= \ \phi(\mu(0)) - \phi(\mu(t))
\end{aligned}$$

by (2.14) and the proof of (2.1) in [25]. □

We present variations on Theorem 2.8 regarding a "partial Γ-convergence" of the energy functionals and condition (2.11) on the error terms $\gamma_\tau^{(m)}$:

Remark 2.10 (Partial Γ-convergence)**.** A *partial* Γ-convergence of ϕ_n to ϕ is sufficient to obtain statement (i) from Theorem 2.8: we need the existence of a recovery sequence (2.6) solely for the initial datum μ^0, cf. Sect. 3.1, Thm. 3.4 in [25]. Also Theorem 2.8(ii) holds true in this case assuming for small $\tau > 0$ and all $\nu \in \{\phi < +\infty\}$ there exists a solution to the minimum problem (2.17) and the function $\phi \circ \mu$ is lower semicontinuous at $s = 0$ so that (2.18), (2.19) and (2.20) are still applicable in its proof.

Remark 2.11 (The error terms)**.** The error order $o(\tau)$ in (2.11) is optimal, cf. Ex. 4.5 in [25]. Further, the proof of Thm. 3.4 in [25] shows that we can extend our theory to a non-uniform distribution of the error terms $\gamma_{\tau_k}^{(m)} > 0$. Theorem 2.8 still holds true if (2.11) is replaced by the general condition

$$\lim_{k\to+\infty} \sum_{m=1}^{N_{\tau_k}} \gamma_{\tau_k}^{(m)} \;=\; 0 \quad \text{for every } N_{\tau_k} \in \mathbb{N} \text{ s.t. } (N_{\tau_k}\tau_k)_{k\in\mathbb{N}} \text{ is bounded,} \tag{2.21}$$

cf. Sect. 3.3 in [25].

Further extensions of Theorem 2.8 regarding a relaxation of the combined compactness condition (ϕ_n2), non-uniform time discretizations and perturbations of the distance term in the relaxed MM scheme (2.7), (2.8) are discussed in Remark 4.1. At the end of Section 4, the main assumptions from the Serfaty-Sandier approach [59, 62] and Ortner's examinations [50] are considered; we can prove (2.9) for every choice $n(\tau) \uparrow +\infty$ in these cases.

3 The KDE-MM-Scheme

Our approach to second and fourth order diffusion equations (1.1) governed by an energy functional ϕ is to carry out the relaxed Minimizing Movement scheme (2.7), (2.8) along a Γ-KDE-Approximation $(\phi_n)_{n\in\mathbb{N}}$ of ϕ. The simple structure of $\{\phi_n < +\infty\}$ as per ($\mathcal{H}$1) in Definition 1.5 brings an advantage for the implementation of the scheme in itself and moreover, it allows a significant simplification of the distance term in (2.8). We call the resultant scheme Kernel-Density-Estimator Minimizing Movement Scheme or KDE-MM-Scheme for short. It constitutes an approximate discrete-time steepest descent movement in $(\mathcal{P}_p(\mathbb{R}^d), \mathcal{W}_p)$ driven by the motion of a finite number of particles / data points in $\mathbb{R}^d$ and converging to solutions of (1.1).

3.1 Definition and Consistency of the KDE-MM-Scheme

Let $p \in (1, +\infty)$ be the conjugate exponent of q from (1.1). We fix a kernel function $\mathcal{K}$ according to Definition 1.3 with finite moment (1.12) of order p.

Definition 3.1 (KDE-MM-Scheme)**.** *Let $(\phi_n)_{n\in\mathbb{N}}$ be a Γ-KDE-Approximation of $\phi : \mathcal{P}_p(\mathbb{R}^d) \to (-\infty, +\infty]$ associated with $\mathcal{K}$ and $n \mapsto h(n)$ according to Definition 1.5. Every time step size $\tau > 0$ is assigned a sample size $n(\tau) \in \mathbb{N}$ and the corresponding bandwidth $h(\tau) := h(n(\tau))$. We define*

$$\Psi(\tau, Y, Z) \ := \ \phi_{n(\tau)}\Big(\frac{1}{n(\tau)}\sum_{i=1}^{n(\tau)} \mathcal{K}_{h(\tau)}(\cdot - z_i)\mathcal{L}^d\Big) \ + \ \frac{1}{p\tau^{p-1}}\sum_{i=1}^{n(\tau)} \frac{|z_i - y_i|^p}{n(\tau)} \quad (3.1)$$

for $Y := \big(y_1, ..., y_{n(\tau)}\big)$ and $Z := \big(z_1, ..., z_{n(\tau)}\big)$ with $y_i,\ z_i \in \mathbb{R}^d$.

For $\tau > 0$ and a given initial datum $Y_\tau^0 := \big(y_{1,\tau}^0, ..., y_{n(\tau),\tau}^0\big)$, $y_{i,\tau}^0 \in \mathbb{R}^d$, our aim is to find a sequence

$$Y_\tau^m := \big(y_{1,\tau}^m, ..., y_{n(\tau),\tau}^m\big), \quad y_{i,\tau}^m \in \mathbb{R}^d, \qquad m \in \mathbb{N} \qquad (3.2)$$

by the scheme

$$\Psi(\tau, Y_\tau^{m-1}, Y_\tau^m) \ \le \ \inf_{Z=(z_1,...,z_{n(\tau)})} \Psi(\tau, Y_\tau^{m-1}, Z) \ + \ \gamma_\tau^{(m)} \qquad (m \ge 1) \quad (3.3)$$

associated with (3.1) and error terms $\gamma_\tau^{(m)} > 0$.

We assign probability measures $\mu_\tau^m := u_\tau^m \mathcal{L}^d \in \mathcal{P}_p(\mathbb{R}^d),\ m \in \mathbb{N}_0$,

$$u_\tau^m \ := \ \frac{1}{n(\tau)}\sum_{i=1}^{n(\tau)} \mathcal{K}_{h(\tau)}(\cdot - y_{i,\tau}^m), \qquad (3.4)$$

to the sequence $(Y_\tau^m)_{m\in\mathbb{N}_0}$; the corresponding piecewise constant interpolations $\mu_\tau : [0, +\infty) \to \mathcal{P}_p(\mathbb{R}^d)$ defined as in (1.4) are called discrete solutions to the KDE-MM-Scheme.

The probabilistic aspect of the KDE-MM-Scheme manifests itself through Hypothesis ($\mathcal{H}3$) in Definition 1.5 and the need for a concrete and feasible recovery sequence corresponding to some initial probability distribution whose exact form is unknown in most cases, cf. Theorem 3.3 below.

If the Γ-KDE-Approximation satisfies the mild coercivity condition ($\phi_n 1$), the relaxed minimum problems (3.3) are solvable and the KDE-MM-Scheme fits into the theory from Section 2.2:

Proposition 3.2 (Basic properties of KDE-MM-Scheme)**.** *Let the* Γ*-KDE-Approximation* $(\phi_n)_{n\in\mathbb{N}}$ *satisfy the coercivity condition* $(\phi_n 1)$*.*

(i) Existence of discrete solutions. *If* $\tau \in \big(0, \big(\frac{1}{pB}\big)^{1/(p-1)}\big)$*, then*

$$\inf_{Z=(z_1,...,z_{n(\tau)})} \Psi(\tau, Y, Z) \ > \ -\infty \qquad \textit{for every} \ \ Y = (y_1, ..., y_{n(\tau)}).$$

(ii) Discrete solutions to the KDE-MM-Scheme are discrete solutions to the relaxed MM scheme (2.7), (2.8). *Let* μ_{τ_k} *be discrete solutions to the KDE-MM-Scheme with time step sizes* $\tau_k \downarrow 0$ *and initial data*

$$\sup_k \{\phi_{n(\tau_k)}(\mu_{\tau_k}(0)), \mathcal{W}_p(\mu_{\tau_k}(0), \mu_\star)\} < +\infty. \tag{3.5}$$

If the corresponding bandwidth parameters $h(\tau_k)$ *and error terms* $\gamma_{\tau_k}^{(m)}$ *satisfy*

$$\lim_{k\to+\infty} \frac{h(\tau_k)}{\tau_k^p} \ = \ 0 \tag{3.6}$$

and (2.21) *respectively, then* μ_{τ_k} *are discrete solutions to* (2.7)*,* (2.8) *associated with error terms* $\bar{\gamma}_{\tau_k}^{(m)} > 0$ *that satisfy* (2.21)*.*

The proof of Proposition 3.2 is postponed to Section 3.3.

An application of Definition 1.5 ($\mathcal{H}3$), Proposition 3.2(ii) and the theory from [25] and Section 2.2 establishes the strong consistency of the KDE-MM-Scheme as an approximation scheme for p-curves of maximal slope w.r.t. ϕ in $(\mathcal{P}_p(\mathbb{R}^d), \mathcal{W}_p)$ that are solutions to (2.13) (weak form of (1.1)):

Theorem 3.3 (Strong consistency of KDE-MM-Scheme)**.** *Let* $(\phi_n)_{n\in\mathbb{N}}$ *be a* Γ*-KDE-Approximation of an energy functional* $\phi : \mathcal{P}_p(\mathbb{R}^d) \to (-\infty, +\infty]$ *according to Definition 1.5. We assume that the coercivity conditions* $(\phi_n 1)$*,* $(\phi_n 2)$ *are satisfied.*

A given sequence $(\tau_k)_{k\in\mathbb{N}}$ *of time step sizes* $\tau_k \downarrow 0$ *is assigned a sequence of sample size parameters* $n(\tau_k) \uparrow +\infty$ *and a corresponding sequence of bandwidth parameters* $h(\tau_k) := h(n(\tau_k)) \downarrow 0$ *in accordance with* (2.9) *and* (3.6)*.*

Let the error terms $\gamma_{\tau_k}^{(m)} > 0$ *satisfy condition* (2.21) *from Remark 2.11.*

If the associated KDE-MM-Scheme is carried out as per Definition 3.1 with initial data $Y_{\tau_k}^0 := \big(X_1, ..., X_{n(\tau_k)}\big)$ *that are formed of an i.i.d. sample* $X_1, ..., X_{n(\tau_k)}$ *from some measure* $\mu^0 \in \{\phi < +\infty\}$*, then with probability* 1*, the statements from Theorem 2.8(i) and (ii) hold true for the corresponding discrete solutions.*

Remark 3.4 (Existence of right parameters for KDE-MM-Scheme)**.** Thanks to Theorem 2.6 and the fact that $h(n) \downarrow 0$ as $n \uparrow +\infty$ by Definition 1.5, we can *always* select $n = n(\tau)$ and $h = h(n(\tau))$ in accordance with (2.9) and (3.6). To be precise, there exists a sequence $(n_\tau)_{\tau>0}$ in $\mathbb{N}$ with $n_\tau \uparrow +\infty$ as $\tau \downarrow 0$ such that the relaxed slope $|\partial^- \phi|$ of ϕ and the (p, τ)-Moreau-Yosida difference quotients $\mathbb{D}^{\mathcal{Y}}_{p,\tau}\phi_{n(\tau)}$ of $\phi_{n(\tau)}$ are connected by (2.9) and the corresponding bandwidths $h(\tau) := h(n(\tau))$ satisfy (3.6) *whenever* the correlation $\tau \mapsto n(\tau)$ between step sizes $\tau > 0$ and sample size parameters $n(\tau) \in \mathbb{N}$ is selected in such a way that $n(\tau) \geq n_\tau$.

In Remark 3.5, an extension of Theorem 3.3 regarding a relaxation of Hypotheses ($\mathcal{H}2$) and ($\mathcal{H}3$) in Definition 1.5 is considered:

Remark 3.5 (Partial Γ-KDE-Approximation)**.** We say that $(\phi_n)_{n\in\mathbb{N}}$ is a partial Γ-KDE-Approximation of ϕ associated with $\mathcal{K}$, the correlation $n \mapsto h(n)$ and a set $\mathcal{I} \subsetneq \{\phi < +\infty\}$ if the bandwidths $h(n) \downarrow 0$ as the sample sizes $n \uparrow +\infty$, Hypothesis ($\mathcal{H}1$) and the Γ-liminf-inequality (2.5) are satisfied and for all $\mu \in \mathcal{I}$, Kernel Density Estimation almost surely yields a recovery sequence (1.13).

Theorem 3.3 still holds true if the KDE-MM-Scheme is performed along a partial Γ-KDE-Approximation with the initial measure μ^0 belonging to the associated set $\mathcal{I}$; we refer the reader to Remark 2.10 for the corresponding extension of Theorem 2.8.

3.2 Applications of the KDE-MM-Scheme

3.2.1 Strong consistency of KDEs in $(\mathcal{P}_p(\mathbb{R}^d), \mathcal{W}_p)$

An essential ingredient for the application of the KDE-MM-Scheme is the strong consistency of Kernel Density Estimators in $(\mathcal{P}_p(\mathbb{R}^d), \mathcal{W}_p)$, cf. Hypothesis ($\mathcal{H}3$) in Definition 1.5.

Proposition 3.6 (Strong consistency of KDEs in $(\mathcal{P}_p(\mathbb{R}^d), \mathcal{W}_p)$)**.** *Let $\mathcal{K}$ be a kernel function according to Definition 1.3 with finite moment* (1.12) *of order p. If $(\hat{\rho}_{n,h(n)})_{n\in\mathbb{N}}$ is a family of Kernel Density Estimators* (1.11) *for a probability distribution $\mu \in \mathcal{P}_p(\mathbb{R}^d)$ associated with $\mathcal{K}$, the sample sizes n and the bandwidths $h(n) \downarrow 0$ as $n \uparrow +\infty$, then $\hat{\mu}_{n,h(n)} := \hat{\rho}_{n,h(n)}\mathcal{L}^d$ satisfies*

$$\lim_{n\to+\infty} \mathcal{W}_p(\hat{\mu}_{n,h(n)}, \mu) \;=\; 0 \qquad \textit{almost surely.}$$

The statement follows from the obvious estimate

$$\mathcal{W}_p\Big(\frac{1}{n}\sum_{i=1}^n \mathcal{K}_h(\cdot - X_i), \frac{1}{n}\sum_{i=1}^n \delta_{X_i}\Big) \;\leq\; h\cdot \mathcal{M}_{\mathcal{K},p}^{1/p}$$

and the strong consistency of empirical measures in $(\mathcal{P}_p(\mathbb{R}^d), \mathcal{W}_p)$. The latter is a consequence of the almost surely weak convergence of empirical measures to the corresponding probability distribution as the sample sizes $n \uparrow +\infty$, see [64], the strong law of large numbers and the equivalence (4.1) between 'weak convergence & convergence of moments of order p' and '$\mathcal{W}_p$-convergence'.

We refer the reader to [29] for estimates of the expected rate of convergence and related concentration inequalities.

3.2.2 Second Order Diffusion Equations

The study of how to apply the theory of the preceding sections (definition of Γ-KDE-Approximation, selection of parameters for KDE-MM-Scheme, validation of chain rule, etc.) to classic second order examples of (1.1) is carried out in every detail in the author's papers [27, 23]. The purpose of this subsection is to give the reader a rough idea thereof.

Example 3.7 (Second order diffusion equation with no-flux boundary condition)**.** Let Ω be an open and bounded subset of $\mathbb{R}^d$, whose boundary $\partial\Omega$ satisfies $\mathcal{L}^d(\partial\Omega) = 0$. As in the preceding sections, q belongs to $(1, +\infty)$ and p denotes its conjugate exponent. In [27, 23], it is proved that the KDE-MM-Scheme, performed as per Definition 3.1 and Theorem 3.3, approximates weak solutions to the partial differential equation

$$\partial_t u - \nabla\cdot\Big(u j_q\Big(\nabla F'(u) + \nabla V + \nabla W * u\Big)\Big) \;=\; 0 \quad \text{in } (0,+\infty)\times\Omega \tag{3.7}$$

with no-flux boundary condition

$$u j_q\Big(\nabla F'(u) + \nabla V + \nabla W * u\Big)\cdot \mathsf{n} \;=\; 0 \quad \text{on } (0,+\infty)\times\partial\Omega, \tag{3.8}$$

where $j_q(v) := |v|^{q-2}v$ for $v \in \mathbb{R}^d$ (with $j_q(0) = 0$). The associated energy functional $\phi : \mathcal{P}_p(\mathbb{R}^d) \to (-\infty, +\infty]$ is defined as

$$\phi(\mu) := \begin{cases} \mathcal{E}(\mu) & \text{if } \mu = u\mathcal{L}^d \text{ and } u \equiv 0 \text{ in } \mathbb{R}^d\setminus\Omega, \\ +\infty & \text{else,} \end{cases}$$

where

$$\mathcal{E}(\mu) := \int_{\mathbb{R}^d} F(u(x))\mathrm{d}x + \int_{\mathbb{R}^d} V\mathrm{d}\mu + \frac{1}{2}\int_{\mathbb{R}^d\times\mathbb{R}^d} W(x-y)u(x)u(y)\mathrm{d}x\mathrm{d}y.$$

Our hypotheses on the functions F, V and W are of a quite general nature being aimed at coercivity properties of the functional $\mathcal{E}$, the characterization of the limiting subdifferential $\partial_l\phi(\mu)$ as $\{\nabla F'(u)+\nabla V+\nabla W*u\}$ and the validation of the chain rule (2.12); they cover not only displacement convex and semi-displacement-convex functionals but also functionals ϕ that do *not* satisfy any convexity criterion along constant speed geodesics in $(\mathcal{P}_p(\mathbb{R}^d), \mathcal{W}_p)$, cf. Remark 2.9.

The natural guess regarding a Γ-KDE-Approximation of ϕ is a sequence of functionals $\phi_n : \mathcal{P}_p(\mathbb{R}^d) \to (-\infty, +\infty]$ defined as

$$\phi_n(\mu) := \begin{cases} \mathcal{E}(\mu) & \text{if } \exists y_1, ..., y_n \in \Omega: \ \mu = \frac{1}{n}\sum_{i=1}^n \mathcal{K}_{h(n)}(\cdot - y_i)\mathcal{L}^d, \\ +\infty & \text{else,} \end{cases}$$

with $h(n) \downarrow 0$ as $n \uparrow +\infty$: in [27, 23], we exemplify the corresponding practical selection of correlations $n \mapsto h(n)$ and $\tau \mapsto n(\tau)$ in accordance with $(\mathcal{H}3)$ and (2.9) and give concrete examples of suitable choices $n = n(\tau)$ and $h = h(n(\tau))$ for an application of Theorem 3.3. We make use of variations on the following statement on the asymptotic behaviour of Kernel Density Estimators proved in [39] (cf. Cor. 15, Thm. 27, Lem. 14, Lem. 11 therein):

Proposition 3.8 (Uniform convergence rates for KDEs)**.** *Let $\mathcal{K}$ be a Lipschitz continuous kernel function according to Definition 1.3 with compact support and $(\hat\rho_{n,h})_{n\in\mathbb{N},h>0}$ an associated family of Kernel Density Estimators* (1.11) *for a probability distribution μ on $\mathbb{R}^d$ with expected value functions*

$$\mathbb{E}[\hat\rho_{n,h}] : \mathbb{R}^d \to [0, +\infty), \qquad x \mapsto \int_{\mathbb{R}^d} \mathcal{K}_h(x-y)\mathrm{d}\mu(y).$$

If μ has compact support, there exists a constant $C_{\mathcal{K},\mu}$ depending on $\mathcal{K}$ and μ such that for every $\alpha \in (0,1)$ and every $n \in \mathbb{N}$ and $h \in (0, \frac{9}{10})$ the following is true with probability at least $1-\alpha$:

$$||\hat\rho_{n,h} - \mathbb{E}[\hat\rho_{n,h}]||_\infty \ \leq \ C_{\mathcal{K},\mu} \cdot \mathfrak{R}(n,h,\alpha) \tag{3.9}$$

where

$$\mathfrak{R}(n,h,\alpha) := \sqrt{\frac{\log(1/h)}{nh^{2d}}} + \sqrt{\frac{\log(1/\alpha)}{nh^{2d}}} + \frac{\log(1/h)}{nh^d} + \frac{\log(1/\alpha)}{nh^d}. \tag{3.10}$$

An application of Proposition 3.8, Borel-Cantelli lemma, the fact that *KDEs for probability densities are almost everywhere asymptotically unbiased* and of the dominated convergence theorem yields suitable choices $h = h(n)$ such that $(\phi_n)_{n\in\mathbb{N}}$ is a Γ-KDE-Approximation of ϕ as per Definition 1.5. Please recall Proposition 3.6 in this context.

The procedure for selecting $n = n(\tau)$ in accordance with the desired relation (2.9) between $|\partial^-\phi|$ and $\big(\mathbb{D}^{\mathfrak{y}}_{p,\tau}\phi_{n(\tau)}\big)_{\tau>0}$ is based on the fact that

$$\frac{1}{q}\big(|\partial^-\phi|(\nu)\big)^q \;\le\; \Big[\mathfrak{G} - \liminf_{\tau\downarrow 0}\mathbb{D}^{\mathfrak{y}}_{p,\tau}\phi\Big](\nu) \quad \text{for all} \;\; \nu \in \{\phi < +\infty\},$$

where $\Big[\mathfrak{G} - \liminf_{\tau\downarrow 0}\mathbb{D}^{\mathfrak{y}}_{p,\tau}\phi\Big](\nu)$ is defined as

$$\inf\Big\{\liminf_{\tau\downarrow 0}\mathbb{D}^{\mathfrak{y}}_{p,\tau}\phi(\nu_\tau) : \; \lim_{\tau\downarrow 0}\mathcal{W}_p(\nu_\tau,\nu) = 0, \; \sup_{\tau}\phi(\nu_\tau) < +\infty\Big\},$$

(cf. Prop. 4.1 in [25] and Theorem 2.6) and on a precise examination and quantification of the differences between the (p,τ)-Moreau-Yosida difference quotients $\mathbb{D}^{\mathfrak{y}}_{p,\tau}\phi$ and $\mathbb{D}^{\mathfrak{y}}_{p,\tau}\phi_n$ corresponding to a fixed time step size τ and a sample size n. For that, we refine the uniform convergence rate (3.9), (3.10) for KDEs taking advantage of the proofs from [39] and the underlying information on the constant $C_{\mathcal{K},\mu}$.

Furthermore, if realistic initial probability densities and solutions corresponding to an application of (3.7), (3.8) in physics, biology, etc. are essentially bounded from above, the concept of *partial* Γ-KDE-Approximation from Remark 3.5 enables us to take account of this extra condition a priori.

We refer the reader to [27, 23] for all details.

Remark 3.9 (Second order diffusion equation: unbounded domain)**.** It is noteworthy that in [39] (cf. Sect. 4 therein) uniform convergence rates such as (3.9) are also established for probability distributions with unbounded support and a wider class of kernel functions (involving more parameters than in Proposition 3.8).

Whilst for Example 3.7 and corresponding Γ-KDE-Approximations the combined compactness property $(\phi_n 2)$ easily follows from Prokhorov's theorem because the PDE domain is bounded, it is typically an obstacle if $\Omega = \mathbb{R}^d$ (the case that the function V, by which the external potential term of ϕ is defined, satisfies $\frac{V(x)}{|x|^p} \uparrow +\infty$ for $|x| \uparrow +\infty$ is an exception); we tackle this obstacle by *weak* and *partial weak* Γ-KDE-Approximations, for which the parameter selection is similar to the above procedure, see Remark 4.1(i).

We end this subsection with a brief account of physical, biological, chemical, etc. processes that can be simulated by the second order diffusion equation from Example 3.7 and the corresponding KDE-MM-Scheme:

Remark 3.10 (Applications of (3.7), (3.8))**.** The general second order diffusion equation (3.7) with no-flux boundary condition (3.8) has a wide range of applications. We name but a few:

- models for studying the motion of a gas in a porous medium, the motion of particles under chemical bonding forces, the motion of a population or the evolution in time of a region occupied by water where groundwater infiltration through a porous stratum occurs (cf. [65, 18, 41, 49, 15, 40, 30, 33, 11])
- the theory of shallow-water flows, e.g. atmospheric flows, tidal flows, storm surges, river flows, tsunamis (cf. [68, 16] and the references therein),
- the theory of heat conduction and heat radiation in plasmas, (cf. [71, 65, 72, 9]),
- the theory of granules affected by their environment, friction and inelastic collisions between granules with different velocities (cf. [8, 7, 63, 13, 14]).

3.2.3 Fourth Order Diffusion Equations

The abstract theory from Section 3.1, Proposition 3.6 and uniform convergence rates for KDE derivatives provide good reasons for studying the KDE-MM-Scheme as a variational approximation technique for fourth order examples of (1.1) from a theoretical and an experimental point of view.

Remark 3.11 (Uniform convergence rates for KDE derivatives)**.** If the kernel function $\mathcal{K}$ belongs to $\mathrm{C}_{\mathrm{c}}^{1,1}(\mathbb{R}^d)$, convergence rates similar to (3.9), (3.10) can be proved for $\left\|\nabla\hat{\rho}_{n,h} - \nabla\mathbb{E}[\hat{\rho}_{n,h}]\right\|_{\infty}$, see Sect. 6 in [39]. Please note that

$$\nabla\mathbb{E}[\hat{\rho}_{n,h}](x) = \int_{\mathbb{R}^d} \nabla\mathcal{K}_h(x-y)\mathrm{d}\mu(y)$$

and $\nabla\mathbb{E}[\hat{\rho}_{n,h}] = \mathcal{K}_h * \nabla\rho$ if μ has a Lebesgue density function ρ with locally integrable weak gradient $\nabla\rho$.

Presumably the best-known fourth order examples of (1.1) (for $q = 2$) are the thin film equation

$$\partial_t u + \nabla \cdot \big(u \nabla \Delta u\big) \;=\; 0 \qquad \Big(\phi(u) := \frac{1}{2}\int |\nabla u|^2 \mathrm{d}x\Big),$$

applied to lubrication theory for describing the motion of a moving contact line (cf. e.g. [10]), and the quantum drift diffusion equation

$$\partial_t u + 4\nabla \cdot \big(u \nabla \frac{\Delta\sqrt{u}}{\sqrt{u}}\big) \;=\; 0 \qquad \Big(\phi(u) := \int \frac{|\nabla u|^2}{u} \mathrm{d}x\Big),$$

in which u stands e.g. for the electron density in a quantum model for semiconductors (cf. [56, 37, 38]).

3.3 Proof of Proposition 3.2

We prove Proposition 3.2.

Proof. An application of Young's inequality and condition ($\phi_n 1$) show that for all $\epsilon \in \big(0, \frac{1}{p}\big)$ there exists a constant $\tilde{C}_\epsilon > 0$ such that

$$\phi_n(\cdot) \;\geq\; -A - B\mathcal{W}_p(\cdot, \mu_\star)^p \;\geq\; -A - \frac{B}{1-p\epsilon}\mathcal{W}_p(\cdot, \tilde{\mu})^p - \tilde{C}_\epsilon \cdot \mathcal{W}_p(\tilde{\mu}, \mu_\star)^p \quad (3.11)$$

for all $\tilde{\mu} \in \mathcal{P}_p(\mathbb{R}^d)$ and $n \in \mathbb{N}$. Statement (i) follows from (3.11) and the fact that

$$\mathcal{W}_p\Big(\frac{1}{n}\sum_{i=1}^n \mathcal{K}_{h(n)}(\cdot - y_i)\mathcal{L}^d, \frac{1}{n}\sum_{i=1}^n \mathcal{K}_{h(n)}(\cdot - z_i)\mathcal{L}^d\Big)^p \;\leq\; \frac{1}{n}\sum_{i=1}^n |y_i - z_i|^p \quad (3.12)$$

for every $y_i, z_i \in \mathbb{R}^d$, which can be easily seen by testing the minimum problem in the definition of $\mathcal{W}_p$ on the measure $\frac{1}{n}\sum_{i=1}^n (\mathrm{id} \times T_i)_\#(\mathcal{K}_{h(n)}(\cdot - y_i)\mathcal{L}^d)$, $T_i(x) := z_i + x - y_i$.

Now, we prove statement (ii). The differences in the distance terms of the relaxed Minimizing Movement scheme (2.7), (2.8) and the KDE-MM-Scheme can be estimated in the following way.

Let $Y^m_{\tau_k} := \big(y^m_{1,\tau_k}, ..., y^m_{n(\tau_k),\tau_k}\big)$, $y^m_{i,\tau_k} \in \mathbb{R}^d$, be a solution (3.2) of one step (3.3) of the KDE-MM-Scheme and $\mu^m_{\tau_k} = u^m_{\tau_k}\mathcal{L}^d$ be the associated measure (3.4); $u^m_{\tau_k}$ equals the convolution between the scaled kernel function $\mathcal{K}_{h(\tau_k)}$

and the measure $\frac{1}{n(\tau_k)}\sum_{i=1}^{n(\tau_k)}\delta_{y^m_{i,\tau_k}}$ (where δ_y denotes the Dirac measure with centre $y\in\mathbb{R}^d$) so that

$$\mathcal{W}_p(\mu^m_{\tau_k},\mu^{m-1}_{\tau_k})^p \;\le\; \frac{1}{n(\tau_k)}\sum_{i=1}^{n(\tau_k)}|y^m_{i,\tau_k}-y^{m-1}_{i,\tau_k}|^p \tag{3.13}$$

according to (3.12). Estimate (3.13) is the first step towards the inequality

$$\Phi(\tau_k,\mu^{m-1}_{\tau_k},\mu^m_{\tau_k}) \;\le\; \inf_{\nu\in\mathcal{P}_p(\mathbb{R}^d)}\Phi(\tau_k,\mu^{m-1}_{\tau_k},\nu) \;+\; \bar{\gamma}^{(m)}_{\tau_k} \tag{3.14}$$

for Φ defined as in (2.8) and a suitable error term $\bar{\gamma}^{(m)}_{\tau_k}>0$.

Moreover, we apply the change of variables formula and Jensen's inequality to $\mathcal{W}_p(\mu_{\tau_k}(0),\delta_0)=\Big(\int_{\mathbb{R}^d}|x|^p\mathrm{d}\mu_{\tau_k}(0)\Big)^{1/p}$ so that we have

$$\mathcal{W}_p(\mu_{\tau_k}(0),\delta_0) \;\ge\; \int_{\mathbb{R}^d}\Big(\frac{1}{n(\tau_k)}\sum_{i=1}^{n(\tau_k)}|h(\tau_k)z+y^0_{i,\tau_k}|^p\Big)^{1/p}\mathcal{K}(z)\mathrm{d}z$$

and using Minkowski's inequality for sequences, we obtain

$$\mathcal{W}_p(\mu_{\tau_k}(0),\delta_0) \;\ge\; \Big(\frac{1}{n(\tau_k)}\sum_{i=1}^{n(\tau_k)}|y^0_{i,\tau_k}|^p\Big)^{1/p} \;-\; h(\tau_k)\cdot C, \tag{3.15}$$

where $C:=\int_{\mathbb{R}^d}|z|\mathcal{K}(z)\mathrm{d}z\le\mathcal{M}^{1/p}_{\mathcal{K},p}<+\infty$ (cf. (1.12)). We may assume w.l.o.g. that $\tau_k\in\big(0,\big(\frac{1}{pB}\big)^{1/(p-1)}\big)$. For all $k,m\in\mathbb{N}$ there exists a constant $R_{k,m}>0$ only depending on the initial data and the error terms and not on the discrete solutions themselves such that

$$\Big(\frac{1}{n(\tau_k)}\sum_{i=1}^{n(\tau_k)}|y^m_{i,\tau_k}|^p\Big)^{1/p} \;\le\; R_{k,m} \tag{3.16}$$

and

$$\sup\{R_{k,m}:\ k,m \text{ s.t. } m\tau_k\le T\} \;<\; +\infty \qquad \text{for every} \qquad T>0; \tag{3.17}$$

in actual fact, it is not difficult to deduce (3.16), (3.17) from the first part of the proof of Thm. 3.4 in [25], (2.21), (3.15), (3.5) and an application of

(3.11) with $\tilde{\mu} := \mathcal{K}_{h(\tau_k)}(\cdot)\mathcal{L}^d$ and fixed ϵ in combination with (3.12) and the fact that $\mathcal{W}_p(\mathcal{K}_{h(\tau_k)}(\cdot)\mathcal{L}^d, \mu_\star) \leq h(\tau_k)\mathcal{M}_{\mathcal{K},p}^{1/p} + \mathcal{W}_p(\delta_0, \mu_\star)$.

Further, we have

$$\mathcal{W}_p(\mu_{\tau_k}^{m-1}, \delta_0) \;\leq\; \Big(\frac{1}{n(\tau_k)} \sum_{i=1}^{n(\tau_k)} |y_{i,\tau_k}^{m-1}|^p\Big)^{1/p} + \underbrace{\mathcal{W}_p(\mathcal{K}_{h(\tau_k)}(\cdot)\mathcal{L}^d, \delta_0)}_{\leq h(\tau_k)\cdot \mathcal{M}_{\mathcal{K},p}^{1/p}}$$

by (3.12), (1.12) and

$$\phi_{n(\tau_k)}(\mu_{\tau_k}^{m-1}) \;\leq\; \phi_{n(\tau_k)}(\mu_{\tau_k}(0)) + \sum_{j=1}^{m-1} \gamma_{\tau_k}^{(j)}$$

by (3.3). Consequently, the first part of the proof of Thm. 3.4 in [25], (3.16), (3.17), (3.5), (2.21), ($\phi_n 1$) and an estimate analogous to (3.15) show that for all $k, m \in \mathbb{N}$ there exists a constant $\bar{R}_{k,m} > 0$ such that (3.17) holds true for $\bar{R}_{k,m}$ and

$$\Big(\frac{1}{n(\tau_k)} \sum_{i=1}^{n(\tau_k)} |y_i|^p\Big)^{1/p} \;\leq\; \bar{R}_{k,m} \tag{3.18}$$

whenever $\nu := \Big(\frac{1}{n(\tau_k)} \sum_{i=1}^{n(\tau_k)} \mathcal{K}_{h(\tau_k)}(\cdot - y_i)\Big)\mathcal{L}^d$ satisfies

$$\phi_{n(\tau_k)}(\nu) + \frac{1}{p\tau_k^{p-1}} \mathcal{W}_p(\nu, \mu_{\tau_k}^{m-1})^p \;\leq\; \inf_{\bar{\nu} \in \mathcal{P}_p(\mathbb{R}^d)} \Phi(\tau_k, \mu_{\tau_k}^{m-1}, \bar{\nu}) \;+\; 1$$

for Φ defined as in (2.8). Obviously (see e.g. Lem. 7.1.10 in [4]), we have

$$\mathcal{W}_p\Big(\nu,\; \frac{1}{n(\tau_k)} \sum_{i=1}^{n(\tau_k)} \delta_{y_i}\Big)^p \;\leq\; h(\tau_k)^p \cdot \mathcal{M}_{\mathcal{K},p}. \tag{3.19}$$

As the measure ν is independent of the numbering order of $y_1, ..., y_{n(\tau_k)}$, we may assume w.l.o.g. that

$$\mathcal{W}_p\Big(\frac{1}{n(\tau_k)} \sum_{i=1}^{n(\tau_k)} \delta_{y_i} \;,\; \frac{1}{n(\tau_k)} \sum_{i=1}^{n(\tau_k)} \delta_{y_{i,\tau_k}^{m-1}}\Big)^p \;=\; \frac{1}{n(\tau_k)} \sum_{i=1}^{n(\tau_k)} |y_i - y_{i,\tau_k}^{m-1}|^p$$

(cf. pp. 5-6 in [66]). We obtain

$$\frac{1}{n(\tau_k)} \sum_{i=1}^{n(\tau_k)} |y_i - y_{i,\tau_k}^{m-1}|^p \;\leq\; \mathcal{W}_p(\nu, \mu_{\tau_k}^{m-1})^p \;+ 2p \cdot h(\tau_k) \cdot \mathcal{M}_{\mathcal{K},p}^{1/p} \cdot \tilde{R}_{k,m}^{p-1}, \tag{3.20}$$

with $\tilde{R}_{k,m} := \bar{R}_{k,m} + R_{k,m-1}$, by applying the estimate (3.19) to both ν and $\mu_{\tau_k}^{m-1}$, (3.16), (3.18), the triangle inequality, Young's inequality and Minkowski's inequality for sequences.

All in all, we infer from (3.13), (3.3), (3.20), the fact that the constants $\tilde{R}_{k,m}$ satisfy (3.17) and from (3.6) that the measures $\mu_{\tau_k}^m = u_{\tau_k}^m \mathcal{L}^d$, $m \in \mathbb{N}$, solve the successive relaxed minimum problems (3.14) and the corresponding error terms

$$\bar{\gamma}_{\tau_k}^{(m)} := \gamma_{\tau_k}^{(m)} + 2\mathcal{M}_{\mathcal{K},p}^{1/p} \cdot \tilde{R}_{k,m}^{p-1} \cdot \frac{h(\tau_k)}{\tau_k^{p-1}}$$

satisfy condition (2.21).

The proof of Proposition 3.2 is complete. □

4 Remarks and Extensions of the Theory

In Sections 1.3 and 3, advantages of the KDE-MM-Scheme regarding a computational implementation are explicated. Please note that such a process always involves a second stage besides the mathematical theory, in which part of the mathematical accuracy is carefully sacrificed for computational feasibility and cost economy. Corresponding further simplifications of the KDE-MM-Scheme are beyond the scope of this introductory paper.

We present variations on Theorem 2.8 and Theorem 3.3 with respect to the hypotheses on the energy functionals, the topology and the time discretizations:

Remark 4.1 (Some extensions of Theorem 2.8 and Theorem 3.3)**.** We discuss extensions of our theory to the case that $(\phi_n)_{n\in\mathbb{N}}$ does not satisfy the combined compactness property ($\phi_n 2$), the case that ϕ does not satisfy ($\phi 1$) or the hypotheses on its limiting subdifferential from Theorem 2.8(ii) and to the cases of non-uniform time discretizations and perturbations of the distance term in the general relaxed Minimizing Movement scheme (2.7), (2.8) and the KDE-MM-Scheme (3.3), (3.1):

(i) Using the topology induced by weak convergence $\nu_n \rightharpoonup \nu$ defined as

$$\lim_{n\to+\infty} \int_{\mathbb{R}^d} f(x)\mathrm{d}\nu_n(x) = \int_{\mathbb{R}^d} f(x)\mathrm{d}\nu(x) \qquad \text{for all } f \in \mathrm{C}_b(\mathbb{R}^d)$$

as an auxiliary topology besides the metric topology $(\mathcal{P}_p(\mathbb{R}^d), \mathcal{W}_p)$, we can omit the combined compactness condition ($\phi_n 2$) because by

Prokhorov's Theorem, $\mathcal{W}_p$-bounded sets are relatively compact w.r.t. weak convergence; Theorem 2.8 still holds true (with pointwise weak instead of locally uniform $\mathcal{W}_p$-convergence of the discrete solutions) if the functionals ϕ_n satisfy $(\phi_n 1)$ and $\mathcal{W}_p(\mu_n, \mu) \to 0$ is replaced with $\sup_n \mathcal{W}_p(\mu_n, \mu) < +\infty,\ \mu_n \rightharpoonup \mu$ in the Γ-liminf inequality (2.5) ('weak Γ-liminf inequality') and in the definitions of $|\partial^- \phi|(\mu)$ and $\partial_l \phi(\mu)$ (cf. Definitions 2.1 and 2.3) and if we adapt the parameter selection (2.9) by correspondingly changing the definition (2.10) of the slight modification $\left[\mathfrak{G} - \liminf_{\tau \downarrow 0} \mathbb{D}^{\mathcal{Y}}_{p,\tau} \phi_{n(\tau)}\right]$ of the Γ-lower limit of the (p,τ)-Moreau-Yosida difference quotients of $\phi_{n(\tau)}$, see Sects. 2, 3.1, 3.2 and 4.1 in [25].

We say that $(\phi_n)_{n \in \mathbb{N}}$ is a weak Γ-KDE-Approximation of ϕ associated with $\mathcal{K}$ and $n \mapsto h(n)$ if $h(n) \downarrow 0$ as $n \uparrow +\infty$, Hypothesis $(\mathcal{H}1)$, the weak Γ-liminf inequality and Hypothesis $(\mathcal{H}3)$ are satisfied; please recall Proposition 3.6 in this context.

Theorem 3.3 still holds true according to the above extension of Theorem 2.8 if the KDE-MM-Scheme is performed along a weak Γ-KDE-Approximation and we only assume the coercivity condition $(\phi_n 1)$. The same is true of a partial weak Γ-KDE-Approximation, whose definition is obvious, cf. Remark 3.5.

Please note that

$$\nu_n \xrightarrow{\mathcal{W}_p} \nu \quad \Leftrightarrow \quad \nu_n \rightharpoonup \nu, \ \lim_{n \to +\infty} \int_{\mathbb{R}^d} |x|^p \mathrm{d}\nu_n(x) = \int_{\mathbb{R}^d} |x|^p \mathrm{d}\nu(x) \quad (4.1)$$

(see e.g. Thm. 7.12 in [66]).

(ii) We assume $(\phi 1)$ only for the sake of a clear presentation with a straightforward notation. Theorem 2.8(ii) can be generalized to the case that $\phi(\mu) < +\infty$ does not necessarily imply $\mu \ll \mathcal{L}^d$. The subdifferential calculus from Definition 2.3 has to be adapted for this case, see Sect. 10.3 and Def. 10.3.1 in [4], and the chain rule (2.12) has to be reformulated correspondingly.

(iii) The hypothesis in part (ii) of Theorem 2.8 that the limiting subdifferential $\partial_l \phi(\mu(t))$ contains at most one element seems none too restrictive, cf. Section 2.1; however, we may assume instead that for $\mathcal{L}^1$-a.e. $t > 0$, the limiting subdifferential $\partial_l\phi(\mu(t))$ contains an element $D_l\phi(\mu(t))$ satisfying (2.19). For that, it suffices to assume one of the following two properties: the "minimal selection" criterion

$$\|D_l\phi(\mu(t))\|_{\mathrm{L}^q(\mu_t;\mathbb{R}^d)} = \min\{\|\zeta\|_{\mathrm{L}^q(\mu_t;\mathbb{R}^d)} : \ \zeta \in \partial_l\phi(\mu(t))\}$$

or the condition that relaxed and local slope coincide, i.e. $|\partial^-\phi| \equiv |\partial\phi|$, cf. Lem. 10.3.4 and Rem. 3.1.7 in [4] and the proofs of Lem. 4.6 in [5] and Lem. 10.1.5 in [4].

(iv) The theory from Sections 2.2 and 3 can be easily extended to time discretizations with non-equi-sized time steps. Every time discretization $\mathcal{T} = \{0 = t_0, t_1, ..\}$ with $t_j \uparrow +\infty$ and $\underline{\tau}(\mathcal{T}) := \inf_{j\in\mathbb{N}}(t_j - t_{j-1}) > 0$ is assigned a parameter / sample size $n(\mathcal{T})$ following the selection procedure corresponding to the smallest time step size $\underline{\tau}(\mathcal{T})$ of $\mathcal{T}$, cf. Theorem 2.6, Remark 3.4 and Section 3.3. Letting $n(\mathcal{T}) \uparrow +\infty$ and $\overline{\tau}(\mathcal{T}) := \sup_{j\in\mathbb{N}}(t_j - t_{j-1}) \downarrow 0$ simultaneously in the associated schemes, we obtain the same statements as in Theorems 2.8 and 3.3.

(v) It is possible to allow for perturbations not only of the energy functional but also of the distance term in (2.8) and to extend the theory from Theorem 2.8 to this case, cf. Sect. 9 in [25].

Finally, we note that the key to our stability theory for steepest descents under Γ-convergence from [25] and Section 2.2 is to focus on the discrete-time steepest descents w.r.t. the Γ-perturbations ϕ_n, on the (p, τ)-Moreau-Yosida difference quotients of ϕ_n and the interplay between parameters and time step sizes as $n \uparrow +\infty$ and $\tau \downarrow 0$ simultaneously, thus compensating for the lack of control over the slopes of ϕ_n.

As is explicated in Sect. 1 in [25], the limit of a sequence of continuous-time steepest descents w.r.t. ϕ_n, $n \in \mathbb{N}$, is in general *not* related to the continuous-time gradient flow of the Γ-limit functional ϕ because the slopes of ϕ_n are not related to the slope of ϕ; in terms of the Minimizing Movement scheme, considerations regarding the limit behaviour of continuous-time steepest descents correspond to first letting the time step sizes $\tau \downarrow 0$ for a fixed parameter $n \in \mathbb{N}$ and only then letting the parameters $n \uparrow +\infty$.

The well-known Serfaty-Sandier approach [59, 62] offers the underlying structure of special cases $\phi_n \xrightarrow{\Gamma} \phi$ in which continuous-time gradient flow solutions w.r.t. ϕ_n do converge to the continuous-time gradient flow w.r.t. ϕ;

the Serfaty-Sandier theory relies on the assumption that the corresponding relaxed slopes satisfy a Γ-liminf inequality

$$\mu_n \to \mu \quad \Rightarrow \quad |\partial^- \phi|(\mu) \le \liminf_{n\to+\infty} |\partial^- \phi_n|(\mu_n).$$

Ortner's examination [50] of a joint Minimizing Movement scheme along $\phi_n \xrightarrow{\Gamma} \phi$ is based on the assumption that

$$\mu_n \to \mu \quad \Rightarrow \quad |\partial \phi|(\mu) \le \liminf_{n\to+\infty} |\partial \phi_n|(\mu_n).$$

In both cases, the relation (2.9) between the relaxed slope $|\partial^- \phi|$ of ϕ and the (p,τ)-Moreau-Yosida difference quotients $\mathbb{D}^{\mathcal{Y}}_{p,\tau}\phi_{n(\tau)}$ of $\phi_{n(\tau)}$ holds true for *every* choice $n = n(\tau) \uparrow +\infty$, cf. Sect. 5, Prop. 5.1, Prop. 5.2 and Sect. 9 in [25]. The same is true if the topology induced by weak convergence is used as an auxiliary topology as in Remark 4.1(i).

References

[1] C. Abraham, G. Biau, and B. Cadre, *On the Asymptotic Properties of a Simple Estimate of the Mode*, ESAIM: Probability and Statistics, 8 (2004), pp. 1–11.

[2] F. Almgren, J. E. Taylor, and L. Wang, *Curvature-Driven Flows: a Variational Approach*, SIAM Journal on Control and Optimization, 31 (1993), pp. 387–438.

[3] L. Ambrosio, *Minimizing Movements*, Rend. Accad. Naz. Sci. XL Mem. Mat. Appl.(5), 19 (1995), pp. 191–246.

[4] L. Ambrosio, N. Gigli, and G. Savaré, *Gradient Flows in Metric Spaces and in the Space of Probability Measures*, Lectures in Mathematics ETH Zürich, Birkhäuser Verlag, Basel, second ed., 2008.

[5] L. Ambrosio, S. Lisini, and G. Savaré, *Stability of Flows Associated to Gradient Vector Fields and Convergence of Iterated Transport Maps*, Manuscripta Mathematica, 121 (2006), pp. 1–50.

[6] E. Arias-Castro, D. Mason, and B. Pelletier, *On the Estimation of the Gradient Lines of a Density and the Consistency of the*

Mean-Shift Algorithm, The Journal of Machine Learning Research, 17 (2016), pp. 1487–1514.

[7] D. Benedetto, E. Caglioti, J. A. Carrillo, and M. Pulvirenti, *A Non-Maxwellian Steady Distribution for One-Dimensional Granular Media*, Journal of Statistical Physics, 91 (1998), pp. 979–990.

[8] D. Benedetto, E. Caglioti, and M. Pulvirenti, *A Kinetic Equation for Granular Media*, ESAIM: Mathematical Modelling and Numerical Analysis, 31 (1997), pp. 615–641.

[9] J. G. Berryman and C. J. Holland, *Nonlinear Diffusion Problem Arising in Plasma Physics*, Physical Review Letters, 40 (1978), pp. 1720–1722.

[10] A. L. Bertozzi, *The Mathematics of Moving Contact Lines in Thin Liquid Films*, Notices of the AMS, 45 (1998), pp. 689–697.

[11] J. Boussinesq, *Recherches Théoriques sur l'Écoulement des Nappes d'Eau Infiltrés dans le Sol et sur le dÉbit de Sources*, Comptes Rendus Acad. Sci. / J. Math. Pures Appl., 10 (1903/04), pp. 5–78.

[12] A. Braides, *Local Minimization, Variational Evolution and Γ-Convergence*, vol. 2094 of Lecture Notes in Mathematics, Springer, 2014.

[13] J. A. Carrillo, R. J. McCann, and C. Villani, *Kinetic Equilibration Rates for Granular Media and Related Equations: Entropy Dissipation and Mass Transportation Estimates*, Revista Matematica Iberoamericana, 19 (2003), pp. 971–1018.

[14] ——, *Contractions in the 2-Wasserstein Length Space and Thermalization of Granular Media*, Archive for Rational Mechanics and Analysis, 179 (2006), pp. 217–263.

[15] S. Chandrasekhar, *Stochastic Problems in Physics and Astronomy*, Reviews of Modern Physics, 15 (1943), pp. 1–89.

[16] V. T. Chow, *Open-Channel Hydraulics*, vol. 131, McGraw-Hill Book Company, 1960.

[17] K. Craig, *Nonconvex Gradient Flow in the Wasserstein Metric and Applications to Constrained Nonlocal Interactions*, Proceedings of the London Mathematical Society, 114 (2017), pp. 60–102.

[18] H. Darcy, *Les Fontaines Publiques de la Ville de Dijon*, V. Dalmont, Paris, (1856), pp. 305–401.

[19] E. De Giorgi, *New Problems on Minimizing Movements*, in Boundary Value Problems for PDE and Applications, C. Baiocchi and J. L. Lions, eds., Masson, 1993, pp. 81–98.

[20] E. De Giorgi, A. Marino, and M. Tosques, *Problems of Evolution in Metric Spaces and Maximal Decreasing Curve*, Atti Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Natur. (8), 68 (1980), pp. 180–187.

[21] M. Degiovanni, A. Marino, and M. Tosques, *Evolution Equations with Lack of Convexity*, Nonlinear Anal., 9 (1985), pp. 1401–1443.

[22] B. T. Fasy, F. Lecci, A. Rinaldo, L. Wasserman, S. Balakrishnan, and A. Singh, *Confidence Sets for Persistence Diagrams*, The Annals of Statistics, (2014), pp. 2301–2339.

[23] F. C. Fleissner, *The Application of the KDE-MM-Scheme to General Second Order Diffusion Equations*, in preparation.

[24] ——, *The JKO-Scheme: from De Giorgi to AI*, in preparation.

[25] F. C. Fleissner, *Gamma-Convergence and Relaxations for Gradient Flows in Metric Spaces: a Minimizing Movement Approach*, ESAIM: Control, Optimisation and Calculus of Variations, 25 (2019).

[26] F. C. Fleissner, *A Minimizing Movement Approach to a Class of Scalar Reaction–Diffusion Equations*, ESAIM: Control, Optimisation and Calculus of Variations, 27 (2021).

[27] F. C. Fleissner, *A Kernel-Density-Estimator Minimizing Movement Scheme for Diffusion Equations*, arXiv preprint arXiv:2310.11961, (2023).

[28] F. C. Fleissner and G. Savaré, *Reverse Approximation of Gradient Flows As Minimizing Movements: a Conjecture by De Giorgi*, Annali

della Scuola Normale Superiore di Pisa. Classe di Scienze, 20 (2020), pp. 677–720.

[29] N. Fournier and A. Guillin, *On the Rate of Convergence in Wasserstein Distance of the Empirical Measure*, Probability Theory and Related Fields, 162 (2015), pp. 707–738.

[30] C. W. Gardiner et al., *Handbook of Stochastic Methods*, vol. 3, Springer Berlin, 1985.

[31] U. Gianazza, G. Savaré, and G. Toscani, *The Wasserstein Gradient Flow of the Fisher Information and the Quantum Drift-Diffusion Equation*, Archive for Rational Mechanics and Analysis, 194 (2009), pp. 133–220.

[32] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, MIT Press, 2016.

[33] M. E. Gurtin and R. C. MacCamy, *On the Diffusion of Biological Populations*, Mathematical Biosciences, 33 (1977), pp. 35–49.

[34] H. Jiang, *Uniform Convergence Rates for Kernel Density Estimation*, in International Conference on Machine Learning, PMLR, 2017, pp. 1694–1703.

[35] R. Jordan, D. Kinderlehrer, and F. Otto, *Free Energy and the Fokker-Planck Equation*, Physica D: Nonlinear Phenomena, 107 (1997), pp. 265–271.

[36] ——, *The Variational Formulation of the Fokker–Planck Equation*, SIAM Journal on Mathematical Analysis, 29 (1998), pp. 1–17.

[37] A. Jüngel, *A Positivity-Preserving Numerical Scheme for a Nonlinear Fourth Order Parabolic System*, SIAM Journal on Numerical Analysis, 39 (2001), pp. 385–406.

[38] ——, *Transport Equations for Semiconductors*, vol. 773 of Lecture Notes in Physics, Springer, 2009.

[39] J. Kim, J. Shin, A. Rinaldo, and L. Wasserman, *Uniform Convergence Rate of the Kernel Density Estimator Adaptive to Intrinsic*

Volume Dimension, in International Conference on Machine Learning, PMLR, 2019, pp. 3398–3407.

[40] H. A. Kramers, *Brownian Motion in a Field of Force and the Diffusion Model of Chemical Reactions*, Physica, 7 (1940), pp. 284–304.

[41] L. S. Leibenzon, *The Motion of a Gas in a Porous Medium. Complete Works*, 2 (Acad. Sciences URSS, Moscow, 1953, First published in Neftanoe i slantsevoe khozyastvo, 10, 1929, and Neftanoe khozyastvo, 8-9, 1930).

[42] J. Li, S. Ray, and B. G. Lindsay, *A Nonparametric Statistical Approach to Clustering via Mode Identification.*, Journal of Machine Learning Research, 8 (2007).

[43] A. Marino, C. Saccon, and M. Tosques, *Curves of Maximal Slope and Parabolic Variational Inequalities on Nonconvex Constraints*, Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4), 16 (1989), pp. 281–330.

[44] D. Matthes, R. J. McCann, and G. Savaré, *A Family of Nonlinear Fourth Order Equations of Gradient Flow Type*, Communications in Partial Differential Equations, 34 (2009), pp. 1352–1397.

[45] U. F. Mayer, *Gradient Flows on Nonpositively Curved Metric Spaces and Harmonic Maps*, Communications in Analysis and Geometry, 6 (1998), pp. 199–253.

[46] R. J. McCann, *A Convexity Principle for Interacting Gases*, Advances in Mathematics, 128 (1997), pp. 153–179.

[47] A. Mielke, *Differential, Energetic, and Metric Formulations for Rate-Independent Processes*, in Nonlinear PDE's and Applications, Springer, 2011, pp. 87–170.

[48] A. Mielke, R. Rossi, and G. Savaré, *Balanced Viscosity Solutions to Infinite-Dimensional Rate-Independent Systems*, Journal of the European Mathematical Society, 18 (2016), pp. 2107–2165.

[49] M. Muskat, *The Flow of Homogeneous Fluids Through Porous Media*, McGraw-Hill, New York, 1937.

[50] C. Ortner, *Two Variational Techniques for the Approximation of Curves of Maximal Slope*, Tech. Rep. NA-05/10, Oxford Comp. Lab. Report, 2005. `http://web2.comlab.ox.ac.uk/oucl/publications/natr/na-05-10.html`.

[51] F. Otto, *Double Degenerate Diffusion Equations As Steepest Descent*, 1996.

[52] ——, *Lubrication Approximation with Prescribed Nonzero Contact Angle*, Communications in Partial Differential Equations, 23 (1998), pp. 2077–2164.

[53] ——, *Dynamics of Labyrinthine Pattern Formation in Magnetic Fluids: A Mean-Field Theory*, Archive for Rational Mechanics and Analysis, 141 (1998, submitted in 1996), pp. 63–103.

[54] ——, *The Geometry of Dissipative Evolution Equations: the Porous Medium Equation*, Communications in Partial Differential Equations, 26 (2001), pp. 101–174.

[55] E. Parzen, *On Estimation of a Probability Density Function and Mode*, The Annals of Mathematical Statistics, 33 (1962), pp. 1065–1076.

[56] R. Pinnau, *The Linearized Transient Quantum Drift Diffusion Model—Stability of Stationary States*, Zeitschrift für Angewandte Mathematik und Mechanik: Applied Mathematics and Mechanics, 80 (2000), pp. 327–344.

[57] A. Rinaldo and L. Wasserman, *Generalized Density Clustering*, The Annals of Statistics, (2010).

[58] M. Rosenblatt, *Remarks on Some Nonparametric Estimates of a Density Function*, The Annals of Mathematical Statistics, (1956), pp. 832–837.

[59] E. Sandier and S. Serfaty, *Gamma-Convergence of Gradient Flows with Applications to Ginzburg-Landau*, Communications on Pure and Applied Mathematics, 57 (2004), pp. 1627–1672.

[60] F. Santambrogio, *Optimal Transport for Applied Mathematicians*, vol. 87 of Progress in Nonlinear Differential Equations and Their Applications, Springer, 2015.

[61] ——, *Euclidean, Metric, and Wasserstein Gradient Flows: an Overview*, Bulletin of Mathematical Sciences, 7 (2017), pp. 87–154.

[62] S. Serfaty, *Gamma-Convergence of Gradient Flows on Hilbert and Metric Spaces and Applications*, Discrete & Continuous Dynamical Systems, 31 (2011), p. 1427.

[63] G. Toscani, *One-Dimensional Kinetic Models of Granular Flows*, ESAIM: Mathematical Modelling and Numerical Analysis, 34 (2000), pp. 1277–1291.

[64] V. S. Varadarajan, *On the Convergence of Sample Probability Distributions*, Sankhyā: The Indian Journal of Statistics (1933-1960), 19 (1958), pp. 23–26.

[65] J. L. Vázquez, *The Porous Medium Equation: Mathematical Theory*, Oxford University Press, 2007.

[66] C. Villani, *Topics in Optimal Transportation*, vol. 58 of Graduate Studies in Mathematics, American Mathematical Society, Providence, RI, 2003.

[67] ——, *Optimal Transport. Old and New*, vol. 338 of Grundlehren der Mathematischen Wissenschaften, Springer-Verlag, Berlin, 2009.

[68] C. B. Vreugdenhil, *Numerical Methods for Shallow-Water Flow*, vol. 13, Springer Science & Business Media, 2013.

[69] D. Wang, X. Lu, and A. Rinaldo, *Optimal Rates for Cluster Tree Estimation Using Kernel Density Estimators*, arXiv preprint arXiv:1706.03113, (2017).

[70] P. Whittle, *On the Smoothing of Probability Density Functions*, Journal of the Royal Statistical Society: Series B (Methodological), 20 (1958), pp. 334–343.

[71] D. V. Widder, *The Heat Equation*, Academic Press New York – London, 1975.

[72] Y. Zel'dovich and Y. Raizer, *Physics of Shock Waves and High-Temperature Hydrodynamic Phenomena II*, Academic Press, New York, 1966.