Technische Universität München
Fakultät für Mathematik
Lehrstuhl für Dynamics

Structure preserving discretization and approximation of gradient flows in Wasserstein-like spaces

Simon Josef Plazotta

Vollständiger Abdruck der von der Fakultät für Mathematik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzender:		Prof. Dr. Folkmar Bornemann
Prüfer der Dissertation:	1.	Prof. Dr. Daniel Matthes
	2.	Prof. Dr. Giuseppe Savaré

Die Dissertation wurde am 14. März 2019 bei der Technischen Universität München eingereicht und durch die Fakultät für Mathematik am 6. Juni 2019 angenommen.

Abstract

The main objective of this Ph.D. thesis is to investigate structure-preserving, temporal semi-discretizations and approximations for PDEs with gradient flow structure with the particular application to evolution problems in the space of probability measures equipped with the L^2-Wasserstein distance. In the spirit of De Giorgi's work on Minimizing Movements [29], we investigate the variational formulation of two particular temporal semi-discretizations and one temporal approximation, namely: time-dependent Minimizing Movement scheme (discretization), second order Backward Differentiation Formula (discretization), and Weighted Energy-Dissipation principle (approximation). The two canonical examples of L^2-Wasserstein gradient flows where we apply these methods are the second-order family of diffusion-aggregation equations given by the non-linear Fokker-Planck equation and the fourth-order Derrida-Lebowitz-Speer-Spohn equation.

Zusammenfassung

Diese Doktorarbeit behandelt verschiedene strukturerhaltende und zeitliche Diskretisierungen und Approximationen von Partiellen Differential Gleichungen mit Gradienten Fluß Struktur und deren Anwendung auf Evolutions Probleme im Raum der Wahrscheinlichkeitsmaße versehen mit der L^2-Wasserstein Distanz. Im Sinne von De Giorgis Arbeit über Minimizing Movements [29] untersuchen wir variationelle Formulierungen von zwei zeitlichen Diskretisierungen und von einer zeitlichen Approximation, namentlich: das zeit-abhängigen impliziten Euler Verfahrens (Diskretisierung), die Rückwärts Differenzen Formel zweiter Ordnung (Diskretisierung), das Prinzip der gewichteten Energie Dissipation (Approximation). Die beiden Hauptbeispiele für L^2-Wasserstein Gradienten Flüße, an welchen wir die oben genannten Methoden anwenden, sind die Familie von Diffusions-Aggregations Gleichungen gegeben durch die nichtlineare Fokker-Planck Gleichung und die Derrida-Lebowitz-Speer-Spohn Gleichung von vierter Ordnung.

Acknowledgements

First of all, I would like to thank my advisor Prof. Dr. Daniel Matthes for his guidance and his constant support. I am deeply thankful for his valuable feedback and for sharing his ingenious intuition throughout the last four years. Thank you very much for finding always – especially during your parental leave – some time to discuss the emerging mathematical and non-mathematical problems.

I want to express my gratitude to Prof. Dr. Giuseppe Savaré for both reviewing this thesis and the kind hospitality at the Università di Pavia during my research stays. Thank you for sharing your knowledge and for your valuable feedback during these times. Furthermore, I would like to thank Prof. Dr. Folkmar Bornemann for taking the chair of the examination committee.

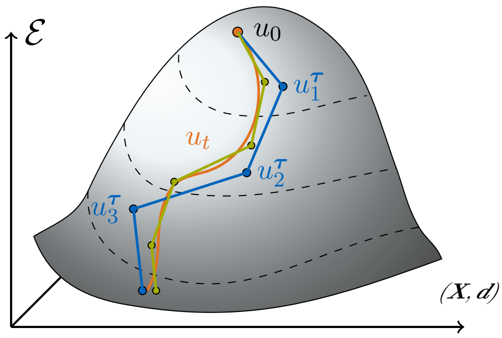
I gratefully acknowledge the DFG Collaboration Research Center TRR 109, "Discretization in Geometry and Dynamics" for funding my position and for its major support during my Ph.D. thesis. In this context, I also want to thank the German Academic Exchange Service (DAAD) and the TUM Graduate School for their scholarships.

My sincere thanks also go to all my friends and colleagues at the mathematical faculty for the friendly and supporting atmosphere. Thank you for all the fertile discussions, the joyful chats, the countless coffee breaks, and all the amusing lunches. Within all my colleagues, a special thanks go to my current and former office mates Benjamin Söllner and Jonathan Kirchner who were always patient with me and all my habits and who were always ready to help whenever I had a stupid or obvious question or other problems.

I also want to thank all my friends outside of university for their patience while listening to my math problems. Thank you for distracting me when my mind was stuck too deep in mathematics or for cheering me up when the math was not working as I wanted.

Especially, I want to thank my family for their patience, their constant encouragement, and their never-ending support no matter what. Thank you so much!

Contents



1 Introduction and Main Results

The main objective of this Ph.D. thesis is to investigate structure-preserving, temporal semi-discretizations and approximations for PDEs with gradient flow structure like

$$\dot{u}_t = -\nabla \mathcal{E}(u_t) \tag{1.0.1}$$

with particular application to evolution problems in the space of probability measures equipped with the L^2-Wasserstein distance. The overall goal was to define novel numerical schemes for approximating solutions of these evolution equations focusing on the preservation of various aspects of the variational structure and the associated qualitative properties of solutions on the approximation level, like: energy monotonicity, auxiliary Lyapunov functionals, positivity, mass preservation, etc.

Gradient Flow Structure of PDEs. Besides the formulation of the gradient flow equation as an ordinary differential equation in the Euclidean space $\mathbb{R}^d$ with the metric induced by the usual inner product of vectors, it is known that also various partial differential equations possess a gradient flow structure with respect to a corresponding infinite-dimensional space. In this setting we mention the well-known heat equation particularly, given by

$$\partial_t \rho_t = \Delta \rho_t \qquad \text{on } \Omega \subseteq \mathbb{R}^d$$

with suitable initial and boundary conditions. With respect to the Hilbert space $L^2(\Omega)$ the driving free energy functional $\mathcal{E}$ of the heat equation is given by the Dirichlet-energy. Other prominent examples in Hilbert spaces are the Allen-Cahn and the Cahn-Hillard equation both driven by the Ginzburg-Landau functional where the first is a gradient flow in $L^2(\Omega)$ and whereas the latter is posed in $H^{-1}(\Omega)$, see [3, 15, 33]. More general systems of reaction-diffusion equations in arbitrary Hilbert spaces have been also investigated, see [49, 74].

L^2-Wasserstein Gradient Flows. In the seminal paper by Jordan, Kinderlehrer and Otto [54] the authors exploited the fact, that the heat equation posses an additional gradient flow structure with respect to the manifold of probability measures $\mathcal{P}_2(\Omega)$ equipped with the so-called L^2-Wasserstein distance. The differential geometry of this dissipative evolution equation and of the extension to porous medium type equations was physically justified by Otto in [78, 79] and rigorously analyzed by McCann in [73], which paved the way for the by now popular and ubiquitous formulation of diffusion equations in metrics related to Optimal Transport. Lastly, this theory was extended by Carrillo, McCann, Villani [21], Carrillo, Gualdani, Jüngel [20], Ambrosio, Gigili, Savaré [4], and Villani [92, 93] to the general non-linear Fokker-Planck equation with confinement and aggregation effects:

$$\partial_t \rho_t = \Delta(\rho_t^m) + \operatorname{div}(\rho_t \nabla V) + \operatorname{div}(\rho_t \nabla (W * \rho_t)).$$

This equation is our main example for this thesis and has a wide variety of applications in statistical mechanics [40, 83], Markov diffusion processes [42, 53], and semiconductor theory [56, 67]. In the past decades, the theory of L^2-Wasserstein gradient flows gained a lot of popularity since it turns out that numerous PDEs posses this differential structure in L^2-Wasserstein like spaces. To name here only few examples: non-local Fokker-Planck equations [19, 31, 91]; Fokker-Planck equations on manifolds [38, 89]; fourth order fluid and quantum models [45, 47, 68]; chemotaxis systems [9, 10, 96]; Poisson-Nernst-Planck equations [58, 95]; multi-component fluid systems [60, 61, 63]; Cahn-Hilliard equations [34, 64]; degenerate cross-diffusion systems [72, 97].

Gradient Flow Theory in Abstract Metric Spaces. In his seminal paper [29], De Giorgi transferred the Euclidean theory of gradient flows to the purely metric theory of curve of steepest descent. The key idea here was to pass from the ordinary gradient flow equation (1.0.1) to a scalar equation whilst preserving the information on the direction. One considers the time derivative of $t \mapsto \mathcal{E}(u_t)$,

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{E}(u_t) = \langle \nabla \mathcal{E}(u_t), \dot{u}_t \rangle \geq -\|\nabla \mathcal{E}(u_t)\| \, \|\dot{u}_t\| \geq -\frac{1}{2}\|\nabla \mathcal{E}(u_t)\|^2 - \frac{1}{2}\|\dot{u}_t\|^2.$$

The first inequality is an equality if and only if $\nabla \mathcal{E}(u_t)$ and $\dot{u}_t$ are anti-parallel and the third inequality is an equality if and only if $\|\nabla \mathcal{E}(u_t)\|$ and $\|\dot{u}_t\|$ are equal. Hence, one can equivalently rewrite the original gradient flow equation (1.0.1) as

$$\frac{1}{2}\|\dot{u}_t\|^2 + \frac{1}{2}\|\nabla \mathcal{E}(u_t)\|^2 = -\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{E}(u_t). \tag{1.0.2}$$

For a more robust formulation, one integrates (1.0.2) in time and replaces the scalar quantities $\|\nabla \mathcal{E}(u_t)\|$ and $\|\dot{u}_t\|$ with the metric surrogates $|\partial \mathcal{E}|(u_t)$ and $|u_t'|$ which are defined in the chapter below, to end up with

$$\frac{1}{2}\int_s^t |u_r'|^2 \, \mathrm{d}r + \frac{1}{2}\int_s^t |\partial \mathcal{E}|^2(u_r) \, \mathrm{d}r = \mathcal{E}(u_s) - \mathcal{E}(u_t) \qquad \text{for a.e. } 0 \leq s \leq t. \tag{1.0.3}$$

Then, one says a curve u_t is a curve of steepest descent with respect to the free energy functional $\mathcal{E}$ if and only if u_t satisfies the Energy Dissipation Equality (EDE) (1.0.3).

Alternatively, a purely metric concept of gradient flows is the Evolution Variational Inequality (EVI). It is only sensible for almost Hilbertian metrics and requires the convexity of the free energy functional $\mathcal{E}$. It is particularly well adapted to deal with gradient flows in the L^2-Wasserstein space $(\mathcal{P}_2(\Omega), \mathbf{W}_2)$. The cornerstone of this ansatz is to consider the derivative of the squared distance of u_t to a fixed reference point v,

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{1}{2}\|u_t - v\|^2 = \langle \dot{u}_t, u_t - v\rangle = -\langle \nabla\mathcal{E}(u_t), u_t - v\rangle \leq \mathcal{E}(v) - \mathcal{E}(u_t).$$

The second equality holds true due to the fact that u_t is a solution to the gradient flow equation (1.0.1) and the inequality is due to the convexity of the free energy functional $\mathcal{E}$. Actually, it is possible to prove under the assumption of smoothness of the free energy function $\mathcal{E}$, that if the following differential comparison principle

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{1}{2}\|u_t - v\|^2 \leq \mathcal{E}(v) - \mathcal{E}(u_t) \tag{1.0.4}$$

is satisfied for each v, then (1.0.4) is actually equivalent to the original formulation of the gradient flow equation (1.0.1). As with the EDE, there is an integrated version of (1.0.4) and that is sufficiently robust to be used for abstract metric gradient flows. For a detailed discussion of the role of the EVI in the gradient flow theory, we refer to [4, Chapter 4.0].

Analysis of Gradient Flows in Abstract Metric Spaces. The cornerstone of the theoretical analysis of these gradient flows in general abstract metric spaces $(\mathbf{X}, \boldsymbol{d})$ is the variational formulation of the implicit Euler scheme introduced by De Giorgi in his seminal papers [29, 30] on Minimizing Movements. Later, in [54] Jordan, Kinderlehrer and Otto used this scheme – which is nowadays called JKO-scheme in this context – to construct solutions to the heat equation. Schematically, the main idea is to apply the following implicit induction formula:

$$u_k^\tau \in \operatorname*{argmin}_{u\in\mathbf{X}} \frac{1}{2\tau}\boldsymbol{d}^2(u_k^\tau, u) + \mathcal{E}(u) \tag{1.0.5}$$

such that – at least in the Euclidean setting – the minimizer u_k^τ satisfy the usual implicit Euler method. There are various "soft" conditions that guarantee well-definedness of this scheme, i.e., the inductive solvability of the minimization problems. This technique was transferred by Ambrosio, Gigli, Savaré in the monograph [4] to mainly investigate gradient flows in abstract metric spaces $(\mathbf{X}, \boldsymbol{d})$ with respect to λ-convex free energy functionals $\mathcal{E}$, see the famous hypothesis [4, Assumption 4.0.4] for more details. Note that λ-convexity can be understood here as a sort of convexity along geodesics which takes the curvature of the underlying space into account, see [73] for the proper notion of convexity in the L^2-Wasserstein case. One of the remarkable strengths of this method in the abstract metric setting is its intrinsic stability properties: the unconditional energy dissipation of the approximated solution in every time step; and the step size independent bounds on the integrated kinetic energy. These bounds are usually sufficient to conclude the convergence of the discrete solution to a continuous curve u_t^* in the limit $\tau \to 0$. Some additional work is needed to prove that u_t^* is indeed a curve of steepest descent in the sense of the energy dissipation equation or the evolution variational inequality.

Numerical Aspects of $\mathbf{L^2}$-Wasserstein Gradient Flow Theory. Besides the theoretical use of the variational formulation of the implicit Euler scheme (1.0.5) to construct a solution for the gradient flow equation, this particular discretization provides also a structure-preserving numerical scheme. Different approaches to actually compute the minimizers of the Minimizing Movement Scheme by means of certain spatial-temporal full-discretization have been investigated: particle schemes [16, 23, 17, 94]; evolving diffeomorphisms [23, 22]; Lagrangian schemes [8, 32, 41, 55, 69, 71]; entropic regularization [80]; primal-dual methods [18].

Aim of the Thesis. In the spirit of De Giorgi's work on Minimizing Movements [29], we seek to exploit the variational structure of the gradient flow equation to construct a sequence of approximate solutions by means of temporal semi-discretizations or approximations. Following this, the key tool to design novel schemes was to cast existing numerical schemes for the gradient flow equation seen as ODE in an equivalent variational formulation which are robust enough to deal with the possibly rough structure of the underlying metric space and of the free energy functional. We emphasize that throughout all of our results, a decisive role is played by functional inequalities for the approximated solutions obtained by exactly this variational formulation. In the end, these functional inequalities turn into the relevant a priori estimates and into the essential information about the structural properties of the approximated solution to derive the limit behavior as the approximation parameter tends to zero.

In the following, we give a brief overview of the different discretization and approximation schemes and on the related results, we developed and analyzed in this thesis.

A Time-Dependent Minimizing Movement Scheme. Chapter 3 is designated to study the particular temporal discretization for non-autonomous evolution problems where the free energy functional $\mathcal{E}_t$ depends additional on time t by means of the variational formulation of the time-dependent implicit Euler method. In the Euclidean setting the main idea to approximate solutions u_t^* to gradient flow problems is to use the time-dependent implicit Euler scheme, given by the induction formula

$$\frac{u_k^{\boldsymbol{\tau}} - u_{k\text{-}1}^{\boldsymbol{\tau}}}{\tau_k} = -\nabla\mathcal{E}_{t_k^{\boldsymbol{\tau}}}(u_k^{\boldsymbol{\tau}}),$$

where $t_k^{\boldsymbol{\tau}} = \sum_{l=1}^k \tau_l$ for $k \geq 1$ for a given partition $\boldsymbol{\tau} := (\tau_1, \tau_2, \ldots)$ of step sizes $\tau_k \in (0, \tau_*)$ and for a given an initial condition $u_0^{\boldsymbol{\tau}}$ that approximates u_0.

In the abstract metric case when $(\mathbf{X}, \boldsymbol{d})$ is just a complete, separable metric space we propose the time-dependent version of the Minimizing Movement scheme. The variational formulation of the time-dependent implicit Euler method reads then

$$u_k^{\boldsymbol{\tau}} \in \underset{w\in\mathbf{X}}{\operatorname{argmin}} \frac{1}{2\tau_k}\boldsymbol{d}^2(u_{k\text{-}1}^{\boldsymbol{\tau}}, w) + \mathcal{E}_{t_k^{\boldsymbol{\tau}}}(w).$$

In the Euclidean setting the minimizer $u_k^{\boldsymbol{\tau}}$ satisfies the implicit Euler formula. We derive existence and convergence results for the approximate solutions $(u_k^{\boldsymbol{\tau}})_{k\in\mathbb{N}}$ as the discretization parameter $\boldsymbol{\tau} = \sup_k \tau_k$ tends to zero.

- **B Time-Homogenization.** In chapter 4 we investigate the high-frequency limit of the family $(u_t^\omega)_\omega$ of solutions to the non-autonomous evolution problems
$$\dot{u}_t^\omega = -\nabla\mathcal{E}_{\omega t}(u_t^\omega), \qquad u_0^\omega = u_0$$
with respect to the convex free energy functionals $\mathcal{E}_{t\omega} = \overline{\mathcal{E}} + \mathcal{P}_{t\omega}$, where $\overline{\mathcal{E}}(u) := \fint \mathcal{E}_t(u)\,\mathrm{d}t$ is the time-average and $\mathcal{P}_{t\omega} = \mathcal{E}_{t\omega} - \overline{\mathcal{E}}$ is the periodic forcing.
A comparison principle for u_t^ω and u_t^∞, i.e., the solution of gradient flow equation driven by the time-averaged free energy functional $\overline{\mathcal{E}}$, yields the convergence result.
- **C Variational Second Order Backward Differentiation Formula.** Our intention in chapter 5 is to design a temporal semi-discretization, which converges at least formally to second order, by means of the second order Backward Differentiation Formula (BDF2). In the Euclidean setting, this temporal discretization of the gradient flow equation reads then
$$\frac{3u_k^\tau - 4u_{k\text{-}1}^\tau + u_{k\text{-}2}^\tau}{2\tau} = -\nabla\mathcal{E}(u_k^\tau),$$
for a given time step size $\tau \in (0, \tau_*)$ and well-prepared initial data $(u_0^\tau, u_1^\tau) \approx u_0$.
We proposed a variational formulation of the BDF2 method to construct a discrete approximation $(u_k^\tau)_{k\in\mathbb{N}}$ for metric gradient flows, which reads as follows:
$$u_{k+1}^\tau \in \operatorname*{argmin}_{w\in\mathbf{X}} \frac{1}{\tau}\boldsymbol{d}^2(u_k^\tau, w) - \frac{1}{4\tau}\boldsymbol{d}^2(u_{k\text{-}1}^\tau, w) + \mathcal{E}(w).$$
In this case, we also prove the existence and convergence of the approximated solution $(u_k^\tau)_{k\in\mathbb{N}}$. Notably, in the abstract metric space case, an additional assumption is needed, namely a sort of convexity of the BDF2-penalized free energy functional.
- **D Weighted Energy-Dissipation Principle.** In chapter 6 we follow a different time-continuous approximation approach by means of the Weighted Energy-Dissipation principle (WED). The main idea here is to perturb the gradient flow equation by an elliptic regularization in time
$$-\varepsilon\partial_t^2 u_t^\varepsilon + \partial_t u_t^\varepsilon = -\nabla\mathcal{E}(u_t^\varepsilon).$$
Even though one loses the gradient flow structure at first glance, the solutions u_t^ε satisfy another crucial variational principle. In particular, the solutions u_t^ε are the minimizer of a global-in-time minimization of the WED-functional Φ_ε, given by
$$u_t^\varepsilon \in \operatorname*{argmin}_{u_t} \int_0^\infty \frac{e^{-t/\varepsilon}}{\varepsilon}\left(\frac{\varepsilon}{2}\left|u_t'\right|^2 + \mathcal{E}(u_t)\right)\,\mathrm{d}t.$$
Here, the minimization is performed over a suitable class of curves u_t emanating from u_0. We are able to adapt this approach to gradient flows in the L^2-Wasserstein space and to prove the existence and convergence of the approximated solution u_t^ε as the approximation parameter ε tends to zero.

In advance of the detailed elaboration of our work we begin with a short summary on notation and motivation of the theory of gradient flows and related topics.

2 Notation and Preliminaries

2.1 Function Spaces

In this thesis, we always denote by $\Omega \subset \mathbb{R}^\mathrm{d}$ an open, bounded, and connected domain with Lipschitz-continuous boundary $\partial\Omega$ with normal derivative $\boldsymbol{n}$ or Ω is equal to $\mathbb{R}^\mathrm{d}$.

Space of Continuous Functions. Let us consider scalar functions $\varphi : \Omega \to \mathbb{R}$ and vector fields $\xi : \Omega \to \mathbb{R}^\mathrm{d}$. If these functions depend additionally on time we write φ_t and ξ_t. For $k \in \mathbb{N} \cup \{\infty\}$ we define by $\mathcal{C}^k(\Omega)$ the set of all k-times continuous differentiable functions. The set of all functions which have in addition compact support is denoted by $\mathcal{C}_c^k(\Omega)$. Furthermore, we denote by $\mathcal{C}_b(\Omega)$ all continuous functions which are bounded on Ω. Analogously, we define $\mathcal{C}^k(\Omega, \mathbb{R}^\mathrm{d})$, $\mathcal{C}_c^k(\Omega, \mathbb{R}^\mathrm{d})$, and $\mathcal{C}_b(\Omega, \mathbb{R}^\mathrm{d})$ for vector fields. The spatial derivative of φ is denoted by $\mathrm{D}\varphi$ and we write $\mathrm{D}^2\varphi$ for the second order derivative. Partial derivatives with respect to one component x_i or t are denoted by ∂_{x_i} and ∂_t, respectively. For higher-order partial derivatives we use a multi-index $\alpha \in \mathbb{N}_0^d$ of order $|\alpha| = \sum_{i=1}^d \alpha_i = k$. Then define $\partial^\alpha\varphi$ of a k-times differentiable function φ by $(\partial_{x_1})^{\alpha_1} \cdots (\partial_{x_d})^{\alpha_d}\varphi$. The gradient $\nabla\varphi$ of φ and the divergence $\mathrm{div}(\xi)$ of ξ are given by $\nabla\varphi = (\mathrm{D}\varphi)^\top$ and $\mathrm{div}(\xi) = \mathrm{tr}(\mathrm{D}\xi)$. We write $\Delta\varphi = \mathrm{div}(\nabla\varphi)$ for the Laplacian of φ.

Lebesgue Integrable Functions. Denote by $\mathcal{L}^\mathrm{d}$ the d-dimensional Lebesgue measure on the domain Ω. We write $\mathcal{L}(\Omega)$ for all Lebesgue-measurable sets of Ω and we call a function φ Lebesgue-measurable if and only if the preimage of every Lebesgue-measurable set of $\mathbb{R}$ under φ is a Lebesgue-measurable set of Ω. For $p \in [1, \infty)$ the L^p-norm of a Lebesgue-measurable function φ is defined by

$$\|\varphi\|_{\mathrm{L}^p(\Omega)} := \Big(\int_\Omega |\varphi(x)|^p \,\mathrm{d}\mathcal{L}^\mathrm{d}(x) \Big)^{1/p}.$$

All Lebesgue-measurable functions with finite L^p-norm form the set of p-integrable functions $\mathrm{L}^p(\Omega)$. The set $\mathrm{L}^p_{\mathrm{loc}}(\Omega)$ of locally p-integrable functions is defined as all Lebesgue-measurable functions φ with $\|\varphi\|_{\mathrm{L}^p(K)} < \infty$ for all compact subsets $K \Subset \Omega$. For $p = \infty$ we introduce the set of essentially bounded functions $\mathrm{L}^\infty(\Omega)$ with the norm

$$\|\varphi\|_{\mathrm{L}^\infty(\Omega)} := \operatorname*{ess\,sup}_{x \in \Omega} |\varphi(x)| \,.$$

For any $p \in [1, \infty)$ the space $\mathrm{L}^p(\Omega)$ is a Banach space with dual $(\mathrm{L}^p(\Omega))^* \cong \mathrm{L}^q(\Omega)$ for the Hölder-conjugated exponent q which is given through the formula $\frac{1}{p} + \frac{1}{q} = 1$. Analogously, define the corresponding spaces of p-integrable vector fields $\mathrm{L}^p(\Omega, \mathbb{R}^\mathrm{d})$.

Functions of Bounded Variation. We recall the basic definitions and properties of functions of bounded variation, following [48]. For a given open domain Ω a function $\varphi \in \mathrm{L}^1(\Omega)$ is called a *function of bounded variation* if and only if

$$V(\varphi,\Omega) := \sup\left\{ \int_\Omega \varphi(x)\ \operatorname{div}\xi(x)\,\mathrm{d}\mathcal{L}^{\mathrm{d}}(x) \mid \xi \in C_c^\infty(\Omega,\mathbb{R}^{\mathrm{d}}),\ \|\xi\|_\infty \le 1 \right\} < \infty.$$

The set of all functions of bounded variation is denoted by $\mathrm{BV}(\Omega)$ with the norm:

$$\|\varphi\|_{\mathrm{BV}(\Omega)} = \|\varphi\|_{\mathrm{L}^1(\Omega)} + V(\varphi,\Omega).$$

For open sets $\Omega \subset \mathbb{R}^{\mathrm{d}}$ the set $\mathrm{BV}(\Omega)$ is a Banach space and the norm is lower semi-continuous with respect to the weak convergence in $\mathrm{L}^1(\Omega)$.

Sobolev Spaces. We say a Lebesgue-measurable scalar function φ is k-times weakly differentiable if for each multi-index α of order k there exists a function $v \in \mathrm{L}^1_{\mathrm{loc}}(\Omega)$ with

$$\int_\Omega \varphi\partial^\alpha\psi(x)\,\mathrm{d}\mathcal{L}^{\mathrm{d}}(x) = (-1)^{|\alpha|}\int_\Omega v\psi(x)\,\mathrm{d}\mathcal{L}^{\mathrm{d}}(x) \quad \forall\,\psi \in \mathcal{C}_c^\infty(\Omega).$$

In this case v is denoted by $\partial^\alpha\varphi$. Then for $k \in \mathbb{N}$ and $p \in [1,\infty)$ the Sobolev space $\mathrm{W}^{k,p}(\Omega)$ is defined as the set of all Lebesgue-measurable functions φ such that the $\mathrm{W}^{k,p}(\Omega)$-norm is finite, i.e.,

$$\|\varphi\|_{\mathrm{W}^{k,p}(\Omega)} := \Big(\|\varphi\|^p_{\mathrm{L}^p(\Omega)} + \sum_{|\alpha|\le k} \|\partial^\alpha\varphi\|^p_{\mathrm{L}^p(\Omega)} \Big)^{1/p} < \infty.$$

For any $k \in \mathbb{N}$ and any $p \in [1,p)$ the Sobolev space $\mathrm{W}^{k,p}(\Omega)$ is a Banach space. Furthermore, for $p = 2$ the spaces $\mathrm{W}^{k,2}(\Omega)$ are Hilbert spaces and are denoted by $\mathrm{H}^k(\Omega)$.

Bochner Spaces. Let us consider time-dependent functions $u_t : [0,\infty) \to \mathbf{X}$ with values in a complete, separable metric space $(\mathbf{X},\boldsymbol{d})$. We define the set $\mathcal{M}(0,T;(\mathbf{X},\boldsymbol{d}))$ as all $\mathbf{X}$-valued functions φ_t which are measurable with respect to $\mathcal{B}(\mathbf{X})$, the Borel-sigma algebra on $\mathbf{X}$ generated by topology with respect to $\boldsymbol{d}$. We say a sequence $(u_t^n)_{n\in\mathbb{N}}$ of measurable $\mathbf{X}$-valued functions converges in $\mathcal{M}(0,T;(\mathbf{X},\boldsymbol{d}))$ – in words, converges in measure – to a limit function u_t^* if and only if

$$\lim_{n\to\infty} \mathcal{L}^1(\{t \in (0,T) \mid \boldsymbol{d}(u_t^n,u_t^*) \ge \varepsilon\}) = 0 \qquad \text{for all } \varepsilon > 0.$$

If $(\mathbf{X},\boldsymbol{d})$ is a vector space V, we define the Bochner spaces $\mathrm{L}^p(0,T;\mathrm{V})$ as the set of all $\mathcal{B}(\mathrm{V})$-measurable functions u_t such that the $\mathrm{L}^p(0,T;\mathrm{V})$-norm is finite, i.e,

$$\|u_t\|_{\mathrm{L}^p(0,T;\mathrm{V})} := \Big(\int_0^T \|u_t\|^p_{\mathrm{V}}\ \mathrm{d}t \Big)^{1/p} < \infty$$

Remark 2.1.1. If $(\mathbf{X},\boldsymbol{d})$ is a Banach space V with $\boldsymbol{d}$ induced by the intrinsic norm, we write only $\mathcal{M}(0,T;\mathrm{V})$ instead of $\mathcal{M}(0,T;(\mathbf{X},\boldsymbol{d}))$. In this case the topology of $\mathrm{L}^p(0,T;\mathrm{V})$ and of $\mathcal{M}(0,T;\mathrm{V})$ coincide on p-uniformly integrable sets, see [85, Proposition 1.10].

2.2 L^2-Wasserstein Spaces $(\mathcal{P}_2(\Omega), \mathbf{W}_2)$

In this section, we briefly review the basic definitions and facts concerning the analysis in the metric space of probability measures. For more details on optimal transport and the connection to the gradient flow theory, we refer to the monographs by Ambrosio et al. [4], Santambrogio [87], and Villani [92, 93].

Space of Probability Measures $\mathcal{P}_2(\Omega)$. Given a domain $\Omega \subseteq \mathbb{R}^{\mathrm{d}}$ as before. By $\mathcal{P}_2(\Omega)$ we denote the set of probability measures on Ω with finite second moment $\boldsymbol{M}_2$, i.e. $\boldsymbol{M}_2(\mu) := \int_\Omega \|x\|^2 \,\mathrm{d}\mu(x) < \infty$. The subset of probability measures μ which are absolutely continuous with respect to the Lebesgue measure $\mathcal{L}^{\mathrm{d}}$, i.e., $\mu = \rho \,\mathrm{d}\mathcal{L}^{\mathrm{d}}$, is denoted by $\mathcal{P}_2^{\mathrm{ac}}(\Omega)$. By abuse of notation, we identify an absolutely continuous measure μ with its Lebesgue density ρ and visa versa. Define $\varphi_\# \mu$ by the pushforward of the measure μ with respect to the measurable function φ, i.e., $\varphi_\#\mu(B) := \mu(\varphi^{-1}(B))$ for all measurable sets B. Given the product space $\Omega_1 \times \Omega_2$, we define the projections π_1 and π_2 by $\pi_i(x_1, x_2) = x_i$ for all $(x_1, x_2) \in \Omega_1 \times \Omega_2$. Hence, given a measure $\boldsymbol{p} \in \mathcal{P}_2(\Omega_1 \times \Omega_2)$ the marginal distributions are given by $(\pi_1)_\# \boldsymbol{p} \in \mathcal{P}_2(\Omega_1)$ and $(\pi_2)_\# \boldsymbol{p} \in \mathcal{P}_2(\Omega_2)$.

We equip $\mathcal{P}_2(\Omega)$ with the topology induced by the *narrow-convergence* of measures, denoted by $\mu_n \rightharpoonup^* \mu$, if and only if

$$\lim_{n\to\infty} \int_\Omega \psi \,\mathrm{d}\mu_n(x) = \int_\Omega \psi \,\mathrm{d}\mu(x) \qquad \psi \in \mathcal{C}_b(\Omega).$$

I.e., narrow convergence is equal to the weak*-convergence of measures, which is induced by the pairing of the continuous and bounded functions $\mathcal{C}_b(\Omega)$ with the corresponding dual space of finitely additive signed Borel measures $\mathcal{M}_f(\Omega)$.

L^2-Wasserstein Space $(\mathcal{P}_2(\Omega), \mathbf{W}_2)$. The main object in the theory of Optimal Transport is the space of probability measures $\mathcal{P}_2(\Omega)$ equipped with the L^2-Wasserstein distance $\mathbf{W}_2$, defined by the *Kantorovich problem*:

$$\mathbf{W}_2^2(\mu, \nu) := \inf_{\boldsymbol{p} \in \Gamma(\mu,\nu)} \int_{\Omega^2} \|x - y\|^2 \,\mathrm{d}\boldsymbol{p}(x, y) \tag{2.2.1}$$

where $\Gamma(\mu, \nu) := \{\boldsymbol{p} \in \mathscr{P}(\Omega \times \Omega) : (\pi_1)_\# \boldsymbol{p} = \mu,\ (\pi_2)_\# \boldsymbol{p} = \nu\}$ is the set of all probability measures $\boldsymbol{p}$ whose marginals are μ and ν. If either $\mu \in \mathcal{P}_2^{\mathrm{ac}}(\Omega)$ or $\nu \in \mathcal{P}_2^{\mathrm{ac}}(\Omega)$, then by [92, Theorem 2.12] the minimizer $\boldsymbol{p}_{opt} \in \Gamma(\mu, \nu)$ of the Kantorovich problem (2.2.1) exists and is called the *optimal transport plan*. Furthermore, if $\mu \in \mathcal{P}_2^{\mathrm{ac}}(\Omega)$ the optimal transport plan $\boldsymbol{p}_{opt}$ is uniquely determined by $\boldsymbol{p}_{opt} = (\mathbb{1}_\Omega \times \mathrm{T}_{opt})_\# \mu$, where the *optimal transport map* T_{opt} is the unique solution of the *Monge Problem*, i.e.,

$$\mathbf{W}_2^2(\mu, \nu) = \int_\Omega \|x - \mathrm{T}_{opt}(x)\|^2 \,\mathrm{d}\mu(x) = \inf_{\mathrm{T}_\#\mu = \nu} \int_\Omega \|x - \mathrm{T}(x)\|^2 \,\mathrm{d}\mu(x). \tag{2.2.2}$$

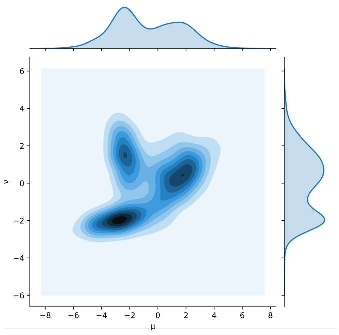


Figure 2.1: Two one-dimensional measures μ and ν plotted on the x and y axes, and one possible joint distribution that defines a transport plan between them [24].

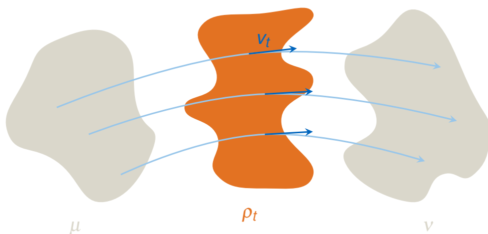


Figure 2.2: Two two-dimensional measures μ and ν are connected by a density-velocity pair $(\rho_t, \mathbf{v}_t)$ which solves the continuity equation in the sense of distributions.

Dynamic Properties of $(\mathcal{P}_2(\Omega), \mathbf{W}_2)$. Another equivalent characterization of the L^2-Wasserstein distance $\mathbf{W}_2$ is given by Benamou and Brenier in [7]: Define the set $\mathfrak{C}(\mu, \nu)$ of density-velocity pairs connecting two measures μ and ν via the continuity equation by

$$\mathfrak{C}(\mu, \nu) := \{(\rho_t, \mathbf{v}_t) \in \mathcal{P}_2(\Omega) \times \mathscr{V}(\Omega) \mid \partial_t \rho_t + \operatorname{div}(\rho_t \mathbf{v}_t) = 0,\ \rho_0 = \mu,\ \rho_1 = \nu\},$$

where $\mathscr{V}(\Omega)$ is the set of all measurable vector fields on Ω. Note, the continuity equation is meant to be satisfied in the weak sense with no-flux boundary condition, i.e., the density-velocity pair $(\rho_t, \mathbf{v}_t)$ with $\mathbf{v}_t \in \mathrm{L}^1(0,1; \mathrm{L}^1(\Omega, \mathrm{d}\rho_t))$ satisfies for each test function $\varphi_t \in \mathcal{C}_c^\infty((0,1) \times \overline{\Omega})$:

$$\int_0^1 \int_\Omega \partial_t \varphi_t(x) + \langle \nabla \varphi_t(x), \mathbf{v}_t(x) \rangle \,\mathrm{d}\rho_t(x) \,\mathrm{d}t = 0.$$

Then, the dynamic formulation of the L^2-Wasserstein distance is given by

$$\mathbf{W}_2^2(\mu, \nu) := \inf_{(\rho_t, \mathbf{v}_t) \in \mathfrak{C}(\mu, \nu)} \int_0^1 \int_\Omega |\mathbf{v}_t(x)|^2 \,\mathrm{d}\rho_t(x) \,\mathrm{d}t. \tag{2.2.3}$$

Note, if $\mu, \nu \in \mathcal{P}_2^{\mathrm{ac}}(\Omega)$ then also the connecting measure ρ_t is absolutely continuous with respect to the Lebesgue measure for each $t \in [0,1]$, see [92, Theorem 8.1].

Equivalent Reformulation of the Dynamic Properties. The for this thesis more convenient version of the Benamou-Brenier Formula (2.2.3) is written in terms of density-flux pairs $(\rho_t, \mathbf{w}_t)$ instead of density-velocity pairs $(\rho_t, \mathbf{v}_t)$. Similarly, define the set

$$\mathcal{C}(\mu, \nu) := \{(\rho_t, \mathbf{w}_t) \in \mathcal{P}_2(\Omega) \times \mathscr{W}(\Omega) \mid \partial_t \rho_t + \operatorname{div}(\mathbf{w}_t) = 0,\ \rho_0 = \mu,\ \rho_1 = \nu\},$$

where $\mathscr{W}(\Omega)$ is the set of $\mathbb{R}^{\mathrm{d}}$-valued signed Borel measures. Here, the continuity equation with no-flux boundary condition is understand to hold as follows: the density-flux pair $(\rho_t, \mathbf{w}_t)$ with $t \mapsto |\mathbf{w}_t|(\Omega) \in \mathrm{L}^1(0,1)$ satisfies for each test function $\varphi_t \in \mathcal{C}_c^\infty((0,1) \times \overline{\Omega})$:

$$\int_0^1 \int_\Omega \partial_t \varphi_t(x) \,\mathrm{d}\rho_t(x) \,\mathrm{d}t + \int_0^1 \int_\Omega \nabla \varphi_t(x) \,\mathrm{d}\mathbf{w}_t(x) \,\mathrm{d}t = 0.$$

Using the representation $\mathbf{v}_t \,\mathrm{d}\rho_t = \mathrm{d}\mathbf{w}_t$, the Benamou-Brenier formula (2.2.3) rewrites to

$$\mathbf{W}_2^2(\mu, \nu) := \inf_{(\rho_t, \mathbf{w}_t) \in \mathcal{C}(\mu,\nu)} \int_0^1 \mathcal{K}(\rho_t, \mathbf{w}_t) \,\mathrm{d}t \tag{2.2.4}$$

where the kinetic energy $\mathcal{K}$ of a density-flux pair $(\rho_t, \mathbf{w}_t)$ is defined by

$$\mathcal{K}(\rho, \mathbf{w}) := \int_\Omega \mathbf{K}(1, \mathbf{v}) \,\mathrm{d}\rho(x) \quad \text{where} \quad \mathbf{K}(t, z) \begin{cases} \|z\|^2 t^{-1} & \text{if } t > 0, \\ 0 & \text{if } (t,z) = (0,0), \\ +\infty & \text{if either } t < 0 \text{ or } t = 0 \wedge z \neq 0, \end{cases}$$

if the vectorial measure $\mathbf{w}$ is absolutely continuous with respect to ρ with density $\mathbf{v}$, i.e., $\mathbf{v}\,\mathrm{d}\rho = \mathrm{d}\mathbf{w}$ otherwise we set $\mathcal{K}(\rho, \mathbf{w}) = \infty$. Note, with this convention the integral-functional $\mathcal{K}$ is lower semi-continuous with respect to narrow convergence, jointly convex and 1-homogenous. Also, if $\rho \in \mathcal{P}_2^{\mathrm{ac}}(\Omega)$, then the feasible class of fluxes $\mathbf{w}$ is given by the absolutely continuous $\mathbb{R}^{\mathrm{d}}$-valued signed Borel measures. In this case, one has $\mathcal{K}(\rho, \mathbf{w}) = \int_\Omega \mathbf{K}(\rho, \mathbf{w}) \,\mathrm{d}x$. As before, if $\mu, \nu \in \mathcal{P}_2^{\mathrm{ac}}(\Omega)$ then also $\rho_t \in \mathcal{P}_2^{\mathrm{ac}}(\Omega)$ for each t.

L^2-Absolutely Continuous Curves and the Continuity Equation. Lastly, we define the set $\mathcal{C}(\mu)$ as all density-flux pairs $(\rho_t, \mathbf{w}_t)$ which solve the free end problem

$$\partial_t \rho_t + \operatorname{div}(\mathbf{w}_t) = 0, \qquad \rho_0 = \mu \tag{2.2.5}$$

in a weak sense with no-flux boundary condition, as before. Then, the link between L^2-absolutely continuous curves $\rho_t \in \mathrm{AC}^2(0, \infty; (\mathcal{P}_2(\Omega), \mathbf{W}_2))$ and the density-flux pairs $(\widehat{\rho}_t, \widehat{\mathbf{w}}_t) \in \mathcal{C}(\rho_0)$ is given by the following characterization, cf. [4, Thm. 8.3.1]:

Theorem 2.2.1 (Absolutely Continuous Curves and the Continuity Equation). *If ρ_t is a L^2-absolutely continuous curve with a locally integrable metric slope $|\rho_t'| \in \mathrm{L}^1_{loc}(0,\infty)$, then there exits a flux $\mathbf{w}_t$ such that $(\rho_t, \mathbf{w}_t) \in \mathcal{C}(\rho_0)$ and*

$$\mathcal{K}(\rho_t, \mathbf{w}_t) \leq |\rho_t'|^2 \qquad \text{for a.e. } t. \tag{2.2.6}$$

Conversely, if a narrowly-continuous curve ρ_t satisfies the continuity equation (2.2.5) *for some flux function $\mathbf{w}_t$ with $t \mapsto \sqrt{\mathcal{K}(\rho_t, \mathbf{w}_t)} \in \mathrm{L}^1_{loc}(0,\infty)$, then ρ_t is L^2-absolutely continuous and*

$$|\rho_t'|^2 \leq \mathcal{K}(\rho_t, \mathbf{w}_t) \qquad \text{for a.e. } t. \tag{2.2.7}$$

2.3 Gradient Flow Theory in Abstract Metric Spaces

In this section, we give a brief overview on the theory of gradient flows in abstract metric spaces $(\mathbf{X}, \boldsymbol{d})$, which is the foundation of the novel discretization and approximation techniques developed in this thesis. For a comprehensive introduction to the theory of gradient flows in metric spaces we refer to the monograph of Ambrosio et al. [4].

Topology. Here and below, $(\mathbf{X}, \boldsymbol{d})$ is separable, complete metric space with a weaker Hausdorff topology $\boldsymbol{\sigma}$ on $\boldsymbol{X}$, that is compatible with $\boldsymbol{d}$, i.e.,

$$u_n \overset{\boldsymbol{d}}{\to} u \implies u_n \overset{\boldsymbol{\sigma}}{\rightharpoonup} u, \qquad (u_n, v_n) \overset{\boldsymbol{\sigma}}{\rightharpoonup} (u, v) \implies \boldsymbol{d}(u, v) \leq \liminf_{n\to\infty} \boldsymbol{d}(u_n, v_n).$$

From now on we use the convention to write

$$u_n \overset{\boldsymbol{d}}{\to} u \quad \text{for the convergence w.r.t. } \boldsymbol{d}, \qquad u_n \overset{\boldsymbol{\sigma}}{\rightharpoonup} u \quad \text{for the convergence w.r.t. } \boldsymbol{\sigma}.$$

Note, this additional weaker topology on $\mathbf{X}$ allows us more flexibility to derive compactness results.

Analysis in Metric Spaces. Given the free energy functional $\mathcal{E} : \mathbf{X} \to \mathbb{R} \cup \{\infty\}$, we say $\mathcal{E}$ is proper if the domain $\mathcal{D}(\mathcal{E}) := \{u \mid \mathcal{E}(u) < \infty\}$ where the free energy functional is finite is not empty. The metric surrogates for the norm of the gradient of the free energy functional $\|\nabla\mathcal{E}\|$ and of the time derivative $\|\dot{u}_t\|$ in (1.0.3) are defined as follows, see [4, Definition 1.1.1&1.2.4] for further details.

Definition 2.3.1 (Local Slope)**.** Given a functional $\mathcal{E} : \mathbf{X} \to \mathbb{R} \cup \{\infty\}$ defined on a metric space $(\mathbf{X}, \boldsymbol{d})$. Then the *local slope* $|\partial\mathcal{E}| : \mathbf{X} \to \mathbb{R} \cup \{\infty\}$ of $\mathcal{E}$ at $u \in \mathbf{X}$ is defined via

$$|\partial\mathcal{E}|(u) := \limsup_{v\to u} \Big(\frac{\mathcal{E}(u) - \mathcal{E}(v)}{\boldsymbol{d}(u, v)}\Big)^+.$$

Definition 2.3.2 (AC Curves)**.** A curve $u_t : [0, \infty) \to \mathbf{X}$ is said to be L^2-*absolutely continuous*, written as $u_t \in \mathrm{AC}^2\left(0, \infty; (\mathbf{X}, \boldsymbol{d})\right)$, if there exists a function $m \in \mathrm{L}^2_{\mathrm{loc}}(0, \infty)$ such that

$$\boldsymbol{d}(u_t, u_s) \leq \int_s^t m(r)\,\mathrm{d}r \qquad \text{for all} \quad 0 \leq s \leq t.$$

It can be shown [4, Theorem 1.1.2] that among all possible choices for m, there is a minimal one, called the *metric derivative* $|u_t'| \in \mathrm{L}^2_{\mathrm{loc}}(0, \infty)$, given by

$$|u_t'| := \lim_{s\to t} \frac{\boldsymbol{d}(u_s, u_t)}{|s - t|} \qquad \text{for a.e. } t.$$

Metric Formulation of the EDE and the EVI. The main definition is that of a *gradient flow* in the energy landscape of a functional $\mathcal{E}_t : [0,\infty) \times \mathbf{X} \to \mathbb{R} \cup \{\infty\}$ with respect to the metric $\boldsymbol{d}$. Here we adopt the notions of the *energy dissipation equality* (EDE) from (1.0.3) and of the *evolution variational inequality* (EVI) from (1.0.4) to non-autonomous versions.

Definition 2.3.3. Given a proper free energy functional $\mathcal{E}_t : [0,\infty) \times \mathbf{X} \to \mathbb{R} \cup \{\infty\}$ and some initial datum $u_0 \in \mathcal{D}(\mathcal{E}_0)$.

We say that $u_t \in \mathrm{AC}^2(0,\infty; (\mathbf{X}, \boldsymbol{d}))$ is a *curve of steepest descent* with respect to $\mathcal{E}_t$ emanating from $u_0 \in \mathcal{D}(\mathcal{E}_0)$ if and only if one of the following holds.

a) **Energy Dissipation Equality.** The following energy balance holds for all $T > 0$:

$$\mathcal{E}_T(u_T) + \frac{1}{2}\int_0^T |u_t'|^2 \,\mathrm{d}t + \frac{1}{2}\int_0^T |\partial\mathcal{E}_t|^2(u_t)\,\mathrm{d}t = \mathcal{E}_0(u_0) + \int_0^T \partial_t\mathcal{E}_t(u_t)\,\mathrm{d}t. \quad (2.3.1)$$

b) **Evolution Variational Inequality.** For arbitrary $0 \le s \le t$ and for every reference point $w \in \mathcal{D}(\mathcal{E}_r)$ at each $r \in [0,\infty)$ the following holds:

$$\frac{1}{2}\boldsymbol{d}^2(w,u_t) - \frac{1}{2}\boldsymbol{d}^2(w,u_s) \le \int_s^t \left[\mathcal{E}_r(w) - \mathcal{E}_r(u_r) - \frac{\lambda}{2}\boldsymbol{d}^2(u_r,w)\right]\mathrm{d}r. \quad (2.3.2)$$

In this case we say that the free energy functional $\mathcal{E}_t$ generates a time-dependent λ-contractive gradient flow on $(\mathbf{X}, \boldsymbol{d})$.

Remark 2.3.4. The EVI is a more restrictive characterization of gradient flows than the EDE. Most notably, the validity of the EVI implies that the gradient flow is λ-contractive on $(\mathbf{X}, \boldsymbol{d})$, so in particular, solutions are uniquely determined by their initial datum. Moreover, if the metric space $(\mathbf{X}, \boldsymbol{d})$ is “almost Euclidean” — for instance, $\mathbf{X}$ is a Hilbert space, or $\mathbf{X}$ is the space $\mathcal{P}_2(\Omega)$ of probability measures endowed with the Wasserstein metric $\mathbf{W}_2$ — then if $\mathcal{E}$ generates a λ-contractive gradient flow then $\mathcal{E}$ is uniformly semi-convex see [28]. Thus, (2.3.2) is not available for gradient flows of non-semi-convex functionals $\mathcal{E}_t$.

2.4 Gradient Flow Theory in the L^2-Wasserstein Space

As our main application of the temporal discretization and approximations in the general framework of abstract metric spaces, we consider two particular classes of L^2-Wasserstein gradient flows, namely: the non-autonomous and non-linear drift-diffusion equation, also called Fokker-Planck (FP) equation,

$$\partial_t \rho_t = \Delta \rho_t^m + \operatorname{div}(\rho_t \nabla V_t) + \operatorname{div}(\rho_t (\nabla W_t * \rho_t)), \tag{2.4.1}$$

with no-flux boundary condition in a domain Ω, as before; and the Derrida-Lebowitz-Speer-Spohn (DLSS) equation

$$\partial_t \rho_t = -\operatorname{div}\Big(\rho_t \nabla \Big(2 \frac{\Delta\sqrt{\rho_t}}{\sqrt{\rho_t}}\Big)\Big), \tag{2.4.2}$$

with no-flux boundary condition in a domain Ω, which is additionally convex. In both cases, the sought-for solution $\rho_t : [0,\infty) \times \Omega \to [0,\infty]$ should be non-negative and preserve mass.

Gradient Flow Structure. In the seminal work of Jordan et al. [54], it has been used that (2.4.1) possesses a gradient flow structure in the L^2-Wasserstein space, see the monographs [4, 87, 92] for a comprehensive introduction to this theory. To be more precise, define the free energy functional $\mathcal{E}_t : [0,\infty) \times \mathcal{P}_2(\Omega) \to \mathbb{R} \cup \{\infty\}$ via

$$\mathcal{E}_t(\mu) := \begin{cases} \int_\Omega \rho \log(\rho) + V_t \rho + \frac{1}{2}(W_t * \rho)\rho \,\mathrm{d}x & \text{if } m = 1, \\ \int_\Omega \frac{1}{m-1}\rho^m + V_t \rho + \frac{1}{2}(W_t * \rho)\rho \,\mathrm{d}x & \text{if } m > 1, \end{cases} \tag{2.4.3}$$

if the measure $\mu = \rho \,\mathrm{d}\mathcal{L}^{\mathrm{d}} \in \mathcal{P}_2^{\mathrm{ac}}(\Omega)$ and otherwise we set $\mathcal{E}_t(\mu) := \infty$. Then, (2.4.1) is equivalent to the coupling of the continuity equation with Darcy's law where the pressure is given by the variational derivative of free energy functional $\mathcal{E}$:

$$\partial_t \rho_t + \operatorname{div}(\rho_t \mathbf{v}_t) = 0, \qquad \mathbf{v}_t = -\mathrm{D}\frac{\delta \mathcal{E}(\rho_t)}{\delta \rho}. \tag{2.4.4}$$

To shorten notion, we introduce the following abbreviation for the different parts of the free energy functional $\mathcal{E}_t$ from (2.4.3), namely: the Boltzman entropy $\mathcal{H}$, the internal energy $\mathcal{U}_m$, the confinement energy $\mathcal{V}_t$, and the interaction energy $\mathcal{W}_t$, all defined by:

$$\mathcal{H}(\mu) := \mathcal{U}_1(\mu) := \int_\Omega \rho(x)\log(\rho(x))\,\mathrm{d}x, \quad \mathcal{U}_m(\rho) := \frac{1}{m-1}\int_\Omega (\rho(x))^m \,\mathrm{d}x,$$
$$\mathcal{V}_t(\rho) := \int_\Omega V_t(x)\,\mathrm{d}\rho(x), \qquad \mathcal{W}_t(\rho) := \frac{1}{2}\int_\Omega\int_\Omega W_t(x-y)\,\mathrm{d}\rho(y)\,\mathrm{d}\rho(x),$$

given a confinement potential $V_t : [0,\infty) \times \Omega \to \mathbb{R}$ and a symmetric interaction potential $W_t : [0,\infty) \times \mathbb{R}^{\mathrm{d}} \to \mathbb{R}$ and if everything is well-defined, otherwise set the value to $+\infty$.

Similarly, fourth order equations like the Derrida-Lebowitz-Speer-Spohn equation (2.4.2) or the Hele-Shaw flow (or general interpolations of these two equations) possess this variational structure, see [46, 47, 68, 78]. The corresponding free energy functional $\mathcal{E}$ for the DLSS equation (2.4.2) is given by the Fisher information:

$$\mathcal{E}(\mu) := \mathcal{I}(\mu) = \int_\Omega \|\nabla\sqrt{\rho}\|^2 \,\mathrm{d}x \tag{2.4.5}$$

provided that $\mu = \rho\,\mathrm{d}\mathcal{L}^\mathrm{d} \in \mathcal{P}_2^\mathrm{ac}(\Omega)$ and $\sqrt{\rho} \in \mathrm{H}^1(\Omega)$, otherwise we set $\mathcal{E}(\rho) = \infty$.

Variations in the L^2-Wasserstein Space. A key tool introduced by Jordan et al. [54] and Otto [78] was that the natural notion of variation in the L^2-Wasserstein space is to perturb measures $\rho_* \in \mathcal{P}_2(\Omega)$ along solutions of the transport equation (2.4.6). Especially having the dynamic formulation (2.2.3) of the L^2-Wasserstein distance in mind, this perturbation is clearly evident.

Definition 2.4.1 (Variation Along the Transport Equation)**.** Given an initial datum $\rho_* \in \mathcal{P}_2^\mathrm{ac}(\Omega)$ and a vector field $\xi \in C_c^\infty(\overline{\Omega}, \mathbb{R}^\mathrm{d})$ with $\xi \cdot \boldsymbol{n} = 0$ on $\partial\Omega$. We define the perturbation ρ^s along the vector field ξ in the auxiliary time s as the solution to the transport equation

$$\partial_s \rho^s + \mathrm{div}(\rho^s\,\xi) = 0, \qquad \rho^0 = \rho_*. \tag{2.4.6}$$

The solution ρ^s is explicitly given by the push-forward of ρ_* under the flow map X^s, i.e., $\rho^s = (\mathrm{X}^s)_\#\rho_*$, such that the flow map X^s satisfies the initial value problem:

$$\frac{\mathrm{d}}{\mathrm{d}s}\mathrm{X}^s(x) = \xi\left(\mathrm{X}^s(x)\right), \qquad \mathrm{X}^0(x) = x.$$

Note that the flow map X^s exists and for each s the flow map X^s is a diffeomorphism on Ω, cf. [14, 47, 77] . Additionally, we have an explicit representation of the perturbed density ρ^s and we can calculate the derivative of $\det(\mathrm{D}\mathrm{X}^s)$, i.e.,

$$\det(\mathrm{D}\mathrm{X}^s)\rho^s \circ \mathrm{X}^s = \rho_*, \quad \text{and} \quad \frac{\mathrm{d}}{\mathrm{d}s}\left[\det(\mathrm{D}\mathrm{X}^s(x))\right]_{s=0} = \mathrm{tr}(\mathrm{D}\xi \circ \mathrm{X}_0) = \mathrm{div}(\xi). \tag{2.4.7}$$

To calculate the first variation of the different energies is a by now standard calculation, see [4, 92, 93] for more details. The first result concerns the differentiability of the L^2-Wasserstein distance if all measures are absolutely continuous.

Lemma 2.4.2. *Let $\eta, \rho_* \in \mathcal{P}_2^{ac}(\Omega)$ and consider the perturbation ρ^s with respect to the transport equation* (2.4.6) *for some given vector field $\xi \in C_c^\infty(\overline{\Omega}, \mathbb{R}^\mathrm{d})$ with $\xi \cdot \boldsymbol{n} = 0$ on $\partial\Omega$. Then, the map $s \mapsto \mathbf{W}_2^2(\eta, \rho^s)$ is differentiable at $s = 0$ with derivative*

$$\frac{\mathrm{d}}{\mathrm{d}s}\left[\mathbf{W}_2^2(\eta, \rho^s)\right]_{s=0} = 2\int_{\Omega^2} \langle \xi(x), x - y\rangle \,\mathrm{d}\boldsymbol{p}(x,y),$$

where $\boldsymbol{p} \in \Gamma(\rho_*, \eta)$ is an optimal transport plan from ρ_* to η.

The first variation of the functionals in the definition of the free energy functional $\mathcal{E}_t$ corresponding to the Fokker-Planck equation (2.4.1) is given as follows.

Lemma 2.4.3. *Let $\rho_* \in \mathcal{P}_2(\Omega)$ and consider the perturbation ρ^s with respect to the transport equation* (2.4.6) *for some given vector field $\xi \in \mathcal{C}_c^\infty(\overline{\Omega}, \mathbb{R}^d)$ with $\xi \cdot \boldsymbol{n} = 0$ on $\partial\Omega$. Then, the first variations of the confinement energy $\mathcal{V}_t$ and the interaction energy $\mathcal{W}_t$ are given by:*

$$\frac{\mathrm{d}}{\mathrm{d}s}\left[\mathcal{V}_t(\rho_s)\right]_{s=0} = \int_\Omega \langle \xi, \nabla V\rangle \,\mathrm{d}\rho_*(x),$$
$$\frac{\mathrm{d}}{\mathrm{d}s}\left[\mathcal{W}_t(\rho_s)\right]_{s=0} = \int_\Omega\int_\Omega \langle \xi(x), \nabla W_t(x-y)\rangle \,\mathrm{d}\rho_*(x)\,\mathrm{d}\rho_*(y).$$

If additionally $\rho \in \mathcal{P}_2^{ac}(\Omega) \cap \mathrm{L}^m(\Omega)$, then the first derivatives at $s = 0$ of the Boltzman entropy $\mathcal{H}$ and the internal energy $\mathcal{U}_m$ energy are given by:

$$\frac{\mathrm{d}}{\mathrm{d}s}\left[\mathcal{H}(\rho_s)\right]_{s=0} = -\int_\Omega \mathrm{div}(\xi)\,\rho_*\,\mathrm{d}x,$$
$$\frac{\mathrm{d}}{\mathrm{d}s}\left[\mathcal{U}_m(\rho_s)\right]_{s=0} = -\int_\Omega \mathrm{div}(\xi)\,\rho_*^m\,\mathrm{d}x,$$

For example, to actually compute the first variation of the heat energy $\mathcal{H}$ we can simplify the difference quotient with the explicit representation of the perturbed density ρ^s and the change of variables $x = \mathrm{X}^s(y)$ as follows

$$\begin{aligned}\frac{1}{s}\left(\mathcal{H}(\rho^s) - \mathcal{H}(\rho_*)\right) &= \frac{1}{s}\Big(\int_\Omega \rho^s \circ \mathrm{X}^s \log(\rho^s \circ \mathrm{X}^s)\det(\mathrm{D}\mathrm{X}^s)\,\mathrm{d}y - \int_\Omega \rho_* \log(\rho_*)\,\mathrm{d}x\Big)\\ &= -\frac{1}{s}\int_\Omega \log(\det(\mathrm{D}\mathrm{X}^s))\,\rho_*\,\mathrm{d}x.\end{aligned}$$

The pointwise limit of the integrand is given by (2.4.7) and we can conclude by a dominated convergence argument

$$\frac{\mathrm{d}}{\mathrm{d}s}\left[\mathcal{H}(\rho_s)\right]_{s=0} = -\int_\Omega \mathrm{div}(\xi)\,\rho_*\,\mathrm{d}x.$$

The differentiability of the Fisher information along solutions to the transport equation has been proven by Gianazza et al. [47, Theorem 4.2] for Ω open, bounded, and convex and by Matthes et al. [68, Lemma 2.5] for $\Omega = \mathbb{R}^d$.

Lemma 2.4.4. *Let $\rho_* \in \mathcal{D}(\mathcal{I})$ and consider the perturbation ρ^s with respect to the transport equation* (2.4.6) *for some given vector field $\xi \in \mathcal{C}_c^\infty(\overline{\Omega}, \mathbb{R}^d)$ with $\xi \cdot \boldsymbol{n} = 0$ on $\partial\Omega$. Then, the first derivative at $s = 0$ of the Fisher information is given by*

$$\frac{\mathrm{d}}{\mathrm{d}s}\left[\mathcal{I}(\rho^s)\right]_{s=0} = -\int_\Omega \frac{1}{2}\langle \nabla(\mathrm{div}\,\xi), \nabla\rho_*\rangle + 2\langle \mathrm{D}\xi\nabla\sqrt{\rho_*}, \nabla\sqrt{\rho_*}\rangle\,\mathrm{d}x.$$

Variations Along the Heat Flow. To derive more information on the L^2-Wasserstein subdifferential of the free energy functionals $\mathcal{E}$ corresponding to the FP equation (2.4.1) or the DLSS equation (2.4.2) we evaluate its derivative along the flow generated by the heat equation starting from a given initial configuration ρ_* with $\mathcal{E}(\rho_*) < +\infty$, see Gianazza et al. [47]. These results have inspired Matthes et al. in [68] to develop the flow interchange technique which allows deducing better a priori bounds for the JKO minimizers. This variation plays a major role in the *Weighted Energy-Dissipation principle* chapter 6 of this thesis.

Definition 2.4.5 (Variations Along the Heat Flow)**.** Given an open, bounded, and convex domain Ω with Lipschitz-continuous boundary $\partial\Omega$ or let $\Omega = \mathbb{R}^{\mathrm{d}}$. Define ρ^s as perturbation of $\rho_* \in \mathcal{P}_2^{\mathrm{ac}}(\Omega)$ as the solution to the heat equation

$$\partial_s \rho^s = \eta\Delta(\rho^s), \qquad \rho^0 = \rho_* \tag{2.4.8}$$

with no-flux boundary condition and for some given diffusivity parameter $\eta \geq 0$.

By the parabolic regularity theory, it is clear, that there exists a smooth and non-negative solution ρ^s. Further, one has an explicit representation of the solutions ρ^s with respect to some Greens function $\mathfrak{G}^s_\eta : [0,\infty) \times \Omega \times \Omega \to [0,\infty]$, i.e., $\rho^s(x) = \int_\Omega \mathfrak{G}^s_\eta(x,y)\rho_*(y)\,\mathrm{d}y$.

Remark 2.4.6. It is also well known, that the heat equation is a 0-contractive gradient flow with respect to the Boltzmann entropy $\mathcal{H}$ in the L^2-Wasserstein space, since by a formal calculation ρ^s solves the continuity equation with velocity field $\mathbf{v}^s = \eta\nabla\log(\rho^s)$. Hence, the map $s \mapsto \mathbf{W}_2^2(\nu,\rho^s)$ is absolutely continuous and ρ^s solves the evolution variational inequality

$$\frac{1}{2}\mathbf{W}_2^2(\nu,\rho^t) - \frac{1}{2}\mathbf{W}_2^2(\nu,\rho^s) \leq \eta \int_s^t \mathcal{H}(\nu) - \mathcal{H}(\rho^r)\,\mathrm{d}r.$$

One can even quantify the derivative of $s \mapsto \mathbf{W}_2^2(\eta,\rho^s)$ since by the regularizing effects of the heat equation the velocity $\eta\nabla\log(\rho^s)$ is sufficiently regular to apply the formulas given in [87, Corollary5.25], [4, Corollary 10.2.7], or [92, Theorem 8.13]. Then, one has for all almost every $s > 0$

$$\frac{\mathrm{d}}{\mathrm{d}s}\mathbf{W}_2^2(\nu,\rho^s) = \eta\int_\Omega \langle x - \mathrm{T}^s_{opt}(x), \nabla\log(\rho^s)\rangle\,\mathrm{d}\rho^s(x) \tag{2.4.9}$$

where T^s_{opt} is the optimal transport map from ρ^s to ν.

The differentiability of the L^2-Wasserstein distance at regular measures $\rho_*, \nu \in \mathcal{P}_2^{\mathrm{ac}}(\Omega)$ with finite Fisher information along the heat flow is a consequence of (2.4.9).

Lemma 2.4.7. *Let $\rho_*, \nu \in \mathcal{P}_2^{ac}(\Omega)$ with $\rho_* \in \mathcal{D}(\mathcal{I})$ and let ρ^s be the associated solution to the heat flow* (2.4.8)*. Then*

$$\limsup_{s\to 0}\frac{1}{s}\left|\mathbf{W}_2^2(\nu,\rho^s) - \mathbf{W}_2^2(\nu,\rho_*)\right| \leq \frac{\eta}{2}\mathbf{W}_2(\eta,\rho_*)\sqrt{\mathcal{I}(\rho_*)}.$$

Proof. By the previous remark 2.4.6 the map $s \mapsto \mathbf{W}_2^2(\eta, \rho^s)$ is absolutely continuous and the derivative of this map at $s > 0$ is given by (2.4.9). Therefore, we obtain for the difference quotient

$$\frac{1}{s}\left[\mathbf{W}_2^2(\nu,\rho^s) - \mathbf{W}_2^2(\nu,\rho_*)\right] = \eta\frac{1}{s}\int_0^s \int_\Omega \langle x - \mathrm{T}^s_{opt}(x), \nabla\log(\rho^s)\rangle \,\mathrm{d}\rho^s(x)\,\mathrm{d}s$$

where T^s_{opt} is the optimal transport map from ρ^s to ν. Applying the Cauchy-Schwarz inequality to the weighted integral over Ω with weight ρ^s yields

$$\begin{aligned}&\frac{1}{s}\left|\mathbf{W}_2^2(\nu,\rho^s) - \mathbf{W}_2^2(\nu,\rho_*)\right| \\ \leq&\eta\frac{1}{s}\int_0^s \Big(\int_\Omega \left\|x - \mathrm{T}^s_{opt}(x)\right\|^2 \mathrm{d}\rho^s(x)\Big)^{1/2}\Big(\int_\Omega \|\nabla\log(\rho^s)\|^2 \,\mathrm{d}\rho^s(x)\Big)^{1/2}\mathrm{d}s.\end{aligned}$$

By the definition of T^s_{opt}, the first inner integral is equals to $\mathbf{W}_2(\nu,\rho^s)$ whereas the second inner integral is equals to $\frac{1}{2}\sqrt{\mathcal{I}(\rho^s)}$. Note, the Fisher information is decreasing along the heat flow ρ^s, i.e., $\mathcal{I}(\rho^s) \leq \mathcal{I}(\rho_*) < \infty$. In conclusion, we obtain

$$\begin{aligned}\limsup_{s\to 0}\frac{1}{s}\left|\mathbf{W}_2^2(\nu,\rho^s) - \mathbf{W}_2^2(\nu,\rho_*)\right| \leq& \limsup_{s\to 0}\frac{\eta}{2}\frac{1}{s}\int_0^s \mathbf{W}_2(\nu,\rho^s)\,\sqrt{\mathcal{I}(\rho_*)}\,\mathrm{d}s \\ =&\frac{\eta}{2}\mathbf{W}_2(\nu,\rho_*)\sqrt{\mathcal{I}(\rho_*)}\end{aligned}$$

where we used in the last equality the continuity of $s \mapsto \mathbf{W}_2^2(\nu,\rho^s)$. □

The derivative of the free energy functional $\mathcal{E}$ with respect to the Fokker-Planck equation is given as follows.

Lemma 2.4.8. *Let $\rho_* \in \mathcal{P}_2^{ac}(\Omega)$ and let ρ^s be the perturbation of ρ_* according to the heat flow* (2.4.8). *If the right derivative $\limsup_{s\to 0}\frac{1}{s}\left[\mathcal{U}_m(\rho^s) - \mathcal{U}_m(\rho_*)\right]$ is finite, then $\rho_*^{m/2} \in \mathrm{H}^1(\Omega)$ and*

$$\limsup_{s\to 0}\frac{1}{s}\left[\mathcal{U}_m(\rho^s) - \mathcal{U}_m(\rho_*)\right] = -\frac{4\eta}{m}\int_\Omega \left\|\nabla(\rho_*^{m/2})\right\|^2 \mathrm{d}x.$$

The derivatives at $s = 0$ of the $\mathcal{V}$ and the interaction energy $\mathcal{W}$ are given by

$$\frac{\mathrm{d}}{\mathrm{d}s}\left[\mathcal{V}(\rho^s)\right]_{s=0} = \eta\int_\Omega \Delta V(x)\,\mathrm{d}\rho_*(x), \quad \frac{\mathrm{d}}{\mathrm{d}s}\left[\mathcal{W}(\rho^s)\right]_{s=0} = \eta\int_\Omega\int_\Omega \Delta W(x-y)\,\mathrm{d}\rho_*(y)\,\mathrm{d}\rho_*(x).$$

Likewise, one can compute the first variation of the Fisher information along the heat flow, see Theorem 5.1 [47] for Ω open, bounded, and convex and Lemma 4.4 [68] for $\Omega = \mathbb{R}^{\mathrm{d}}$.

Lemma 2.4.9. *Let $\rho_* \in \mathcal{D}(\mathcal{I})$ and let ρ^s be the associated solution to the heat flow* (2.4.8). *If the right derivative $\limsup_{s\to 0}\frac{1}{s}\left[\mathcal{I}(\rho^s) - \mathcal{I}(\rho_*)\right]$ is finite, then $\sqrt{\rho_*} \in \mathrm{H}^2(\Omega)$ and*

$$\limsup_{s\to 0}\frac{1}{s}\left[\mathcal{I}(\rho^s) - \mathcal{I}(\rho_*)\right] \leq -C\eta\int_\Omega \left\|\mathrm{D}^2\sqrt{\rho_*}\right\|^2 \mathrm{d}x$$

where $C > 0$ is some universal constant depending only on d and Ω.

2.5 Auxiliary Theorems

Extension of the Aubin-Lions Theorem. We state the main tool to derive convergence of discretizations or approximations of gradient flows. This auxiliary theorem is an extension of the Aubin-Lions compactness Theorem to metric spaces, cf. [85, Theorem 2]. So given a complete, separable metric space $(\mathbf{X}, \boldsymbol{d})$ we define suitable surrogates for *tighness* and *integral equi-continuity* which ensures that a sequence $(u_t^n)_{n\in\mathbb{N}}$ of measurable $\mathbf{X}$-valued functions converges in $\mathcal{M}(0,T;(\mathbf{X},\boldsymbol{d}))$.

Definition 2.5.1 (Normal Coercive Integrand&Pseudo-distance)**.** Given an auxiliary functional $\mathcal{A}_t : (0,T) \times \mathbf{X} \to \mathbb{R} \cup \{\infty\}$ and a positive function $g : \mathbf{X} \times \mathbf{X} \to \mathbb{R} \cup \{\infty\}$. Then, $\mathcal{A}_t$ is called normal coercive integrand if:

a) $\mathcal{A}_t$ is $\mathcal{L}(0,T) \otimes \mathcal{B}(\mathbf{X})$-measurable;

b) the map $u \mapsto \mathcal{A}_t(u)$ is lower semi-$\boldsymbol{d}$-continuous with for each $t \in [0,T]$;

c) the map $u \mapsto \mathcal{A}_t(u)$ has compact sublevels in $\mathbf{X}$ with respect to the topology induced by the distance $\boldsymbol{d}$ for each $t \in [0,T]$.

We call g a pseudo-distance on $\mathbf{X}$ with respect to the auxiliary functional $\mathcal{A}_t$ if:

a) $g(u,v) = 0$ for $u, v \in \mathcal{D}(\mathcal{A}_t)$ implies $u = v$;

b) the map $(u,v) \mapsto g(u,v)$ is lower semi-$\boldsymbol{d}$-continuous.

The two main examples of normal coercive integrands and pseudo-distances which will be used in this thesis are as follows:

Lemma 2.5.2. *Let $\Omega \subset \mathbb{R}^{\mathrm{d}}$ be an open, bounded and connected domain with Lipschitz-continuous boundary $\partial\Omega$ and consider $\mathbf{X} = \mathrm{L}^m(\Omega)$ with the strong topology. Then, the following auxiliary functional $\mathcal{A} : \mathrm{L}^m(\Omega) \to [0,\infty]$ and positive function $g : \mathrm{L}^m(\Omega) \times \mathrm{L}^m(\Omega) \to [0,\infty]$ defined via*

$$\mathcal{A}(\varphi) := \begin{cases} \|\varphi^m\|_{\mathrm{BV}(\Omega)} & \text{if } \varphi^m \in \mathrm{BV}(\Omega) \text{ and } \varphi \in \mathcal{P}_2^{ac}(\Omega), \\ +\infty & \text{else,} \end{cases}$$

$$g(\varphi,\psi) := \begin{cases} \mathbf{W}_2(\varphi,\psi) & \text{if } \varphi,\psi \in \mathcal{P}_2^{ac}(\Omega), \\ +\infty & \text{else,} \end{cases}$$

are a normal coercive integrand and a pseudo-distance in the sense of definition 2.5.1.

Proof. The auxiliary functional $\mathcal{A}$ is clearly $\mathcal{B}(\mathrm{L}^m(\Omega))$-measurable. The lower semi-continuity can be derived as follows. Given a sequence $(\varphi_n)_{n\in\mathbb{N}} \subset \mathrm{L}^m(\Omega)$ converging in the strong $\mathrm{L}^m(\Omega)$-topology to a limit function $\varphi_* \in \mathrm{L}^m(\Omega)$. Since the $\mathrm{BV}(\Omega)$-norm is lower semi-continuous with respect to strong $\mathrm{L}^1(\Omega)$-convergence the map $\varphi \mapsto \|\varphi^m\|_{\mathrm{BV}(\Omega)}$ is lower semi-continuous in the strong $\mathrm{L}^m(\Omega)$-topology. Next, we derive the compactness of the sublevels of $\mathcal{A}$ by Rellich's compactness theorem. I.e., for any sequence $(\varphi_n)_{n\in\mathbb{N}} \subset$

$\mathrm{L}^m(\Omega)$ with $\sup_n \|\varphi_n^m\|_{\mathrm{BV}(\Omega)} < \infty$ we can extract a (non-relabeled) subsequence such that $(\varphi_n)_{n\in\mathbb{N}}$ converges in the strong $\mathrm{L}^m(\Omega)$-topology to a limit function $\varphi_* \in \mathrm{L}^m(\Omega)$.

Lastly, we prove that g is indeed a pseudo-distance on $\mathrm{L}^m(\Omega)$. Given two functions $\varphi, \psi \in \mathrm{L}^m(\Omega) \cap \mathcal{P}_2^{\mathrm{ac}}(\Omega)$, then it follows from $0 = g(\varphi,\psi) = \mathbf{W}_2(\varphi,\psi)$ that $\varphi = \psi$ in the sense of measures and clearly also $\varphi = \psi$ almost every where since $\varphi, \psi \in \mathcal{P}_2^{\mathrm{ac}}(\Omega)$. Hence $\varphi = \psi$ in $\mathrm{L}^m(\Omega)$. The joint lower semi-continuity of g follows also from the fact, that the $\mathrm{L}^m(\Omega)$-topology is finer than the topology induced by the weak*-convergence of measures if Ω is bounded, i.e., if $\varphi_n \to \varphi_*$ in $\mathrm{L}^m(\Omega)$ then also $\varphi_n \rightharpoonup^* \varphi_*$. Since the L^2-Wasserstein distance is lower semi-continuous with respect to the weak*-convergence of measures, we obtain the desired result. □

Lemma 2.5.3. *Let $\Omega \subset \mathbb{R}^{\mathrm{d}}$ be an open, bounded and connected domain with Lipschitz-continuous boundary $\partial\Omega$ or let $\Omega = \mathbb{R}^{\mathrm{d}}$ and consider $\mathbf{X} = \mathrm{L}^1(\Omega)$. Then, the following auxiliary functional $\mathcal{A} : \mathrm{L}^1(\Omega) \to [0,\infty]$ and a positive function $g : \mathrm{L}^1(\Omega) \times \mathrm{L}^1(\Omega) \to [0,\infty]$ defined via*

$$\mathcal{A}(\varphi) := \begin{cases} \|\sqrt{\varphi}\|_{\mathrm{H}^1(\Omega)} + \boldsymbol{M}_2(\varphi) & \text{if } \sqrt{\varphi} \in \mathrm{H}^1(\Omega) \text{ and } \varphi \in \mathcal{P}_2^{ac}(\Omega), \\ +\infty & \text{else,} \end{cases}$$

$$g(\varphi,\psi) := \begin{cases} \mathbf{W}_2(\varphi,\psi) & \text{if } \varphi,\psi \in \mathcal{P}_2^{ac}(\Omega), \\ +\infty & \text{else,} \end{cases}$$

are a normal coercive integrand and a pseudo-distance in the sense of definition 2.5.1.

Proof. Also in this case, $\mathcal{A}$ is $\mathcal{B}(\mathrm{L}^1(\Omega))$-measurable. To prove the lower semi-continuity of $\mathcal{A}$ consider a sequence $(\varphi_n)_{n\in\mathbb{N}} \subset \mathrm{L}^1(\Omega)$ converging in the strong $\mathrm{L}^1(\Omega)$-topology to a limit function $\varphi_* \in \mathrm{L}^1(\Omega)$. Since the map $\varphi \mapsto \|\varphi\|_{\mathrm{H}^1(\Omega)}$ is lower semi-continuous with respect to the strong $\mathrm{L}^2(\Omega)$-convergence and since the mapping $\varphi \mapsto \sqrt{\varphi}$ is $\mathrm{L}^1(\Omega)$-$\mathrm{L}^2(\Omega)$-continuous, we obtain the lower semi-continuity with respect to the strong $\mathrm{L}^1(\Omega)$-topology of the map $\varphi \mapsto \|\sqrt{\varphi}\|_{\mathrm{H}^1(\Omega)}$. The lower semi-continuity of the map $\varphi \mapsto \boldsymbol{M}_2(\varphi)$ with respect to the strong $\mathrm{L}^1(\Omega)$-topology follows by the fact that convergence in the strong $\mathrm{L}^1(\Omega)$ convergence implies weak*-convergence in the sense of measures. Since $\boldsymbol{M}_2$ is lower semi-continuous with respect to the weak*-convergence, $\boldsymbol{M}_2$ is also lower semi-continuous with respect to the $\mathrm{L}^1(\Omega)$-topology. Hence, the auxiliary functional $\mathcal{A}$ is lower semi-continuous with respect to the strong $\mathrm{L}^1(\Omega)$-topology. The compactness of the sublevels of $\mathcal{A}$ is split into two parts. Firstly, consider an open, bounded and connected subset $\Omega \subset \mathbb{R}^{\mathrm{d}}$. Fix some C and consider a sequence $(\varphi_n)_{n\in\mathbb{N}} \subset \mathrm{L}^1(\Omega)$ with $\sup_n \mathcal{A}(\varphi_n) < \infty$. Define the auxiliary sequence $(u_n)_{n\in\mathbb{N}} \subset \mathrm{L}^2(\Omega)$ by $u_n := \sqrt{\varphi_n}$. With this definition we clearly have $\sup_n \|u_n\|_{\mathrm{H}^1(\Omega)} < \infty$ and hence we can conclude with Rellich's compactness theorem that $(u_n)_{n\in\mathbb{N}}$ converges strongly in the $\mathrm{L}^2(\Omega)$-topology to a limit $u_* \in \mathrm{L}^2(\Omega)$. Clearly, we also have $\varphi_n \to (u_*)^2$ strongly in $\mathrm{L}^1(\Omega)$. The case $\Omega = \mathbb{R}^{\mathrm{d}}$ is already considered in [68, Lemma 2.2]. By the same arguments as before, we can extract for every sequence $(\varphi_n)_{n\in\mathbb{N}} \subset \mathrm{L}^1(\Omega)$ a (non-relabeled) subsequence such that there is a limit function $\varphi_* \in \mathrm{L}^1(\mathbb{R}^\mathrm{d})$ with

$$\varphi_n \rightharpoonup^* \varphi_* \qquad \text{and} \qquad \varphi_n \to \varphi_* \quad \text{strongly in } \mathrm{L}^1(K) \text{ for each compact subset } K \Subset \mathbb{R}^\mathrm{d}.$$

Since $\|\varphi_n\|_{\mathrm{L}^1(\mathbb{R}^\mathrm{d})} = 1 = \|\varphi_*\|_{\mathrm{L}^1(\mathbb{R}^\mathrm{d})}$ one can conclude the strong convergence in the $\mathrm{L}^1(\mathbb{R}^\mathrm{d})$-topology.

That g is a pseudo-distance on $\mathrm{L}^1(\Omega)$ can be derived as follows. Clearly $0 = g(\varphi, \psi) = \mathbf{W}_2(\varphi, \psi)$ implies $\varphi = \psi$ in the sense of measures and almost everywhere. By the same argument as in the previous case, the pseudo-distance g is lower semi-continuous with respect to the $\mathrm{L}^1(\Omega)$-convergence. □

Having defined normal coercive integrands $\mathcal{A}$ and pseudo-distances g we can state the extension of Aubin-Lions compactness Theorem to metric spaces, cf. [85, Theorem 2].

Theorem 2.5.4 (Extension of the Aubin-Lions Theorem)**.** *Let $(\mathbf{X}, \boldsymbol{d})$ be a complete, separable metric space, let $\mathcal{A}_t : (0,T) \times \mathbf{X} \to \mathbb{R} \cup \{\infty\}$ be a normal coercive integrand, and let $g : \mathbf{X} \times \mathbf{X} \to \mathbb{R} \cup \{\infty\}$ be a pseudo-distance on $\mathbf{X}$. Let $(u_t^n)_{n\in\mathbb{N}}$ be a sequence of measurable functions $u_t^n : (0,T) \to \mathbf{X}$ such that*

$$\sup_{n\in\mathbb{N}} \int_0^T \mathcal{A}_t(u_t^n)\,\mathrm{d}t < \infty, \qquad \lim_{h\searrow 0} \sup_{n\in\mathbb{N}} \int_0^{T-h} g\left(u_{t+h}^n, u_t^n\right)\,\mathrm{d}t = 0. \tag{2.5.1}$$

Then, $(u_t^n)_{n\in\mathbb{N}}$ possesses a subsequence converging in $\mathcal{M}\left(0,T; (\mathbf{X}, \boldsymbol{d})\right)$.

Remark 2.5.5. Note, one can replace the usual weak integral equi-continuity condition in (2.5.1) given in the original version of the theorem by the relaxed averaged weak integral equi-continuity

$$\liminf_{h\searrow 0} \limsup_{n\to\infty} \frac{1}{h} \int_0^h \int_0^{T-s} g(u_{t+s}^n, u_t^n)\,\mathrm{d}t\,\mathrm{d}s = 0. \tag{2.5.2}$$

This condition is sufficient in the proof of Theorem 2.5.4.

Monotonicity of the Kinetic Energy. It is worthwhile to point out that the kinetic energy $\mathcal{K}$ shares also a monotonicity property with respect to non-negative integral kernels $\mathfrak{G}$. Define the integral transform $\mathcal{T}_\mathfrak{G}[\rho]$ with respect to the measurable integral kernel $\mathfrak{G}$ for a scalar measure $\rho : \Omega \to \mathbb{R}$ by

$$\mathcal{T}_\mathfrak{G}[\rho](x) = \int_\Omega \mathfrak{G}(x,y)\,\mathrm{d}\rho(y).$$

For a $\mathbb{R}^\mathrm{d}$-valued measure $\mathbf{w} \in \mathscr{W}(\Omega)$ we define the vector integral transform $\vec{\mathcal{T}}_\mathfrak{G}[\mathbf{w}]$ componentwise, i.e., $\vec{\mathcal{T}}_\mathfrak{G}[\mathbf{w}] := (\mathcal{T}_\mathfrak{G}[\mathbf{w}_i])_i$.

Lemma 2.5.6. *Let $\rho \in \mathcal{P}_2^{ac}(\Omega)$ and let $\mathbf{w} \in \mathscr{W}(\Omega)$ be a $\mathbb{R}^{\mathrm{d}}$-valued signed Borel-measure. Then,*

$$\mathcal{K}\big(\mathcal{T}_{\mathfrak{G}}[\rho], \vec{\mathcal{T}}_{\mathfrak{G}}[\mathbf{w}]\big) \le \mathcal{K}(\rho, \mathbf{w}) \tag{2.5.3}$$

for any non-negative measurable integral kernel $\mathfrak{G} : \Omega \times \Omega \to [0, \infty]$ with $\int_\Omega \mathfrak{G}(x, y)\, \mathrm{d}x = 1$ for each $y \in \Omega$.

The proof of this lemma is almost identical to the proof in of Lemma 8.1.10 in [4] where one replaces $\theta = \rho(x - \cdot)\mu$ by $\theta = \mathfrak{G}(x, \cdot)\rho$. For the sake of completeness, we shall give the proof in the following.

Proof. We use Jensen inequality in the following form: if $\Phi : \mathbb{R}^{m+1} \to [0, \infty]$ is convex, lower semi-continuous and positively 1-homogenous, then

$$\Phi\Big(\int_\Omega \psi(x)\, \mathrm{d}\Omega(x)\Big) \le \int_\theta \Phi(\psi(x))\, \mathrm{d}\theta(x)$$

for any Borel map $\psi : \mathbb{R}^{\mathrm{d}} \to \mathbb{R}^{m+1}$ and any positive and finite measure θ in $\mathbb{R}^{\mathrm{d}}$ (by rescaling θ to a probability measure and looking at the image measure $\psi_{\#}\theta$ the formula reduces to the standard Jensen inequality).

Without loss of generality, we can assume $\mathbf{w}$ is absolutely continuous. Fix $x \in \Omega$ and apply the inequality above with $\psi(\rho, \mathbf{w}) := (1, \mathbf{w}/\rho)$, $\mathrm{d}\theta = \mathfrak{G}(x, \cdot)\, \mathrm{d}\rho$ and $\Phi(\rho, \mathbf{w}) = \mathbf{K}(\rho, \mathbf{w})$ from the definition of the kinetic energy $\mathcal{K}$. Expanding a fraction in the definition of $\vec{\mathcal{T}}_{\mathfrak{G}}[\mathbf{w}]$ with $\rho(y)$ allows us to apply the Jensen inequality to obtain

$$\begin{aligned}
\frac{\big\|\vec{\mathcal{T}}_{\mathfrak{G}}[\mathbf{w}](x)\big\|^2}{\mathcal{T}_{\mathfrak{G}}[\rho](x)} &= \mathbf{K}\Big(\int_\Omega 1 \cdot \mathfrak{G}(x, y)\, \mathrm{d}\rho(y), \int_\Omega \frac{\mathbf{w}(y)}{\rho(y)} \cdot \mathfrak{G}(x, y)\, \mathrm{d}\rho(y)\Big) \\
&\le \int_\Omega \mathbf{K}\Big(1, \frac{\mathbf{w}(y)}{\rho(y)}\Big) \cdot \mathfrak{G}(x, y)\, \mathrm{d}\rho(y) \\
&= \int_\Omega \frac{\|\mathbf{w}(y)\|^2}{\rho(y)^2} \cdot \mathfrak{G}(x, y)\, \mathrm{d}\rho(y).
\end{aligned}$$

An integration with respect to x leads to

$$\mathcal{K}\big(\mathcal{T}_{\mathfrak{G}}[\rho], \vec{\mathcal{T}}_{\mathfrak{G}}[\mathbf{w}]\big) = \int_\Omega \frac{\Big\|\vec{\mathcal{T}}_{\mathfrak{G}}[\mathbf{w}](x)\Big\|^2}{\mathcal{T}_{\mathfrak{G}}[\rho](x)}\, \mathrm{d}x \le \int_\Omega \int_\Omega \frac{\|\mathbf{w}(y)\|^2}{\rho(y)^2} \cdot \mathfrak{G}(x, y)\, \mathrm{d}\rho(y)\, \mathrm{d}x = \mathcal{K}(\rho, \mathbf{w})$$

where we canceled in the last step $\rho(y)$ and used Fubini's theorem with the fact that $\int_\Omega \mathfrak{G}(x, y)\, \mathrm{d}x = 1$ for each $y \in \Omega$. Hence, this is the desired monotonicity property of the kinetic energy. □

3 Time-Dependent Minimizing Movement Scheme

This part of the thesis is based on the first part of the joint work with Jonathan Zinsl [82]. The aim of this chapter is to study a time-dependent formulation of the Minimizing Movement scheme for non-autonomous evolution problems where the sought-for solution u_t is a curve of steepest descent with respect to the time-dependent free energy functional $\mathcal{E}_t$ in the complete, separable metric space $(\mathbf{X}, \boldsymbol{d})$. As a particular example, we apply this theory also to the non-autonomous Fokker-Planck equation with non-flux boundary condition in a domain $\Omega \subseteq \mathbb{R}^{\mathrm{d}}$ seen as L^2-Wasserstein gradient flow.

Main Idea in Short. In the simplest setting, when $\mathbf{X} = \mathbb{R}^{\mathrm{d}}$, $\boldsymbol{d}$ is the Euclidean metric, and $\mathcal{E}_t \in \mathcal{C}^\infty([0,\infty) \times \mathbb{R}^{\mathrm{d}})$, the main idea to approximate solutions u_t^* to evolution problems with gradient flow structure as

$$\dot{u}_t = -\nabla \mathcal{E}_t(u_t)$$

is to use the *time-dependent version of the implicit Euler scheme.* This method is given as follows. Given a partition $\boldsymbol{\tau} := (\tau_1, \tau_2, \ldots)$ of step sizes $\tau_k \in (0, \tau_*)$ and given an initial condition $u_0^{\boldsymbol{\tau}}$ that approximates u_0, define inductively a discrete solution $(u_k^{\boldsymbol{\tau}})_{k\in\mathbb{N}}$ by the implicit formula

$$\frac{u_k^{\boldsymbol{\tau}} - u_{k\text{-}1}^{\boldsymbol{\tau}}}{\tau_k} = -\nabla \mathcal{E}_{t_k^{\boldsymbol{\tau}}}(u_k^{\boldsymbol{\tau}}), \tag{3.0.1}$$

where $t_k^{\boldsymbol{\tau}} = \sum_{l=1}^{k} \tau_l$ for $k \geq 1$. Under these strong hypothesis on the free energy functional $\mathcal{E}_t$, the time-dependent implicit Euler method (3.0.1) is well-defined, i.e., the initial datum $u_0^{\boldsymbol{\tau}}$ determines an entire sequence $(u_k^{\boldsymbol{\tau}})_{k\in\mathbb{N}}$. It is further well-know that this is a first order approximation of a true solution u_t^*, i.e., one has $u_k^{\boldsymbol{\tau}} = u^*_{t_k^{\boldsymbol{\tau}}} + \mathcal{O}(\tau)$ as $\tau \to 0$.

In the abstract case when $(\mathbf{X}, \boldsymbol{d})$ is just a complete, separable metric space we cast (3.0.1) in a variational way, i.e., we propose the time-dependent version of the so-called Minimizing Movement scheme, cf. [29, 30, 54]. The variational formulation of the time-dependent implicit Euler formula (3.0.1) reads then

$$u_k^{\boldsymbol{\tau}} \in \operatorname*{argmin}_{w\in\mathbf{X}} \frac{1}{2\tau_k} \boldsymbol{d}^2(u_{k\text{-}1}^{\boldsymbol{\tau}}, w) + \mathcal{E}_{t_k^{\boldsymbol{\tau}}}(w). \tag{3.0.2}$$

In the Euclidean setting the minimizer $u_k^{\boldsymbol{\tau}}$ satisfies the implicit Euler formula (3.0.1).

Contribution&Method. Most of our results follow from a careful generalization of the autonomous theory on metric gradient flows by Ambrosio, Gigli and Savaré [4], also in view of the theory by Rossi, Mielke, and Savaré [84] for the non-autonomous case under stricter assumptions. Similar results for non-autonomous gradient flows in abstract metric spaces have been obtained – independently at the same time – by Ferreira and Guevara [39]. Therein the main tool to prove the existence and convergence of the scheme (3.1.2) is the assumption of convexity of the free energy functional $\mathcal{E}_t$. In the end, their approach yields that the approximation converges to a solution u_t^* in the sense of the EVI (2.3.2). In contrast, we follow a different approach where we exploit the compactness of the free energy functional $\mathcal{E}_t$. However, with this slightly weaker assumption, we are only, but canonically, able to prove the convergence of the scheme to a solution u_t^* in the sense of the EDE.

Additionally, we seek to construct a solution ρ_t for the particular example of the L^2-Wasserstein gradient flow, given by the Fokker-Planck equation. However, we investigate this problem under weak assumption on V_t and W_t, such that $\mathcal{E}_t$ does not posses convexity properties along geodesics in the space $(\mathcal{P}_2(\Omega), \mathbf{W}_2)$, cf. [73]. Hence, the results on contractive gradient flows by Ambrosio et al. [4] (in the autonomous case) and Ferreira and Guevara [39] (in the non-autonomous case) are not immediately applicable. Nevertheless, the variational formulation of the time-dependent implicit Euler method is robust enough to prove also in this particular case the existence and convergence of the approximation $\overline{\rho}_t^{\boldsymbol{\tau}}$ to a limit curve ρ_t^*. We want to emphasize, that in this framework we are able to prove additional regularity properties of the discrete solution which imply a stronger notion of convergence. This strong convergence results in combination with the discrete Euler-Lagrange equations yields that the limit curve ρ_t^* is indeed a weak solution of the non-autonomous and non-linear Fokker-Planck equation (3.2.1).

Main Results. Our main result concerning the limit behavior as $\boldsymbol{\tau} \to 0$ of the piecewise constant interpolation $\overline{u}_t^{\boldsymbol{\tau}}$ of the discrete solution $u_k^{\boldsymbol{\tau}}$ reads as follows:

Theorem 3.0.1 (Curves of Steepest Descent for Abstract Metric Gradient Flows). *Assume $(\mathbf{X}, \boldsymbol{d})$ is a complete, separable metric space, $\mathcal{E}_t$ satisfies (E0)–(E5) specified in Assumptions 3.1.1&3.1.2&3.1.3, and given a partition $\boldsymbol{\tau} = (\tau_1, \tau_2, \ldots)$ which satisfies (I1) from Assumption 3.1.6. Then,*

a) ***Existence of Discrete Solutions.*** *For each approximation $u_0^{\boldsymbol{\tau}}$ of the initial datum $u_0 \in \mathcal{D}(\mathcal{E}_0)$ which satisfies (I2) from Assumption 3.1.6 there exists a discrete solution $(u_k^{\boldsymbol{\tau}})_{k\in\mathbb{N}}$ of* (3.1.2).

b) ***Step Size Independent Estimates.*** *For fixed time horizon $T > 0$ there is a constant C, independent of $\boldsymbol{\tau}$, such that the corresponding discrete solution $(u_k^{\boldsymbol{\tau}})_{k\in\mathbb{N}}$ satisfies for all N with $t_N^{\boldsymbol{\tau}} \leq T$:*

$$\sum_{k=1}^{N} \frac{1}{2\tau_k} \boldsymbol{d}^2(u_{k-1}^{\boldsymbol{\tau}}, u_k^{\boldsymbol{\tau}}) \leq C, \qquad \mathcal{E}_{t_N^{\boldsymbol{\tau}}}(u_N^{\boldsymbol{\tau}}) \leq C, \qquad \boldsymbol{d}^2(u_*, u_N^{\boldsymbol{\tau}}) \leq C.$$

Furthermore, given a sequence of partitions $(\boldsymbol{\tau}_n)_{n\in\mathbb{N}}$ with $\sup_k \tau_{k,n} \to 0$ and a sequence of approximations $(u_0^{\boldsymbol{\tau}_n})_{n\in\mathbb{N}}$ of the initial datum u_0 which satisfy (I1)&(I2). Then,

c) **Convergence in the $\boldsymbol{\sigma}$-topology.** There exists $u_t^* \in AC^2(0,\infty;(\mathbf{X},\boldsymbol{d}))$ such that for a (non-relabelled) subsequence of $(\boldsymbol{\tau}_n)_{n\in\mathbb{N}}$

$$\overline{u}_t^{\boldsymbol{\tau}_n} \overset{\boldsymbol{\sigma}}{\rightharpoonup} u_t^* \qquad \forall\, t \in [0,\infty).$$

d) **Solution of the Non-autonomous Gradient Flow.** The limit curve u_t^* from c) is a solution to the time dependent gradient flow (3.1.1) in the sense of the energy dissipation equality, see definition 2.3.3.

Our main results concerning the well-posedness and the limit behavior as $\boldsymbol{\tau} \searrow 0$ of the *interpolated solution* $\overline{\rho}_t^{\boldsymbol{\tau}}$ is stated in the following theorem.

Theorem 3.0.2 (Existence of Solutions for the Non-linear Fokker-Planck Equation)**.** *Let $\Omega \subset \mathbb{R}^{\mathrm{d}}$ be either an open, bounded, and connected domain with Lipschitz continuous boundary $\partial\Omega$ or let $\Omega = \mathbb{R}^{\mathrm{d}}$. Further, assume $m \geq 1$ and that V_t and W_t satisfy (F1)–(F3) defined in Assumption 3.2.1 and let a partition $\boldsymbol{\tau}$ be given that satisfy (I1) specified in Assumption 3.1.6. Then, we have additionally to the results of theorem 3.0.1:*

a) ***Step Size Independent $\mathrm{L}^2(0,T;\mathrm{BV}(\Omega))$-estimate.*** *For each $T > 0$ there exists a non-negative constant C, independent of $\boldsymbol{\tau}$ such that for each $\boldsymbol{\tau} \in (0,\tau_*)$:*

$$\|(\overline{\rho}_t^{\boldsymbol{\tau}})^m\|_{\mathrm{L}^2(0,T;\mathrm{BV}(\Omega))} \leq C.$$

Furthermore, given a sequence of partitions $(\boldsymbol{\tau}_n)_{n\in\mathbb{N}}$ with $\sup_k \tau_{k,n} \to 0$ and an approximation $(\rho_0^{\boldsymbol{\tau}_n})_{n\in\mathbb{N}}$ of the initial datum ρ_0 which satisfies (I1)&(I2) defined in Assumption 3.1.6.

b) ***Strong Convergence in $\mathrm{L}^p(0,T;\mathrm{L}^m(\Omega))$.*** *There exists a further (non-relabelled) subsequence $(\tau_n)_{n\in\mathbb{N}}$ such that for all $T > 0$, any $p \in [1,\infty)$,*

1) if Ω bounded:

$$\overline{\rho}_t^{\boldsymbol{\tau}_n} \to \rho_t^* \qquad \textit{strongly in } \mathrm{L}^p(0,T;\mathrm{L}^m(\Omega)) \textit{ as } n \to \infty.$$

2) if $\Omega = \mathbb{R}^{\mathrm{d}}$: for any bounded set $\Theta \subset \mathbb{R}^{\mathrm{d}}$:

$$\overline{\rho}_t^{\boldsymbol{\tau}_n} \to \rho_t^* \qquad \textit{strongly in } \mathrm{L}^p(0,T;\mathrm{L}^m(\Theta)) \textit{ as } n \to \infty.$$

c) ***Solution of the Non-linear Fokker-Planck Equation.*** *The limit curve ρ_t^* from b) satisfies the non-linear Fokker-Planck equation with no-flux boundary condition* (3.2.1) *in the weak sense of* (3.2.11).

3.1 Application to Gradient Flows in Abstract Metric Spaces

This section is devoted to study the temporal discretization of *non-autonomous evolution problems* of the form

$$\dot{u}_t = -\nabla_{\mathbf{X}} \mathcal{E}_t(u_t), \tag{3.1.1}$$

where the sought-for solution $u_t : [0, \infty) \to \mathbf{X}$ is a curve of steepest descent with respect to the *time-dependent free energy functional* $\mathcal{E}_t$ emanating from $u_0 \in \mathcal{D}(\mathcal{E}_0)$ in a complete, separable metric space $(\mathbf{X}, \boldsymbol{d})$. More precisely, we seek to construct with the variational formulation of the time-dependent Implicit Euler method, given by

$$\frac{u_k^{\boldsymbol{\tau}} - u_{k\text{-}1}^{\boldsymbol{\tau}}}{\tau_k} = -\nabla \mathcal{E}_{t_k^{\boldsymbol{\tau}}}(u_k^{\boldsymbol{\tau}}),$$

an approximation $\overline{u}_t^{\boldsymbol{\tau}}$ which converges in the discrete-to-continuous limit to a solution u_t^* to the non-autonomous gradient flow equation (3.1.1) in the sense of the EDE, see definition 2.3.3.

Method. We adapt the variational formulation of the time-independent Implicit Euler method, known as the Minimizing Movement scheme, to the time-dependent case. Our scheme to approximate the true solution u_t^* reads than as follows:

Scheme. For a partition $\boldsymbol{\tau} := (\tau_1, \tau_2, \ldots)$ of step sizes $\tau_k \in (0, \tau_*)$ let an initial condition $u_0^{\boldsymbol{\tau}}$ be given that approximates u_0. Then define inductively a discrete solution $(u_k^{\boldsymbol{\tau}})_{k\in\mathbb{N}}$ such that each $u_k^{\boldsymbol{\tau}}$ with $k = 1, 2, \ldots$ is a minimizer of the Moreau-Yosida-penalized energy functional

$$w \mapsto \Phi(\tau, t_k^{\boldsymbol{\tau}}, u_{k\text{-}1}^{\boldsymbol{\tau}}; w) := \frac{1}{2\tau_k} \boldsymbol{d}^2(u_{k\text{-}1}^{\boldsymbol{\tau}}, w) + \mathcal{E}_{t_k^{\boldsymbol{\tau}}}(w), \tag{3.1.2}$$

where $t_k^{\boldsymbol{\tau}} = \sum_{l=1}^{k} \tau_l$ for $k \geq 1$.

Define the corresponding piecewise constant interpolation in time $\overline{u}_t^{\boldsymbol{\tau}} : [0, \infty) \to \mathbf{X}$ of the discrete solution via

$$\overline{u}_0^{\boldsymbol{\tau}} = u_0^{\boldsymbol{\tau}}, \qquad \overline{u}_t^{\boldsymbol{\tau}} = u_k^{\boldsymbol{\tau}} \quad \text{for } t \in (t_{k\text{-}1}^{\boldsymbol{\tau}}, t_k^{\boldsymbol{\tau}}] \text{ and } k \in \mathbb{N}.$$

Strategy of the Proof. We begin with deriving some basic properties of the penalization in section 3.1.2, namely: lower bounds, a priori estimates, the existence of minimizer, continuity and approximation properties, and differentiability properties. Next, in section 3.1.3 we deduce from the variational formulation of the scheme the necessary structural properties, like estimates on the kinetic energy and on the internal energy and the derivation of the discrete Euler-Lagrange equations. Finally, we prove in section 3.1.4 the main theorem 3.0.1, i.e., the convergence as $\boldsymbol{\tau}$ tends to zero of the approximation $\overline{u}_t^{\boldsymbol{\tau}}$ to a curve of steepest descent u_t^* by means of an Arzelá-Ascoli type argument.

3.1.1 Setup and Assumptions

Given a separable, complete metric space $(\mathbf{X}, \boldsymbol{d})$, we shall introduce a weaker Hausdorff topology $\boldsymbol{\sigma}$ on $\boldsymbol{X}$, which is compatible with $\boldsymbol{d}$, which allows us more flexibility to derive compactness results. From now on we propose the convention to write

$$u_n \xrightarrow{\boldsymbol{d}} u \quad \text{for the convergence w.r.t. } \boldsymbol{d}, \qquad u_n \overset{\boldsymbol{\sigma}}{\rightharpoonup} u \quad \text{for the convergence w.r.t. } \boldsymbol{\sigma}.$$

Compatibility of $\boldsymbol{\sigma}$ with $\boldsymbol{d}$ means in this context

$$u_n \xrightarrow{\boldsymbol{d}} u \implies u_n \overset{\boldsymbol{\sigma}}{\rightharpoonup} u, \qquad (u_n, v_n) \overset{\boldsymbol{\sigma}}{\rightharpoonup} (u, v) \implies \boldsymbol{d}(u,v) \leq \liminf_{n\to\infty} \boldsymbol{d}(u_n, v_n).$$

Additionally, we shall work throughout the rest of this chapter with the following assumptions to the functional $\mathcal{E}_t$.

Assumption 3.1.1 (Chain Rule Inequality)**.** The free energy functional $\mathcal{E}_t : [0,\infty) \times \mathbf{X} \to \mathbb{R} \cup \{\infty\}$ satisfies the following chain rule condition.

(E0) The local slope $|\partial \mathcal{E}_t|$ of $\mathcal{E}_t$ at time t is lower semi-$\boldsymbol{\sigma}$-continuous and for any curve $u_t \in \mathrm{AC}^2\left(0,\infty; (\mathbf{X}, \boldsymbol{d})\right)$ with $|\partial \mathcal{E}_t|(u_t)|u_t'| \in \mathrm{L}^1_{\mathrm{loc}}\left(0,\infty\right)$ and $\sup_{t\in[0,T]} \mathcal{E}_t(u_t) < \infty$, the map $t \mapsto \mathcal{E}_t(u_t)$ is absolutely continuous, and for all $0 \leq s \leq t$:

$$\mathcal{E}_s(u_s) + \int_s^t \partial_t \mathcal{E}_r(u_r)\,\mathrm{d}r \leq \mathcal{E}_t(u_t) + \int_s^t |\partial \mathcal{E}_r|(u_r)|u_r'|\,\mathrm{d}r. \tag{3.1.3}$$

Assumption 3.1.2 (Space-Regularity of $\mathcal{E}_t$)**.** The free-energy functional $\mathcal{E}_t : [0,\infty) \times \mathbf{X} \to \mathbb{R} \cup \{\infty\}$ is proper and satisfies the following regularity conditions in space:

(E1) **Lower Semi-continuity.** $\mathcal{E}_t$ is sequentially lower semi-$\boldsymbol{\sigma}$-continuous on $\boldsymbol{d}$-bounded sets for each $t \in [0,\infty)$:

$$\sup_{n,m} \boldsymbol{d}(u_n, u_m) < \infty, \quad u_n \overset{\boldsymbol{\sigma}}{\rightharpoonup} u \qquad \implies \qquad \mathcal{E}_t(u) \leq \liminf_{n\to\infty} \mathcal{E}_t(u_n).$$

(E2) **Coercivity.** There exist $\tau_* > 0$ and $u_* \in \mathbf{X}$ such that:

$$c_* := \inf_{t\in[0,\infty)} \inf_{w\in\mathbf{X}} \frac{1}{2\tau_*} \boldsymbol{d}^2(u_*, w) + \mathcal{E}_t(w) > -\infty.$$

(E3) **Compactness.** For each $t > 0$, every $\boldsymbol{d}$-bounded set contained in a sublevel of $\mathcal{E}_t$ is relatively sequentially $\boldsymbol{\sigma}$-compact, i.e.:

$$\text{if } (u_n)_{n\in\mathbb{N}} \subset \mathbf{X} \text{ with } \sup_n \mathcal{E}_t(u_n) < \infty, \text{ and } \sup_{n,m} \boldsymbol{d}(u_n, u_m) < \infty, \text{ then}$$
$$(u_n)_{n\in\mathbb{N}} \text{ contains a } \boldsymbol{\sigma}\text{-convergent subsequence.}$$

Assumption 3.1.3 (Time-Regularity of $\mathcal{E}_t$)**.** The free-energy functional $\mathcal{E}_t : [0,\infty) \times \mathbf{X} \to \mathbb{R} \cup \{\infty\}$ satisfies the following regularity conditions in time:

(E4) **Absolute Continuity in Time.** There exists a non-negative function $\alpha_r \in \mathrm{L}^1_{\mathrm{loc}}(0,\infty)$ such that for all $u \in \mathbf{X}$ and for all $0 \leq s \leq t$, it holds that:

$$|\mathcal{E}_t(u) - \mathcal{E}_s(u)| \leq (1 + \boldsymbol{d}^2(u^*, u)) \int_s^t \alpha_r \, \mathrm{d}r.$$

Moreover, α_t has at most countable many points which are not Lebesgue points.

(E5) **Differentiability in Time.** For all $u \in \mathbf{X}$, the partial derivative $\partial_t \mathcal{E}_t(u)$ of the map $t \mapsto \mathcal{E}_t(u)$ exists and is $\boldsymbol{\sigma}$-continuous on $\boldsymbol{d}$-bounded sets:

$$\sup_{n,m} \boldsymbol{d}(u_n, u_m) < \infty, \quad u_n \overset{\boldsymbol{\sigma}}{\rightharpoonup} u, \quad t_n \to t \qquad \Longrightarrow \qquad \lim_{n\to\infty} \partial_t \mathcal{E}_{t_n}(u_n) = \partial_t \mathcal{E}_t(u).$$

Remark 3.1.4. The Chain Rule Inequality (E0) is the reversed version of the time-dependent EDE (1.0.3) in the definition of the energy dissipation equality. This assumption is not restrictive since one has

$$\frac{\mathrm{d}}{\mathrm{d}t} \mathcal{E}_t(u_t) = \partial_t \mathcal{E}_t(u_t) + \langle \nabla \mathcal{E}(u_t), u_t \rangle \geq \partial_t \mathcal{E}_t(u_t) - \|\mathcal{E}(u_t)\| \, \|\dot{u}_t\|$$

by the chain rule and the Cauchy-Schwarz inequality. Integrating this inequality with respect to t yields exactly the Chain Rule Inequality (3.1.3).

Remark 3.1.5. (E1)–(E3) are the time-dependent versions of the known *lower semi-continuity, compactness, and coercivity conditions (LSCC)* which are standard in the gradient flow theory in abstract metric spaces. (E4) and (E5) are then the necessary additional time-dependent assumption on $\mathcal{E}_t$.

Later, we have to specify further assumptions on the partition $\boldsymbol{\tau}$ and the approximation $u_0^{\boldsymbol{\tau}}$ of the initial datum u_0. For sake of completeness, these are given now.

Assumption 3.1.6. The partition $\boldsymbol{\tau} = (\tau_1, \tau_2, \ldots)$ and the approximation $u_0^{\boldsymbol{\tau}}$ of the initial datum u_0 satisfies:

(I1) Given α_t from (E4), then

$$\sup_k 4\alpha_k^{\boldsymbol{\tau}} < 1 \qquad \text{with} \qquad \alpha_k^{\boldsymbol{\tau}} := \Big(\frac{\tau_*}{2} \int_{t_k^{\boldsymbol{\tau}}}^{t_{k+1}^{\boldsymbol{\tau}}} \alpha_r \, \mathrm{d}r + \frac{\tau_k}{\tau_*} \Big). \tag{3.1.4}$$

(I2) There exists a a constant d_1 such that the discrete initial datum $u_0^{\boldsymbol{\tau}}$ satisfy:

$$\mathcal{E}_0(u_0^{\boldsymbol{\tau}}) \leq d_1, \quad \boldsymbol{d}(u_0^{\boldsymbol{\tau}}, u_0) \leq d_1 \sqrt{\tau}, \quad \text{and} \quad \limsup_{\tau \to 0} \mathcal{E}_0(u_0^{\boldsymbol{\tau}}) \leq \mathcal{E}_0(u_0).$$

3.1.2 Time-Dependent Moreau-Yosida Approximation and Resolvent

Assume in the sequel that (E1)–(E5) holds. Define the *Moreau-Yosida functional*

$$\Phi(\tau, t, u; \cdot) : \mathbf{X} \to \mathbb{R} \cup \{\infty\}; \;\Phi(\tau, t, u; w) := \frac{1}{2\tau}\boldsymbol{d}^2(u, w) + \mathcal{E}_t(w) \tag{3.1.5}$$

and furthermore define the *Moreau-Yosida approximation* of $\mathcal{E}_t$ by

$$\phi(\tau, t, u) := \inf_{w\in\mathbf{X}} \Phi(\tau, t, u; w) = \inf_{w\in\mathbf{X}} \frac{1}{2\tau}\boldsymbol{d}^2(u, w) + \mathcal{E}_t(w). \tag{3.1.6}$$

The well-posedness of the *Minimizing Movement scheme* (3.1.2) is equivalent to the existence of a minimizer of Φ. The set of all minimizers is called the *resolvent* J_t^τ and is given by

$$J_t^\tau(u) = \{w \in \mathbf{X} \mid \Phi(\tau, t, u; w) = \phi(\tau, t, u)\}. \tag{3.1.7}$$

Remark 3.1.7. By the construction of the Moreau-Yosida approximation, we have the following monotonicity

$$\phi(\sigma, t, u) \le \phi(\tau, t, u) \le \mathcal{E}_t(u) \qquad \text{for } \sigma \ge \tau > 0. \tag{3.1.8}$$

Before we prove the existence of minimizers of the Moreau-Yosida approximation, we state an auxiliary inequality which will be used several times in the rest of this section.

Lemma 3.1.8 (Bounds)**.** *Let c_*, τ_*, u_* be the constants from (E1)–(E3). Then, for all $\tau \in (0, \tau_*)$, $t \in [0, \infty)$ and all $u, w \in \mathbf{X}$, we have that:*

$$\phi(\tau, t, u) \ge c_* - \frac{1}{\tau_* - \tau}\boldsymbol{d}^2(u_*, u), \tag{3.1.9}$$

$$\boldsymbol{d}^2(w, u) \le \frac{4\tau\tau_*}{\tau_* - \tau}\Big(\Phi(\tau, t, u; w) - c_* + \frac{1}{\tau_* - \tau}\boldsymbol{d}^2(u_*, u)\Big). \tag{3.1.10}$$

Proof. We use the Cauchy-type inequality

$$(a + b)^2 \le (1 + \varepsilon)\, a^2 + \Big(1 + \frac{1}{\varepsilon}\Big)b^2 \qquad \forall\; a, b \ge 0,\; \varepsilon > 0,$$

with $a = d(w, u),\; b = d(u_*, u)$ and $\varepsilon = \frac{\tau_*-\tau}{\tau_*+\tau}$, to get:

$$\frac{1}{2\tau_*}\boldsymbol{d}^2(u_*, w) \le \frac{1}{\tau_* + \tau}\boldsymbol{d}^2(w, u) + \frac{1}{\tau_* - \tau}\boldsymbol{d}^2(u_*, u).$$

This yields for every $u, w \in \mathbf{X}$ and $\tau < \tau_*$:

$$\begin{aligned}\Phi(\tau, t, u; w) &= \frac{\tau_* - \tau}{2\tau(\tau_* + \tau)}\boldsymbol{d}^2(u, w) + \frac{1}{\tau_* + \tau}\boldsymbol{d}^2(u, w) + \mathcal{E}_t(w)\\ &\ge \frac{\tau_* - \tau}{4\tau_*\tau}\boldsymbol{d}^2(u, w) + \frac{1}{2\tau_*}\boldsymbol{d}^2(u_*, w) - \frac{1}{\tau_* - \tau}\boldsymbol{d}^2(u_*, u) + \mathcal{E}_t(w)\\ &\ge \frac{\tau_* - \tau}{4\tau_*\tau}\boldsymbol{d}^2(u, w) - \frac{1}{\tau_* - \tau}\boldsymbol{d}^2(u_*, u) + c_*,\end{aligned}$$

from which (3.1.10) and, after taking the infimum with respect to $w \in \mathbf{X}$, (3.1.9) follows. □

In the following we show that indeed the *time-dependent Minimizing Movement scheme* (3.1.2) is well-defined in the abstract metric setting for sufficiently small τ.

Theorem 3.1.9 (Existence of Minimizers). *For all $\tau \in (0, \tau_*)$, for all $t \in [0, \infty)$ and for all $u \in \mathbf{X}$, there exists a minimizer $w_* \in \mathcal{D}(\mathcal{E}_t)$ of the map $w \mapsto \Phi(\tau, t, u; w)$, i.e.,*

$$J_t^\tau(u) \neq \emptyset.$$

Proof. Fix $\tau \in (0, \tau_*)\,, t \in [0, \infty)\,, u \in \mathbf{X}$ and note that by Lemma 3.1.8 the Moreau-Yosida functional is bounded from below for each $u \in \mathbf{X}$. Since $\mathcal{E}_t$ is proper, the infimum is not equal to infinity. So choose a minimizing sequence $(w_n)_{n\in\mathbb{N}}$ in $\mathbf{X}$ of the map $w \mapsto \Phi(\tau, t, u; w)$ and without loss of generality $\sup_n \Phi(\tau, t, u; w_n) < \infty$. So, we can deduce from (3.1.10)

$$\boldsymbol{d}^2(w_n, u) \leq \frac{4\tau\tau_*}{\tau_* - \tau}\Big(\Phi(\tau, t, u; w_n) - c_* + \frac{1}{\tau_* - \tau}\boldsymbol{d}^2(u_*, u)\Big) < \infty.$$

Thus the sequence $(w_n)_{n\in\mathbb{N}}$ is $\boldsymbol{d}$-bounded. Furthermore, the $\boldsymbol{\sigma}$-compactness of the sequence w_n follows by the upper estimate on $\mathcal{E}_t$

$$\mathcal{E}_t(w_n) \leq \frac{1}{2\tau}\boldsymbol{d}^2(u, w_n) + \mathcal{E}_t(w_n) = \Phi(\tau, t, u; w_n) \leq c < \infty.$$

Hence, we can extract a $\boldsymbol{\sigma}$-convergent subsequence, which converges to some $w_* \in \mathcal{D}(\mathcal{E}_t)$ with respect to the weak topology $\boldsymbol{\sigma}$. By lower semi-$\boldsymbol{\sigma}$-continuity of $\mathcal{E}_t$, we conclude that indeed w_* is a minimizer of the map $w \mapsto \Phi(\tau, t, u; w)$ and thus $J_t^\tau(u) \neq \emptyset$. □

Further facts about the time-dependent Moreau-Yosida approximation and of the resolvent [4], for instance, a priori estimates, continuity results and differentiability properties are proven in the sequel.

Lemma 3.1.10 (Apriori Estimate). *Let $u \in \mathbf{X}$ and define $u_s^\sigma \in J_s^\sigma(u)$ and $u_t^\tau \in J_t^\tau(u)$ with $\sigma < \tau$ and $s < t$, then it holds that*

$$\boldsymbol{d}^2(u, u_s^\sigma) \leq \boldsymbol{d}^2(u, u_t^\tau) + \frac{2\tau\sigma}{\tau - \sigma}\left(2 + \boldsymbol{d}^2(u_*, u_s^\sigma) + \boldsymbol{d}^2(u_*, u_t^\tau)\right)\int_s^t \alpha_r \,\mathrm{d}r.$$

Proof. Fix $u_s^\sigma \in J_s^\sigma(u)$ and $u_t^\tau \in J_t^\tau(u)$ with $\sigma < \tau, s < t$ and exploit once again the variational definition of the resolvent and of the Moreau-Yosida approximation to get

$$\begin{aligned}
\Phi(\sigma, s, u; u_s^\sigma) &\leq \Phi(\sigma, s, u; u_t^\tau) \\
&= \Big(\frac{1}{2\sigma} - \frac{1}{2\tau}\Big)\boldsymbol{d}^2(u, u_t^\tau) + \Phi(\tau, s, u; u_t^\tau) \\
&= \Big(\frac{1}{2\sigma} - \frac{1}{2\tau}\Big)\boldsymbol{d}^2(u, u_t^\tau) + \Phi(\tau, t, u; u_t^\tau) + \mathcal{E}_s(u_t^\tau) - \mathcal{E}_t(u_t^\tau) \\
&\leq \Big(\frac{1}{2\sigma} - \frac{1}{2\tau}\Big)\boldsymbol{d}^2(u, u_t^\tau) + \Phi(\tau, t, u; u_s^\sigma) + \mathcal{E}_s(u_t^\tau) - \mathcal{E}_t(u_t^\tau).
\end{aligned}$$

Subtract $\Phi(\tau, t, u; u_s^\sigma)$ from both sides to obtain

$$\Big(\frac{1}{2\sigma} - \frac{1}{2\tau}\Big)\boldsymbol{d}^2(u, u_s^\sigma) + \mathcal{E}_s(u_s^\sigma) - \mathcal{E}_t(u_s^\sigma) \le \Big(\frac{1}{2\sigma} - \frac{1}{2\tau}\Big)\boldsymbol{d}^2(u, u_t^\tau) + \mathcal{E}_s(u_t^\tau) - \mathcal{E}_t(u_t^\tau).$$

Since $\sigma < \tau$, we can multiply with $\frac{2\tau\sigma}{\tau-\sigma}$ to get

$$\begin{aligned}\boldsymbol{d}^2(u, u_s^\sigma) &\le \boldsymbol{d}^2(u, u_t^\tau) + \frac{2\tau\sigma}{\tau - \sigma}\left(\mathcal{E}_s(u_t^\tau) - \mathcal{E}_t(u_t^\tau) + \mathcal{E}_t(u_s^\sigma) - \mathcal{E}_s(u_s^\sigma)\right) \\ &\le \boldsymbol{d}^2(u, u_t^\tau) + \frac{2\tau\sigma}{\tau - \sigma}\left(2 + \boldsymbol{d}^2(u_*, u_s^\sigma) + \boldsymbol{d}^2(u_*, u_t^\tau)\right)\int_s^t \alpha_r \,\mathrm{d}r,\end{aligned}$$

where in the last step, we used (E4) twice. □

Lemma 3.1.11 (Continuity of the Resolvent)**.** *Let $u \in \mathcal{D}(\mathcal{E}_t)$ and $\tau \in (0, \tau_*)$. Given the convergent sequences $\tau_n \searrow 0$, $t_n \to t$ and $u_n \xrightarrow{\boldsymbol{d}} u$, define a sequence of minimizers $w_n \in J_{t_n}^{\tau_n}(u_n)$. If in addition $\mathcal{E}_{t_n}(u_n) \le C$, then we have*

$$w_n \xrightarrow{\boldsymbol{d}} u \qquad \text{as} \quad n \to \infty.$$

Proof. We can assume without loss of generality that $\tau_n < \tau_*$. Use the monotonicity of the Moreau-Yosida approximation (3.1.8) and the estimate (3.1.10) with $w = w_n$ and $u = u_n$, to obtain:

$$\begin{aligned}\boldsymbol{d}^2(w_n, u_n) &\le \frac{4\tau_n\tau_*}{\tau_* - \tau_n}\Big(\Phi(\tau_n, t_n, u_n; w_n) - c_* + \frac{1}{\tau_* - \tau_n}\boldsymbol{d}^2(u_*, u_n)\Big) \\ &\le \frac{4\tau_n\tau_*}{\tau_* - \tau_n}\Big(\mathcal{E}_{t_n}(u_n) - c_* + \frac{1}{\tau_* - \tau_n}\boldsymbol{d}^2(u_*, u_n)\Big).\end{aligned}$$

By assumption $\mathcal{E}_{t_n}(u_n)$ is bounded from above, so we can further estimate to obtain

$$\boldsymbol{d}^2(w_n, u_n) \le \frac{4\tau_n\tau_*}{\tau_* - \tau_n}\Big(C - c_* + \frac{1}{\tau_* - \tau_n}\boldsymbol{d}^2(u_*, u_n)\Big).$$

Taking the limit $n \to \infty$ yields the desired convergence $w_n \xrightarrow{\boldsymbol{d}} u$. □

Lemma 3.1.12 (Continuity of the Moreau-Yosida Approximation)**.** *The map $(\tau, t, u) \mapsto \phi(\tau, t, u)$ is $\boldsymbol{d}$-continuous on $(0, \tau_*) \times [0, \infty) \times \mathcal{D}(\mathcal{E}_t)$.*

Proof. Choose a sequence $(\tau_n, t_n, u_n)_{n\in\mathbb{N}}$ in $(0, \tau_*) \times [0, \infty) \times \mathcal{D}(\mathcal{E}_{t_n})$ with $\tau_n \to \tau \in (0, \tau_*)$, $t_n \to t \in [0, \infty)$ and $u_n \xrightarrow{\boldsymbol{d}} u \in \mathcal{D}(\mathcal{E}_t)$. Then, it follows for an arbitrary $w \in \mathbf{X}$ that

$$\limsup_{n\to\infty} \phi(\tau_n, t_n, u_n) \le \limsup_{n\to\infty} \frac{1}{2\tau_n}\boldsymbol{d}^2(u_n, w) + \mathcal{E}_{t_n}(w) = \frac{1}{2\tau}\boldsymbol{d}^2(u, w) + \mathcal{E}_t(w)$$

thanks to (E4). Taking the infimum over w on the r.h.s. yields the upper semi-$\boldsymbol{d}$-continuity of ϕ.

To prove the lower semi-continuity, choose $w_n \in J_{t_n}^{\tau_n}(u_n)$ and first of all notice that this sequence is bounded, since by the the upper estimate for the Moreau-Yosida approximation (3.1.10)

$$\begin{aligned}\boldsymbol{d}^2(w_n, u_n) &\leq \frac{4\tau_n\tau_*}{\tau_* - \tau_n}\Big(\Phi(\tau_n, t_n, u_n; w_n) - c_* + \frac{1}{\tau_* - \tau_n}\boldsymbol{d}^2(u_*, u_n)\Big) \\ &= \frac{4\tau_n\tau_*}{\tau_* - \tau_n}\Big(\phi(\tau_n, t_n, u_n) - c_* + \frac{1}{\tau_* - \tau_n}\boldsymbol{d}^2(u_*, u_n)\Big).\end{aligned}$$

Since every term on the r.h.s. is bounded the sequence $(w_n)_{n\in\mathbb{N}}$ is $\boldsymbol{d}$-bounded and by the continuity of the resolvent (Lemma 3.1.11) we also have that w_n converges to u in $\boldsymbol{d}$. Now, the variational definition of w_n yields

$$\begin{aligned}\liminf_{n\to\infty} \phi(\tau_n, t_n, u_n) &= \liminf_{n\to\infty} \Big(\frac{1}{2\tau_n}\boldsymbol{d}^2(u_n, w_n) + \mathcal{E}_{t_n}(w_n)\Big) \\ &\geq \liminf_{n\to\infty} \Big(\frac{1}{2\tau_n}\left(\boldsymbol{d}^2(u_n, u) + \boldsymbol{d}^2(u, w_n) - 2d(u_n, u)d(u, w_n)\right) + \mathcal{E}_{t_n}(w_n)\Big) \\ &= \liminf_{n\to\infty} \Big(\frac{1}{2\tau}\boldsymbol{d}^2(u, w_n) + \mathcal{E}_t(w_n) - \mathcal{E}_t(w_n) + \mathcal{E}_{t_n}(w_n)\Big).\end{aligned}$$

Lastly, the absolute continuity of $\mathcal{E}_t$ and the $\boldsymbol{d}$-boundedness of w_n yields

$$\begin{aligned}\liminf_{n\to\infty} \phi(\tau_n, t_n, u_n) &\geq \liminf_{n\to\infty} \Big(\frac{1}{2\tau}\boldsymbol{d}^2(u, w_n) + \mathcal{E}_t(w_n) - \mathcal{E}_t(w_n) + \mathcal{E}_{t_n}(w_n)\Big) \\ &\geq \liminf_{n\to\infty} \phi(\tau, t, u) - \limsup_{n\to\infty}(1 + \boldsymbol{d}^2(u^*, w_n)) \int_t^{t_n} \alpha_r \,\mathrm{d}r \\ &= \phi(\tau, t, u).\end{aligned}$$

Hence, lower semi-continuity and upper semi-continuity of the Moreau-Yosida approximation ϕ yields the desired result. □

Lemma 3.1.13 (Approximation Property of the Moreau-Yosida Approximation)**.** *For all $u \in \mathcal{D}(\mathcal{E}_t)$ and all sequences $(\tau_n, t_n) \subset (0, \tau_*) \times [0, \infty)$ with $\tau_n \searrow 0$ and $t_n \to t$, we have that*

$$\lim_{n\to\infty} \phi(\tau_n, t_n, u) = \mathcal{E}_t(u).$$

Proof. Given $u \in \mathcal{D}(\mathcal{E}_t)$, choose an element $w_n \in J_{t_n}^{\tau_n}(u)$. Thus, by the continuity of the resolvent (Lemma 3.1.11), $w_n \overset{\boldsymbol{d}}{\to} u$ as $n \to \infty$. The lower semi-$\boldsymbol{\sigma}$-continuity of $\mathcal{E}_t$ for each fixed t and (E4) yields a lower bound for the limit, i.e.:

$$\liminf_{n\to\infty} \phi(\tau_n, t_n, u) = \liminf_{n\to\infty} \frac{1}{2\tau_n}\boldsymbol{d}^2(w_n, u) + \mathcal{E}_{t_n}(w_n) \geq \liminf_{n\to\infty} \mathcal{E}_{t_n}(w_n) \geq \mathcal{E}_t(u).$$

The reverse inequality follows from the monotonicity of the Moreau-Yosida approximation (3.1.8), and with (E4):

$$\limsup_{n\to\infty} \phi(\tau_n, t_n, u) \leq \limsup_{n\to\infty} \mathcal{E}_{t_n}(u) = \mathcal{E}_t(u). \qquad \square$$

Lemma 3.1.14 (Joint Differentiability of the Moreau-Yosida Approximation)**.** *For every $u \in \mathcal{D}(\mathcal{E}_t)$ and $t \in [0,\infty)$, the map $\tau \to \phi(\tau, t+\tau, u)$ is locally Lipschitz continuous on $(0,\tau_*)$ and differentiable except on a countable set $\mathcal{S}_t^u$. For every $\tau \in (0,\tau_*)\setminus\mathcal{S}_t^u$ we have:*

$$\frac{d}{d\tau}\phi(\tau, t+\tau, u) = -\frac{1}{2\tau^2}\boldsymbol{d}^2(u,w) + \partial_t\mathcal{E}_{t+\tau}(w) \qquad \forall\ w \in J_{t+\tau}^{\tau}(u). \tag{3.1.11}$$

Proof. Fix $t > 0$ and $\sigma < \tau$ in $(0,\tau_*)$ and choose $u^\tau \in J_{t+\tau}^\tau(u)$ and $u^\sigma \in J_{t+\sigma}^\sigma(u)$ and exploit the variational definition of the Moreau-Yosida approximation to obtain

$$\begin{aligned}
\phi(\tau, t+\tau, u) - \phi(\sigma, t+\sigma, u) &\leq \Phi(\tau, t+\tau, u; u^\sigma) - \Phi(\sigma, t+\sigma, u; u^\sigma) \\
&= \Big(\frac{1}{2\tau} - \frac{1}{2\sigma}\Big)\boldsymbol{d}^2(u,u^\sigma) + \mathcal{E}_{t+\tau}(u^\sigma) - \mathcal{E}_{t+\sigma}(u^\sigma) \\
&\leq \frac{\sigma-\tau}{2\tau\sigma}\boldsymbol{d}^2(u,u^\sigma) + (1+\boldsymbol{d}^2(u_*,u^\sigma))\int_{t+\sigma}^{t+\tau}\alpha_r\,\mathrm{d}r.
\end{aligned}$$

Analogously, a lower bound can be established by reversing the role of u^τ and u^σ:

$$\begin{aligned}
\phi(\tau, t+\tau, u) - \phi(\sigma, t+\sigma, u) &\geq \Phi(\tau, t+\tau, u, u^\tau) - \Phi(\sigma, t+\sigma, u, u^\tau) \\
&= \Big(\frac{1}{2\tau} - \frac{1}{2\sigma}\Big)\boldsymbol{d}^2(u,u^\tau) + \mathcal{E}_{t+\tau}(u^\tau) - \mathcal{E}_{t+\sigma}(u^\tau) \\
&\geq \frac{\sigma-\tau}{2\tau\sigma}\boldsymbol{d}^2(u,u^\tau) - (1+\boldsymbol{d}^2(u_*,u^\tau))\int_{t+\sigma}^{t+\tau}\alpha_r\,\mathrm{d}r.
\end{aligned}$$

Note that by estimate (3.1.10) and by the monotonicity of the Moreau-Yosida approximation we have

$$\begin{aligned}
\boldsymbol{d}^2(u^\tau,u) &\leq \frac{4\tau\tau_*}{\tau_*-\tau}\Big(\Phi(\tau, t+\tau, u, u^\tau) - c_* + \frac{1}{\tau_*-\tau}\boldsymbol{d}^2(u_*,u)\Big) \\
&= \frac{4\tau\tau_*}{\tau_*-\tau}\Big(\phi(\tau, t+\tau, u) - c_* + \frac{1}{\tau_*-\tau}\boldsymbol{d}^2(u_*,u)\Big) \\
&\leq \frac{4\tau\tau_*}{\tau_*-\tau}\Big(\mathcal{E}_{t+\tau}(u) - c_* + \frac{1}{\tau_*-\tau}\boldsymbol{d}^2(u_*,u)\Big) \\
&\leq \frac{4\tau\tau_*}{\tau_*-\tau}\Big(\mathcal{E}_t(u) + (1+\boldsymbol{d}^2(u_*,u))\int_t^{t+\tau}\alpha_r\,\mathrm{d}r - c_* + \frac{1}{\tau_*-\tau}\boldsymbol{d}^2(u_*,u)\Big).
\end{aligned}$$

Thus, $\boldsymbol{d}^2(u^\tau,u)$, $\boldsymbol{d}^2(u^\tau,u_*)$, $\boldsymbol{d}^2(u^\sigma,u)$, and $\boldsymbol{d}^2(u^\sigma,u_*)$ are locally bounded by some constant independent of τ and σ and therefore $\tau \mapsto \phi(\tau, t+\tau, u)$ is locally absolutely continuous. To calculate the derivative for a point $\tau \in (0,\tau_*)\setminus\mathcal{S}_t^u$ in the set of differentiability, divide the previous inequalities by $\tau - \sigma$ such that we have for $\sigma < \tau$:

$$-\frac{1}{2\tau\sigma}\boldsymbol{d}^2(u, u^\tau) + \frac{\mathcal{E}_{t+\tau}(u^\tau) - \mathcal{E}_{t+\sigma}(u^\tau)}{\tau - \sigma} \le \frac{\phi(\tau, t+\tau, u) - \phi(\sigma, t+\sigma, u)}{\tau - \sigma}.$$

The left-sided limit $\sigma \nearrow \tau$ yields a lower bound for the derivative $\tau \mapsto \phi(\tau, t+\tau, u)$ and analogously we gain from the inequality for $\tau < \sigma$ and the right-sided limit $\sigma \searrow \tau$ an upper bound. Since u^τ was arbitrarily chosen in the resolvent, the value of the derivative is independent of the u^τ and therefore the desired formula is true. □

Lemma 3.1.15 (Local Slope Estimate)**.** *Given $u \in \mathbf{X}, \tau > 0, t \in [0, \infty)$ and $w \in J_t^\tau(u)$, we have*

$$|\partial\mathcal{E}_t|(w) \le \frac{1}{\tau}\boldsymbol{d}(u, w). \tag{3.1.12}$$

Proof. Given $u \in \mathbf{X}$ and $w \in J_t^\tau(u)$. Since w is the minimizer of $w \mapsto \Phi(\tau, t, u; w)$ we have for an arbitrary $v \in \mathbf{X}$

$$\mathcal{E}_t(w) - \mathcal{E}_t(v) \le \frac{1}{2\tau}\boldsymbol{d}^2(u, v) - \frac{1}{2\tau}\boldsymbol{d}^2(u, w) \le \frac{1}{2\tau}\boldsymbol{d}(v, w)(\boldsymbol{d}(u, v) + \boldsymbol{d}(u, w)).$$

Dividing the equation by $\boldsymbol{d}(v, w)$ we get

$$\limsup_{v \overset{d}{\to} w} \frac{(\mathcal{E}_t(w) - \mathcal{E}_t(v))^+}{\boldsymbol{d}(v, w)} \le \limsup_{v \overset{d}{\to} w} \frac{1}{2\tau}(\boldsymbol{d}(u, v) + \boldsymbol{d}(u, w)) = \frac{\boldsymbol{d}(u, w)}{\tau}$$

which is the desired local slope estimate (3.1.12). □

3.1.3 Properties of the Time-Dependent MMS

By the previous section, the sequence $(u_k^{\boldsymbol{\tau}})$ given by the time-dependent Minimizing Movement scheme is well defined for every partition $\boldsymbol{\tau} = (\tau_1, \tau_2, \ldots)$ with $\tau_k \in (0, \tau_*)$. Next, we derive the classical estimates on energy and distance and the discrete energy dissipation inequality. These additionally require (I1) and (I2) to hold for the partition $\boldsymbol{\tau}$ and the approximation $u_0^{\boldsymbol{\tau}}$ of the initial datum u_0.

Theorem 3.1.16 (Classical Estimates)**.** *Fix a time horizon $T > 0$. Then, there is a constant C, independent of the partition $\boldsymbol{\tau}$, such that the corresponding discrete solution $(u_k^{\boldsymbol{\tau}})_{k\in\mathbb{N}}$ satisfies*

$$\sum_{k=1}^{N} \frac{1}{2\tau_k}\boldsymbol{d}^2(u_k^{\boldsymbol{\tau}}, u_{k\text{-}1}^{\boldsymbol{\tau}}) \le C, \qquad \mathcal{E}_{t_N^{\boldsymbol{\tau}}}(u_N^{\boldsymbol{\tau}}) \le C, \qquad \boldsymbol{d}^2(u_*, u_N^{\boldsymbol{\tau}}) \le C \tag{3.1.13}$$

for all $N \in \mathbb{N}$ with $t_N^{\boldsymbol{\tau}} \le T$.

Proof. For a given partition $\boldsymbol{\tau} = (\tau_1, \tau_2, \ldots)$ with $\tau_k \in (0, \tau_*)$ consider the discrete solution $(u_k^{\boldsymbol{\tau}})_{k\in\mathbb{N}}$ obtained by the Minimizing Movement scheme. Since $u_k^{\boldsymbol{\tau}}$ is a minimizer for $\Phi(\tau, t_k^{\boldsymbol{\tau}}, u_{k\text{-}1}^{\boldsymbol{\tau}}; \cdot)$ it satisfies the discrete variational inequality:

$$\frac{1}{2\tau_k}\boldsymbol{d}^2(u_{k\text{-}1}^{\boldsymbol{\tau}}, u_k^{\boldsymbol{\tau}}) + \mathcal{E}_{t_k^{\boldsymbol{\tau}}}(u_k^{\boldsymbol{\tau}}) \leq \frac{1}{2\tau_k}\boldsymbol{d}^2(u_{k\text{-}1}^{\boldsymbol{\tau}}, u_{k\text{-}1}^{\boldsymbol{\tau}}) + \mathcal{E}_{t_k^{\boldsymbol{\tau}}}(u_{k\text{-}1}^{\boldsymbol{\tau}}).$$

Rearrange and sum these inequalities from $k = 1$ to $k = N$ and exploit the telescopic sum to obtain

$$\begin{aligned}\sum_{k=1}^{N} \frac{1}{2\tau_k}\boldsymbol{d}^2(u_{k\text{-}1}^{\boldsymbol{\tau}}, u_k^{\boldsymbol{\tau}}) &\leq \sum_{k=1}^{N} \left[\mathcal{E}_{t_k^{\boldsymbol{\tau}}}(u_{k\text{-}1}^{\boldsymbol{\tau}}) - \mathcal{E}_{t_k^{\boldsymbol{\tau}}}(u_k^{\boldsymbol{\tau}})\right] \\ &= \mathcal{E}_0(u_0^{\boldsymbol{\tau}}) - \mathcal{E}_{t_N^{\boldsymbol{\tau}}}(u_N^{\boldsymbol{\tau}}) + \sum_{k=0}^{N-1} \left[\mathcal{E}_{t_{k+1}^{\boldsymbol{\tau}}}(u_k^{\boldsymbol{\tau}}) - \mathcal{E}_{t_k^{\boldsymbol{\tau}}}(u_k^{\boldsymbol{\tau}})\right].\end{aligned}$$

Since $\mathcal{E}_t$ is absolutely continuous in time, we get with (E4), (I1), and (I2)

$$\begin{aligned}&\sum_{k=1}^{N} \frac{1}{2\tau_k}\boldsymbol{d}^2(u_{k\text{-}1}^{\boldsymbol{\tau}}, u_k^{\boldsymbol{\tau}}) \\ \leq & d_1 + \frac{1}{2\tau_*}\boldsymbol{d}^2(u_*, u_N^{\boldsymbol{\tau}}) - c_* + \sum_{k=0}^{N-1} \left(1 + \boldsymbol{d}^2(u_*, u_k^{\boldsymbol{\tau}})\right) \int_{t_k^{\boldsymbol{\tau}}}^{t_{k+1}^{\boldsymbol{\tau}}} \alpha_r \,\mathrm{d}r.\end{aligned} \tag{3.1.14}$$

Furthermore, using Young's inequality with $\varepsilon = \frac{\tau_*}{2}$, we get

$$\begin{aligned}\frac{1}{2}\boldsymbol{d}^2(u_*, u_N^{\boldsymbol{\tau}}) - \frac{1}{2}\boldsymbol{d}^2(u_*, u_0^{\boldsymbol{\tau}}) &= \sum_{k=1}^{N} \frac{1}{2}\boldsymbol{d}^2(u_*, u_k^{\boldsymbol{\tau}}) - \frac{1}{2}\boldsymbol{d}^2(u_*, u_{k\text{-}1}^{\boldsymbol{\tau}}) \\ &\leq \sum_{k=1}^{N} \boldsymbol{d}(u_k^{\boldsymbol{\tau}}, u_{k\text{-}1}^{\boldsymbol{\tau}})\boldsymbol{d}(u_k^{\boldsymbol{\tau}}, u_*) \\ &\leq \frac{\tau_*}{2}\sum_{k=1}^{N} \frac{1}{2\tau_k}\boldsymbol{d}^2(u_{k\text{-}1}^{\boldsymbol{\tau}}, u_k^{\boldsymbol{\tau}}) + \frac{1}{\tau_*}\sum_{k=1}^{N} \tau_k\boldsymbol{d}^2(u_*, u_k^{\boldsymbol{\tau}}).\end{aligned}$$

Inserting the auxiliary inequality (3.1.14) from above yields

$$\begin{aligned}&\frac{1}{2}\boldsymbol{d}^2(u_*, u_N^{\boldsymbol{\tau}}) - \frac{1}{2}\boldsymbol{d}^2(u_*, u_0^{\boldsymbol{\tau}}) \\ \leq & \frac{\tau_*}{2}\Big(d_1 + \frac{1}{2\tau_*}\boldsymbol{d}^2(u_*, u_N^{\boldsymbol{\tau}}) - c_* + \sum_{k=0}^{N-1} \left(1 + \boldsymbol{d}^2(u_*, u_k^{\boldsymbol{\tau}})\right) \int_{t_k^{\boldsymbol{\tau}}}^{t_{k+1}^{\boldsymbol{\tau}}} \alpha_r \,\mathrm{d}r\Big) + \frac{1}{\tau_*}\sum_{k=1}^{N} \tau_k\boldsymbol{d}^2(u_*, u_k^{\boldsymbol{\tau}}) \\ \leq & \frac{\tau_*}{2}\Big(d_1 - c_* + \left(1 + \boldsymbol{d}^2(u_*, u_k^{\boldsymbol{\tau}})\right) \int_0^T \alpha_r \,\mathrm{d}r\Big) \\ & + \frac{1}{4}\boldsymbol{d}^2(u_*, u_N^{\boldsymbol{\tau}}) + \sum_{k=1}^{N} \Big(\frac{\tau_*}{2}\int_{t_k^{\boldsymbol{\tau}}}^{t_{k+1}^{\boldsymbol{\tau}}} \alpha_r \,\mathrm{d}r + \frac{\tau_k}{\tau_*}\Big)\boldsymbol{d}^2(u_*, u_k^{\boldsymbol{\tau}}).\end{aligned}$$

Rearrange the inequality to obtain

$$\begin{aligned}\boldsymbol{d}^2(u_*, u_N^{\boldsymbol{\tau}}) \leq& 2\boldsymbol{d}^2(u_*, u_0^{\boldsymbol{\tau}}) + 2\tau_* \left(d_1 - c^* + (1 + \boldsymbol{d}^2(u_*, u_0^{\boldsymbol{\tau}})) \int_0^T \alpha_r \,\mathrm{d}r \right) \\ &+ 4 \sum_{k=1}^N \left(\frac{\tau_*}{2} \int_{t_k^{\boldsymbol{\tau}}}^{t_{k+1}^{\boldsymbol{\tau}}} \alpha_r \,\mathrm{d}r + \frac{\tau_k}{\tau_*} \right) \boldsymbol{d}^2(u_*, u_k^{\boldsymbol{\tau}}) \\ =:& \widetilde{C}(T, \tau_*) + 4 \sum_{k=1}^N \alpha_k \boldsymbol{d}^2(u_*, u_k^{\boldsymbol{\tau}}).\end{aligned}$$

Since by assumption $\sup_k 4\alpha_k < 1$ one can apply the discrete version of Gronwall's lemma [4, Lemma 3.2.4] to conclude

$$\begin{aligned}\boldsymbol{d}^2(u_*, u_N^{\boldsymbol{\tau}}) \leq& \widehat{C}(T, \tau_*) \exp\left[\widehat{c}(T, \tau_*) \sum_{k=1}^{N-1} \alpha_k \right] \\ \leq& \widehat{C}(T, \tau_*) \exp\left[\widehat{c}(T, \tau_*) \left(\frac{\tau_*}{2} \int_0^T \alpha_r \,\mathrm{d}r + \frac{T}{\tau_*} \right) \right].\end{aligned}$$

Hence, we have proven the $\boldsymbol{d}$-boundedness of the discrete solution. With this result and with the first chain of inequalities we can deduce the estimate on the kinetic energy, i.e.,

$$\begin{aligned}\sum_{k=1}^N \frac{1}{2\tau_k} \boldsymbol{d}^2(u_{k-1}^{\boldsymbol{\tau}}, u_k^{\boldsymbol{\tau}}) &\leq d_1 + \frac{1}{2\tau_*} \boldsymbol{d}^2(u_*, u_N^{\boldsymbol{\tau}}) - c_* + \sum_{k=0}^{N-1} \left(1 + \boldsymbol{d}^2(u_*, u_k^{\boldsymbol{\tau}})\right) \int_{t_k^{\boldsymbol{\tau}}}^{t_{k+1}^{\boldsymbol{\tau}}} \alpha_r \,\mathrm{d}r \\ &\leq d_1 + \frac{1}{2\tau_*} C(T, \tau_*) - c_* + (1 + C(T, \tau_*)) \int_0^T \alpha_r \,\mathrm{d}r.\end{aligned}$$

Again, using this inequality yields the upper bound for $\mathcal{E}_{t_N^{\boldsymbol{\tau}}}(u_N^{\boldsymbol{\tau}})$, since

$$\begin{aligned}\mathcal{E}_{t_N^{\boldsymbol{\tau}}}(u_N^{\boldsymbol{\tau}}) &\leq \mathcal{E}_{t_N^{\boldsymbol{\tau}}}(u_N^{\boldsymbol{\tau}}) + \sum_{k=1}^N \frac{1}{2\tau_k} \boldsymbol{d}^2(u_{k-1}^{\boldsymbol{\tau}}, u_k^{\boldsymbol{\tau}}) \\ &\leq \mathcal{E}_0(u_0^{\boldsymbol{\tau}}) - \mathcal{E}_{t_N^{\boldsymbol{\tau}}}(u_N^{\boldsymbol{\tau}}) + \sum_{k=0}^{N-1} \left[\mathcal{E}_{t_{k+1}^{\boldsymbol{\tau}}}(u_k^{\boldsymbol{\tau}}) - \mathcal{E}_{t_k^{\boldsymbol{\tau}}}(u_k^{\boldsymbol{\tau}}) \right] \\ &\leq \mathcal{E}_0(u_0^{\boldsymbol{\tau}}) + \frac{1}{2\tau_*} \boldsymbol{d}^2(u_*, u_N^{\boldsymbol{\tau}}) - c_* + \sum_{k=0}^{N-1} \left(1 + \boldsymbol{d}^2(u_*, u_k^{\boldsymbol{\tau}})\right) \int_{t_k^{\boldsymbol{\tau}}}^{t_{k+1}^{\boldsymbol{\tau}}} \alpha_r \,\mathrm{d}r \\ &\leq d_1 + \frac{1}{2\tau_*} C(T, \tau_*) - c_* + (1 + C(T, \tau_*)) \int_0^T \alpha_r \,\mathrm{d}r.\end{aligned}$$

Hence, we have the three desired $\boldsymbol{\tau}$-independent estimates. □

Next, we define the so called *De Giorgi interpolation* $\widetilde{u}_t^{\boldsymbol{\tau}} : [0,\infty) \to \mathbf{X}$, via

$$\widetilde{u}_0^{\boldsymbol{\tau}} = u_0^{\boldsymbol{\tau}}, \qquad \widetilde{u}_{t_k^{\boldsymbol{\tau}}}^{\boldsymbol{\tau}} = u_k^{\boldsymbol{\tau}}, \qquad \widetilde{u}_{t_k^{\boldsymbol{\tau}}+\sigma}^{\boldsymbol{\tau}} \in J_{t_k^{\boldsymbol{\tau}}+\sigma}^{\sigma}(u_k^{\boldsymbol{\tau}}) \qquad \text{for } \sigma \in (0,\tau_{k+1}), \ k \in \mathbb{N}_0,$$

which satisfies the following *discrete energy inequality*.

Theorem 3.1.17 (Discrete Energy Inequality)**.** *The De Giorgi interpolation $\widetilde{u}_t^{\boldsymbol{\tau}}$ satisfies:*

$$\begin{aligned}&\mathcal{E}_{t_N^{\boldsymbol{\tau}}}(\widetilde{u}_{t_N^{\boldsymbol{\tau}}}^{\boldsymbol{\tau}}) + \sum_{k=1}^{N} \frac{1}{2\tau_k}\boldsymbol{d}^2(u_{k-1}^{\boldsymbol{\tau}}, u_k^{\boldsymbol{\tau}}) + \frac{1}{2}\int_0^{t_N^{\boldsymbol{\tau}}} |\partial\mathcal{E}_t|^2(\widetilde{u}_t^{\boldsymbol{\tau}})\,\mathrm{d}t \\ \le&\mathcal{E}_0(u_0^{\boldsymbol{\tau}}) + \int_0^{t_N^{\boldsymbol{\tau}}} \partial_t\mathcal{E}_t(\widetilde{u}_t^{\boldsymbol{\tau}})\,\mathrm{d}t.\end{aligned} \tag{3.1.15}$$

Proof. From Lemma 3.1.14 we know that the map $\sigma \mapsto \phi(\sigma, t+\sigma, u)$ is locally absolutely continuous and we can compute the derivative at almost all σ, which is given by

$$\frac{d}{d\sigma}\phi(\sigma, t+\sigma, u) = -\frac{1}{2\sigma^2}\boldsymbol{d}^2(u,w) + \partial_t\mathcal{E}_{t+\sigma}(w) \qquad \forall\ w \in J_{t+\sigma}^{\sigma}(u).$$

Choose $t = t_k^{\boldsymbol{\tau}}$, $u = u_k^{\boldsymbol{\tau}}$ and use $\widetilde{u}_{t_k^{\boldsymbol{\tau}}+\sigma}^{\boldsymbol{\tau}} \in J_{t_k^{\boldsymbol{\tau}}+\sigma}^{\sigma}(u_k^{\boldsymbol{\tau}})$ when integrating the equation with respect to σ from $\varepsilon > 0$ to τ_k:

$$\phi(\tau, t_{k+1}^{\boldsymbol{\tau}}, u_k^{\boldsymbol{\tau}}) - \phi(\varepsilon, t_k^{\boldsymbol{\tau}}+\varepsilon, u_k^{\boldsymbol{\tau}}) = \int_\varepsilon^{\tau_k} -\frac{1}{2\sigma^2}\boldsymbol{d}^2(u_k^{\boldsymbol{\tau}}, \widetilde{u}_{t_k^{\boldsymbol{\tau}}+\sigma}^{\boldsymbol{\tau}}) + \partial_t\mathcal{E}_{t_k^{\boldsymbol{\tau}}+\sigma}(\widetilde{u}_{t_k^{\boldsymbol{\tau}}+\sigma}^{\boldsymbol{\tau}})\,\mathrm{d}\sigma.$$

Use $\widetilde{u}_{t_{k+1}^{\boldsymbol{\tau}}}^{\boldsymbol{\tau}} \in J_{t_{k+1}^{\boldsymbol{\tau}}}^{\boldsymbol{\tau}}(u_k^{\boldsymbol{\tau}})$ and Lemma 3.1.13 to perform the limit $\varepsilon \searrow 0$ to obtain

$$\frac{1}{2\tau_k}\boldsymbol{d}^2(u_{k+1}^{\boldsymbol{\tau}}, u_k^{\boldsymbol{\tau}}) + \mathcal{E}_{t_{k+1}^{\boldsymbol{\tau}}}(u_{k+1}^{\boldsymbol{\tau}}) - \mathcal{E}_{t_k^{\boldsymbol{\tau}}}(u_k^{\boldsymbol{\tau}}) = \int_0^{\tau_k} -\frac{1}{2\sigma^2}\boldsymbol{d}^2(u_k^{\boldsymbol{\tau}}, \widetilde{u}_{t_k^{\boldsymbol{\tau}}+\sigma}^{\boldsymbol{\tau}}) + \partial_t\mathcal{E}_{t_k^{\boldsymbol{\tau}}+\sigma}(\widetilde{u}_{t_k^{\boldsymbol{\tau}}+\sigma}^{\boldsymbol{\tau}})\,\mathrm{d}\sigma.$$

Apply Lemma 3.1.15 with $t = t_k^{\boldsymbol{\tau}} + \sigma, u = u_k^{\boldsymbol{\tau}}$ to obtain

$$\frac{1}{2\tau_k}\boldsymbol{d}^2(u_{k+1}^{\boldsymbol{\tau}}, u_k^{\boldsymbol{\tau}}) + \mathcal{E}_{t_{k+1}^{\boldsymbol{\tau}}}(u_{k+1}^{\boldsymbol{\tau}}) - \mathcal{E}_{t_k^{\boldsymbol{\tau}}}(u_k^{\boldsymbol{\tau}}) \le \frac{1}{2}\int_{t_k^{\boldsymbol{\tau}}}^{t_{k+1}^{\boldsymbol{\tau}}} -|\partial\mathcal{E}_t|^2(\widetilde{u}_t^{\boldsymbol{\tau}}) + \partial_t\mathcal{E}_t(\widetilde{u}_t^{\boldsymbol{\tau}})\,\mathrm{d}t.$$

Summation from $k = 0$ to $N-1$ yields the desired discrete energy inequality. □

3.1.4 Convergence

In this section we complete the proof of the Main Theorem 3.0.1, i.e., firstly, we prove the convergence in the weak $\boldsymbol{\sigma}$-topology of the piecewise constant interpolation $\overline{u}_t^{\boldsymbol{\tau}}$ to a limit curve u_t^*; secondly, we prove that the De Giorgi interpolation $\widetilde{u}_t^{\boldsymbol{\tau}}$ converges also in the $\boldsymbol{\sigma}$-topology to the same limit curve u_t^*; thirdly, we pass in the discrete energy inequality (3.1.15) to the limit $\boldsymbol{\tau} \to 0$ to show that the limit curve u_t^* is indeed a curve of steepest descent for $\mathcal{E}_t$ emanating from u_0.

Theorem 3.1.18 (Convergence of the Piecewise Constant Interpolation)**.** *Given a sequence of partitions* $(\boldsymbol{\tau}_n)_{n\in\mathbb{N}}$ *with* $\sup_k \tau_{k,n} \to 0$ *and satisfying (I1), and let* $(u_0^{\boldsymbol{\tau}_n})_{n\in\mathbb{N}}$ *be an approximation for* $u_0 \in \mathcal{D}(\mathcal{E}_0)$ *satisfying (I2). Then, there exists an* L^2*-absolutely continuous limit curve* $u_t^* \in \mathrm{AC}^2(0,\infty;(\mathbf{X},\boldsymbol{d}))$ *such for a (non-relabelled) subsequence of* $(\overline{u}_t^{\boldsymbol{\tau}_n})_{n\in\mathbb{N}}$*:*

$$\overline{u}_t^{\boldsymbol{\tau}_n} \overset{\boldsymbol{\sigma}}{\rightharpoonup} u_t^* \qquad \forall\, t \in [0,\infty)\,.$$

Proof. Fix some $T > 0$ and define the discrete derivative A_t^n as

$$A_t^n := \frac{1}{\tau_{n,k}} \boldsymbol{d}(u_{k\text{-}1}^{\boldsymbol{\tau}_n}, u_k^{\boldsymbol{\tau}_n}) \qquad \text{for} \quad t \in \left[t_{k\text{-}1}^{\boldsymbol{\tau}_n}, t_k^{\boldsymbol{\tau}_n}\right). \tag{3.1.16}$$

Using the classical estimates for the Minimizing Movement scheme of Theorem 3.1.16, we get for all $t_N^{\boldsymbol{\tau}_n} < T$

$$\int_0^{t_N^{\boldsymbol{\tau}_n}} (A_t^n)^2 \,\mathrm{d}t = \sum_{k=0}^{N-1} \int_{t_{k\text{-}1}^{\boldsymbol{\tau}_n}}^{t_k^{\boldsymbol{\tau}_n}} \left(\frac{1}{\tau_{n,k}} \boldsymbol{d}(u_{k\text{-}1}^{\boldsymbol{\tau}_n}, u_k^{\boldsymbol{\tau}_n})\right)^2 \mathrm{d}t = \sum_{k=1}^{N} \frac{1}{\tau_{n,k}} \boldsymbol{d}^2(u_{k\text{-}1}^{\boldsymbol{\tau}_n}, u_k^{\boldsymbol{\tau}_n}) \le 2C(T).$$

Thus $A_t^n \in \mathrm{L}^2(0,T)$ and the $\mathrm{L}^2(0,T)$-norm of A_t^n is uniformly bounded in $\boldsymbol{\tau}_n$. Therefore A_t^n possesses a $\mathrm{L}^2(0,T)$-weakly convergent non-relabelled subsequence with limit $A_t \in \mathrm{L}^2(0,T)$. Choose $0 \le s \le t \le T$ arbitrary and define $k_t^n := \max\left\{k \mid t_k^{\boldsymbol{\tau}_n} \le t\right\}$, then

$$\boldsymbol{d}(\overline{u}_s^{\boldsymbol{\tau}_n}, \overline{u}_t^{\boldsymbol{\tau}_n}) \le \sum_{k=k_s^n+1}^{k_t^n} \boldsymbol{d}(u_{k\text{-}1}^{\boldsymbol{\tau}_n}, u_k^{\boldsymbol{\tau}_n}) = \sum_{k=k_s^n+1}^{k_t^n} \int_{t_{k\text{-}1}^{\boldsymbol{\tau}_n}}^{t_k^{\boldsymbol{\tau}_n}} \frac{1}{\tau_{n,k}} \boldsymbol{d}(u_{k\text{-}1}^{\boldsymbol{\tau}_n}, u_k^{\boldsymbol{\tau}_n}) \,\mathrm{d}r = \int_{t_{k_s^n}^{\boldsymbol{\tau}_n}}^{t_{k_t^n}^{\boldsymbol{\tau}_n}} A_r^n \,\mathrm{d}r.$$

Taking the limit $n \to \infty$ yields now

$$\limsup_{n\to\infty} \boldsymbol{d}(\overline{u}_s^{\boldsymbol{\tau}_n}, \overline{u}_t^{\boldsymbol{\tau}_n}) \le \int_s^t A_r \,\mathrm{d}r. \tag{3.1.17}$$

On the other hand, by the classical estimates (3.1.13) for the Minimizing Movement scheme one has for $t \in (t_{k\text{-}1}^{\boldsymbol{\tau}}, t_k^{\boldsymbol{\tau}}]$

$$\mathcal{E}_0(\overline{u}_t^{\boldsymbol{\tau}_n}) = \mathcal{E}_0(u_k^{\boldsymbol{\tau}_n}) \le \mathcal{E}_{t_k^{\boldsymbol{\tau}}}(u_k^{\boldsymbol{\tau}_n}) + (1 + \boldsymbol{d}^2(u_*, u_k^{\boldsymbol{\tau}_n})) \int_0^{t_k^{\boldsymbol{\tau}}} \alpha_t \,\mathrm{d}t \le C + (1+C) \int_0^T \alpha_t \,\mathrm{d}t.$$

Hence, the piecewise constant interpolation $\overline{u}_t^{\boldsymbol{\tau}_n}$ is contained in some sublevel of $\mathcal{E}_0$ uniformly in $t \in [0,T]$. Estimate (3.1.13) additionally ensures the uniform $\boldsymbol{d}$-boundedness of $\overline{u}_{\boldsymbol{\tau}_n}(t)$ and therefore, using the $\boldsymbol{\sigma}$-compactness of $\mathcal{E}_0$, $\overline{u}_t^{\boldsymbol{\tau}_n}$ is contained in some $\boldsymbol{\sigma}$-compact set K for all $t \in [0,T]$ and for all $n \in \mathbb{N}$. Therefore, we can apply the refined Arzelá-Ascoli Theorem [4, Proposition 3.3.1] yielding the existence of a non-relabelled subsequence and a limit curve $u_t^* : [0,T] \to \mathbf{X}$ such that $\overline{u}_t^{\boldsymbol{\tau}_n} \overset{\boldsymbol{\sigma}}{\rightharpoonup} u_t^*$ for each fixed $t \in [0,T]$. Consequently, by a diagonal argument we can extend u_t^* on $[0,\infty)$ such that $u_t^* \in \mathrm{AC}^2(0,\infty;(\mathbf{X},\boldsymbol{d}))$ and $\overline{u}_t^{\boldsymbol{\tau}_n} \overset{\boldsymbol{\sigma}}{\rightharpoonup} u_t^*$ for all $t \in [0,\infty)$. □

Theorem 3.1.19 (Convergence of the De Giorgi Interpolation). *Under the same assumptions as in Theorem 3.1.18. Let $\widetilde{u}_t^{\tau_n}$ be the corresponding De Giorgi interpolation and let u_t^* be the limit curve of the piecewise constant interpolation $(\overline{u}_t^{\tau_n})_{n\in\mathbb{N}}$. Then, there exists a non-relabelled subsequence of $(\widetilde{u}_t^{\tau_n})_{n\in\mathbb{N}}$ such that*

$$\widetilde{u}_t^{\tau_n} \overset{\boldsymbol{\sigma}}{\rightharpoonup} u_t^* \qquad \forall\, t \in [0,\infty)\setminus\mathcal{N}_\alpha. \tag{3.1.18}$$

where $\mathcal{N}_\alpha$ is the set of all non-Lebesgue points of α_t.

Proof. Fix $T>0$ and define in the same manner $k_t^n := \max\{k \mid t_k^{\tau_n} \le t\}$. As before, we prove that the family of De Giorgi interpolations $\widetilde{u}_t^{\tau_n}$ is contained in some $\boldsymbol{\sigma}$-compact set $\widetilde{K}$. The $\boldsymbol{d}$-boundedness of $\widetilde{u}_t^{\tau_n}$ follows by (3.1.10), since for $t = t_{k_t^n}^{\tau_n} + \sigma$

$$\begin{aligned}
\boldsymbol{d}^2(\widetilde{u}_t^{\tau_n}, \overline{u}_t^{\tau_n}) = \boldsymbol{d}^2(\widetilde{u}_t^{\tau_n}, u_{k_t^n}^{\tau_n}) &\le \frac{4\sigma\tau_*}{\tau_*-\sigma}\Big(\Phi(\sigma,t,u_{k_t^n}^{\tau_n},\widetilde{u}_t^{\tau_n}) - c_* + \frac{1}{\tau_*-\sigma}\boldsymbol{d}^2(u_*,u_{k_t^n}^{\tau_n})\Big)\\
&= \frac{4\sigma\tau_*}{\tau_*-\sigma}\Big(\phi(\sigma,t,u_{k_t^n}^{\tau_n}) - c_* + \frac{1}{\tau_*-\sigma}\boldsymbol{d}^2(u_*,u_{k_t^n}^{\tau_n})\Big)\\
&\le \frac{4\sigma\tau_*}{\tau_*-\sigma}\Big(\mathcal{E}_t(u_{k_t^n}^{\tau_n}) - c_* + \frac{1}{\tau_*-\sigma}\boldsymbol{d}^2(u_*,u_{k_t^n}^{\tau_n})\Big).
\end{aligned}$$

The first term on the right hand side is bounded by the constant given in theorem 3.1.16. By the computation in the proof of theorem 3.1.18 the discrete solution $u_k^{\tau_n}$ is bounded by some constant independent of $t\in[0,T]$ and $\tau\in(0,\tau_*)$. The second term is bounded by the classical estimates (3.1.13), hence the De Giorgi interpolation $\widetilde{u}_{\tau_n}$ is locally $\boldsymbol{d}$-bounded.

Next we prove the boundedness of $\mathcal{E}_t(\widetilde{u}_t^{\tau_n})$, to do so, we use $\widetilde{u}_t^{\tau_n} \in J_t^\sigma(u_{k_t^n}^{\tau_n})$ and the monotonicity property (3.1.8) to obtain

$$\mathcal{E}_t(\widetilde{u}_t^{\tau_n}) \le \frac{1}{2\sigma}\boldsymbol{d}^2(u_{k_t^n}^{\tau_n},\widetilde{u}_t^{\tau_n}) + \mathcal{E}_t(\widetilde{u}_t^{\tau_n}) = \phi(\sigma,t,u_{k_t^n}^{\tau_n}) \le \mathcal{E}_t(u_{k_t^n}^{\tau_n}).$$

Hence, we exploit (E4) twice and the $\boldsymbol{d}$-boundedness of $\widetilde{u}_t^{\tau_n}$ to get an estimate for $\mathcal{E}_0(\widetilde{u}_t^{\tau_n})$:

$$\begin{aligned}
\mathcal{E}_0(\widetilde{u}_t^{\tau_n}) \le& \mathcal{E}_t(\widetilde{u}_t^{\tau_n}) + (1+\boldsymbol{d}^2(u_*,\widetilde{u}_t^{\tau_n}))\int_0^t \alpha_t\,\mathrm{d}t\\
\le& \mathcal{E}_t(u_{k_t^n}^{\tau_n}) + (1+\boldsymbol{d}^2(u_*,\widetilde{u}_t^{\tau_n})))\int_0^t \alpha_t\,\mathrm{d}t\\
\le& \mathcal{E}_{t_{k_t^n}^{\tau_n}}(u_{k_t^n}^{\tau_n}) + (1+\boldsymbol{d}^2(u_*,u_{k_t^n}^{\tau_n}))\int_{t_{k_t^n}^{\tau_n}}^t \alpha_t\,\mathrm{d}t + (1+\boldsymbol{d}^2(u_*,\widetilde{u}_t^{\tau_n})))\int_0^t \alpha_t\,\mathrm{d}t.
\end{aligned}$$

Again, the first terms is bounded by (3.1.13), the second term is also bounded by (3.1.13) and the $\mathrm{L}^1(0,T)$-norm of α_t, and the third term is bounded by the $\mathrm{L}^1(0,T)$-norm of α_t and the $\boldsymbol{d}$-boundedness of $\widetilde{u}_t^{\tau_n}$. Hence, the De Giorgi interpolations $\widetilde{u}_t^{\tau_n}$ are contained in some sublevel of $\mathcal{E}_0$, which is by assumption (E3) compact in the $\boldsymbol{\sigma}$-topology.

To apply the refined Arzelá-Ascoli theorem [4, Proposition 3.3.1], it remains to prove an estimate in terms of the modulus of continuity. Note that by Lemma 3.1.10, one has

$$\begin{aligned}&\boldsymbol{d}^2(\overline{u}_t^{\boldsymbol{\tau}_n}, \widetilde{u}_t^{\boldsymbol{\tau}_n})\\ \le&\boldsymbol{d}^2(u_{k_t^n}^{\boldsymbol{\tau}_n}, u_{k_t^n+1}^{\boldsymbol{\tau}_n}) + 2\tau_{n,k}\frac{t - t_{k_t^n}^{\boldsymbol{\tau}_n}}{t_{k_t^n+1}^{\boldsymbol{\tau}_n} - t}\int_t^{t_{k_t^n+1}^{\boldsymbol{\tau}_n}} \alpha_t \,\mathrm{d}t \Big(2 + \boldsymbol{d}^2(u_*, \widetilde{u}_t^{\boldsymbol{\tau}_n}) + \boldsymbol{d}^2(u_*, u_{k_t^n}^{\boldsymbol{\tau}_n})\Big)\\ \le&\boldsymbol{\tau}_n C(T)\Big(1 + \frac{t - t_{k_t^n}^{\boldsymbol{\tau}_n}}{t_{k_t^n+1}^{\boldsymbol{\tau}_n} - t}\int_t^{t_{k_t^n+1}\boldsymbol{\tau}_n} \alpha_t \,\mathrm{d}t\Big).\end{aligned}$$

Here, we estimated the first term by (3.1.13) and the last term using the $\boldsymbol{d}$-boundedness of $\overline{u}_t^{\boldsymbol{\tau}_n}$ and $\widetilde{u}_t^{\boldsymbol{\tau}_n}$. By the assumption (E4), the limit behavior of the last term is given by

$$\limsup_{n\to\infty} \frac{t - t_{k_t^n}^{\boldsymbol{\tau}_n}}{t_{k_t^n+1}^{\boldsymbol{\tau}_n} - t}\int_t^{t_{k_t^n+1}\boldsymbol{\tau}_n} \alpha_t \,\mathrm{d}t = 0 \qquad \text{for each } t \in [0,\infty)\backslash\mathcal{N}_\alpha,$$

where $\mathcal{N}_\alpha$ is the set of all $t \in [0,\infty)$ which are not Lebesgue-points for α_t. Therefore, we can deduce for $s,t \in [0,T]\,\backslash\mathcal{N}_\alpha$

$$\begin{aligned}\limsup_{n\to\infty} \boldsymbol{d}^2(\widetilde{u}_t^{\boldsymbol{\tau}_n}, \widetilde{u}_s^{\boldsymbol{\tau}_n}) \le& \limsup_{n\to\infty} 3\left(\boldsymbol{d}^2(\widetilde{u}_t^{\boldsymbol{\tau}_n}, \overline{u}_t^{\boldsymbol{\tau}_n}) + \boldsymbol{d}^2(\overline{u}_t^{\boldsymbol{\tau}_n}, \overline{u}_s^{\boldsymbol{\tau}_n}) + \boldsymbol{d}^2(\overline{u}_s^{\boldsymbol{\tau}_n}, \widetilde{u}_s^{\boldsymbol{\tau}_n})\right)\\ \le& \limsup_{n\to\infty} 6\boldsymbol{\tau}_n C(T) + \limsup_{n\to\infty} 3\boldsymbol{d}^2(\overline{u}_t^{\boldsymbol{\tau}_n}, \overline{u}_s^{\boldsymbol{\tau}_n})\\ \le& 3\int_s^t A_t \,\mathrm{d}t.\end{aligned}$$

Hence, we can apply the refined version of the Arzelà-Ascoli theorem [4, Proposition 3.3.1], to conclude the pointwise convergence of the sequence $(\widetilde{u}_t^{\boldsymbol{\tau}_n})_{n\in\mathbb{N}}$ of De Giorgi interpolations on $[0,T]$ with respect to the $\boldsymbol{\sigma}$-topology to an absolutely continuous curve $\widetilde{u}_t^* \in \mathrm{AC}^2(0,\infty;(\mathbf{X},\boldsymbol{d}))$ for a non-relabelled subsequence. An additional diagonal argument yields the convergence with respect to the $\boldsymbol{\sigma}$-topology on $[0,\infty)$ for a further (non-relabelled) subsequence.

In particular, the two limit curves u_t^* and $\widetilde{u}_t^*$ have to agree, since by the compatibility assumption to the weak topology $\boldsymbol{\sigma}$ one has at least on the set $[0,\infty)\,\backslash\mathcal{N}_\alpha$

$$\begin{aligned}\boldsymbol{d}^2(\widetilde{u}_t^*, u_t^*) \le& \liminf_{n\to\infty} \boldsymbol{d}^2(\widetilde{u}_t^{\boldsymbol{\tau}_n}, \overline{u}_t^{\boldsymbol{\tau}_n})\\ \le& \liminf_{k\to\infty} \boldsymbol{\tau}_n C(T)\Big(1 + \frac{t - t_{k_t^n}^{\boldsymbol{\tau}_n}}{t_{k_t^n+1}^{\boldsymbol{\tau}_n} - t}\int_t^{t_{k_t^n+1}\boldsymbol{\tau}_n} \alpha_t \,\mathrm{d}t\Big)\\ =& 0.\end{aligned}$$

Hence, the two limit curves $\widetilde{u}_t^*$ and u_t^* have to agree for each $t \ge 0$. $\square$

Theorem 3.1.20 (Existence of a Curve of Steepest Descent)**.** *Under the same assumptions as in Theorem 3.1.18 and that $\mathcal{E}$ satisfies (E0). The limit curve u_t^* from there is a curve of steepest descent with respect $\mathcal{E}_t$ in the sense of definition 2.3.3, i.e., $u_t^* \in \mathrm{AC}^2(0,\infty;(\mathbf{X},\boldsymbol{d}))$ and u_t^* satisfies the energy balance for each $T>0$:*

$$\mathcal{E}_T(u_T^*) + \frac{1}{2}\int_0^T \left|u_t'\right|^2 \mathrm{d}t + \frac{1}{2}\int_0^T |\partial\mathcal{E}_t|^2(u_t^*)\,\mathrm{d}t = \mathcal{E}_0(u_0^*) + \int_0^T \partial_t\mathcal{E}_t(u_t^*)\,\mathrm{d}t.$$

Proof. Given the limit curve $u_t^* \in \mathrm{AC}^2(0,\infty;(\mathbf{X},\boldsymbol{d}))$ obtained by the time-dependent Minimizing Movement scheme starting at $u_0^{\boldsymbol{\tau}}$ using the sequence of partitions $(\boldsymbol{\tau}_n)_{n\in\mathbb{N}}$ of of step sizes $\tau_{n,k} \in (0,\tau_*)$ such that $\sup_k \tau_{n,k} \to 0$ and (I1)&(I2) are satisfied. Without loss of generality $\overline{u}_t^{\boldsymbol{\tau}_n} \overset{\boldsymbol{\sigma}}{\rightharpoonup} u_t^*$ and $\widetilde{u}_t^{\boldsymbol{\tau}_n} \overset{\boldsymbol{\sigma}}{\rightharpoonup} u_t^*$ on the whole sequence. We know that the De Giorgi interpolation satisfies the discrete energy inequality (3.1.15), i.e., for $N = \max\{k \mid t_k^{\boldsymbol{\tau}_n} \le T\}$

$$\begin{aligned}&\mathcal{E}_{t_N^{\boldsymbol{\tau}_n}}(\widetilde{u}_{t_N^{\boldsymbol{\tau}_n}}^{\boldsymbol{\tau}_n}) + \sum_{k=1}^{N}\frac{1}{2\tau_{n,k}}\boldsymbol{d}^2(u_{k\text{-}1}^{\boldsymbol{\tau}_n},u_k^{\boldsymbol{\tau}_n}) + \frac{1}{2}\int_0^{t_N^{\boldsymbol{\tau}_n}} |\partial\mathcal{E}_t|^2(\widetilde{u}_t^{\boldsymbol{\tau}_n})\,\mathrm{d}t\\ \le&\mathcal{E}_0(u_0^{\boldsymbol{\tau}_n}) + \int_0^{t_N^{\boldsymbol{\tau}_n}} \partial_t\mathcal{E}_t(\widetilde{u}_t^{\boldsymbol{\tau}_n})\,\mathrm{d}t.\end{aligned}$$

Fix $T \in [0,\infty)$ and compute the limes inferior of the l.h.s. of the equation above. Since $\widetilde{u}_{t_N^{\boldsymbol{\tau}_n}}^{\boldsymbol{\tau}_n} = \overline{u}_T^{\boldsymbol{\tau}_n} \overset{\boldsymbol{\sigma}}{\rightharpoonup} u_T^*$, we have by the lower semi-$\boldsymbol{\sigma}$-continuity in space and the absolute continuity in time of $\mathcal{E}_t$:

$$\mathcal{E}_T(u_T^*) \le \liminf_{n\to\infty}\left[\mathcal{E}_T(\widetilde{u}_{t_N^{\boldsymbol{\tau}_n}}^{\boldsymbol{\tau}_n}) - (1+\boldsymbol{d}^2(u_*,\widetilde{u}_{t_N^{\boldsymbol{\tau}_n}}^{\boldsymbol{\tau}_n}))\int_{t_N^{\boldsymbol{\tau}_n}}^T \alpha_t\,\mathrm{d}t\right] \le \liminf_{n\to\infty}\mathcal{E}_{t_N^{\boldsymbol{\tau}}}(\widetilde{u}_{t_N^{\boldsymbol{\tau}}}^{\boldsymbol{\tau}}).$$

In the proof of Theorem 3.1.18, we have seen that the discrete derivative A_t^n converges weakly to A_t in $\mathrm{L}^2(0,T)$, with A_t is one possible modulus of continuity in the definition of absolute continuity. Furthermore, since the metric derivative $|(u_t^*)'|$ is the smallest modulus of continuity, one has $|(u_t^*)'| \le A_t$ almost everywhere. The weak lower semi-continuity of the $\mathrm{L}^2(0,T)$-norm implies then:

$$\frac{1}{2}\int_0^T |(u_t^*)'|^2\,\mathrm{d}t \le \liminf_{n\to\infty}\frac{1}{2}\int_0^{t_{k_T^n}^{\boldsymbol{\tau}_n}} (A_t^n)^2\,\mathrm{d}t = \liminf_{n\to\infty}\sum_{k=1}^{k_T^n}\frac{1}{2\tau_{n,k}}\boldsymbol{d}^2(u_{k\text{-}1}^{\boldsymbol{\tau}_n},u_k^{\boldsymbol{\tau}_n}).$$

The sequence $(\widetilde{u}_t^{\boldsymbol{\tau}_n})_{n\in\mathbb{N}}$ of De Giorgi interpolations converges weakly almost everywhere in t, so using Fatou's lemma and the lower semi-$\boldsymbol{\sigma}$-continuity of the local slope $|\partial\mathcal{E}_t|$ yields for the last term on the l.h.s.

$$\frac{1}{2}\int_0^T |\partial\mathcal{E}_t|^2(u_t^*)\,\mathrm{d}t \le \frac{1}{2}\int_0^T \liminf_{n\to\infty}|\partial\mathcal{E}_t|^2(\widetilde{u}_t^{\boldsymbol{\tau}_n})\,\mathrm{d}t \le \liminf_{n\to\infty}\frac{1}{2}\int_0^{t_{k_T^n}^{\boldsymbol{\tau}_n}} |\partial\mathcal{E}_t|^2(\widetilde{u}_t^{\boldsymbol{\tau}_n})\,\mathrm{d}t.$$

At last, we compute the limit of the right-hand side by applying the dominated convergence theorem. Clearly, we have pointwise convergence of $\partial_t \mathcal{E}_t(\widetilde{u}_t^{\tau_n})$ to $\partial_t \mathcal{E}_t(u_t^*)$ almost everywhere. Since the De Giorgi interpolation $\widetilde{u}_t^{\tau_n}$ is locally contained in a $\boldsymbol{\sigma}$-compact set and $\partial_t \mathcal{E}_t$ is $\boldsymbol{\sigma}$-continuous, the integrand is uniformly bounded by some constant. Hence we can conclude with the dominated convergence theorem that

$$\int_0^T \partial_t \mathcal{E}_t(u_t^*)\,\mathrm{d}t = \lim_{n\to\infty} \int_0^{t_{k_T^n}^{\tau_n}} \partial_t \mathcal{E}_t(\widetilde{u}_t^{\tau_n})\,\mathrm{d}t.$$

Lastly, by (I2) we have

$$\lim_{n\to\infty} \mathcal{E}_0(u_0^{\tau_n}) = \mathcal{E}_0(u_0).$$

Summarized, we have the following energy inequality:

$$\mathcal{E}_T(u_T^*) + \frac{1}{2}\int_0^T |(u_t^*)'|^2\,\mathrm{d}t + \frac{1}{2}\int_0^T |\partial\mathcal{E}_t|^2(u_t^*)\,\mathrm{d}t \le \mathcal{E}_0(u_0) + \int_0^T \partial_t\mathcal{E}_t(u_t^*)\,\mathrm{d}t.$$

The reversed inequality follows now by the chain rule assumption (E0), since the energy inequality above yields an upper estimate for the $\mathrm{L}^1(0,T)$-norm of $|\partial\mathcal{E}_t|(u_t^*)|(u_t^*)'|$, i.e.,

$$\begin{aligned}\int_0^T |\partial\mathcal{E}_t|(u_t^*)|(u_t^*)'|\,\mathrm{d}t &\le \frac{1}{2}\int_0^T |(u_t^*)'|^2\,\mathrm{d}t + \frac{1}{2}\int_0^T |\partial\mathcal{E}_t|^2(u_t^*)\,\mathrm{d}t \\ &\le \mathcal{E}_0(u_0) - \mathcal{E}_T(u_T^*) + \int_0^T \partial_t\mathcal{E}_t(u_t^*)\,\mathrm{d}t.\end{aligned}$$

The right-hand-side is is always finite and therefore $|\partial\mathcal{E}_t|(u_t^*)|(u_t^*)'| \in \mathrm{L}^1_{\mathrm{loc}}(0,\infty)$. The boundedness from above of $\mathcal{E}_t(u_t^*)$ is given by

$$\mathcal{E}_t(u_t^*) \le \liminf_{n\to\infty} \mathcal{E}_{t_N^{\tau_n}}(\overline{u}_T^{\tau_n}) = \mathcal{E}_{t_N^{\tau_n}}(u_N^{\tau_n}) \le C$$

thanks to the classical estimate (3.1.13). Thus, we can apply the chain rule inequality (3.1.3) to obtain

$$\begin{aligned}\int_0^T |\partial\mathcal{E}_t|(u_t^*)|(u_t^*)'|\,\mathrm{d}t &\le \frac{1}{2}\int_0^T |(u_t^*)'|^2\,\mathrm{d}t + \frac{1}{2}\int_0^T |\partial\mathcal{E}_t|^2(u_t^*)\,\mathrm{d}t \\ &\le \mathcal{E}_0(u_0) - \mathcal{E}_T(u_T^*) + \int_0^T \partial_t\mathcal{E}_t(u_t^*)\,\mathrm{d}t \\ &\le \int_0^T |\partial\mathcal{E}_t|(u_t^*)|(u_t^*)'|\,\mathrm{d}t.\end{aligned}$$

Therefore all inequalities have to be equalities and u_t^* is in fact a curve of steepest descent for the functional $\mathcal{E}_t$. □

3.2 Application to Non-autonomous Fokker-Planck Equation

As a particular application of the temporal discretization in the general framework of abstract metric spaces, we consider a particular example, namely the non-autonomous and non-linear drift-diffusion equation

$$\partial_t \rho_t = \Delta \rho_t^m + \mathrm{div}(\rho_t \nabla V_t) + \mathrm{div}(\rho_t (\nabla W_t * \rho_t)), \tag{3.2.1}$$

with non-flux boundary condition in a domain Ω, which is an open, bounded and connected domain Ω with Lipschitz continuous boundary $\partial\Omega$ and normal derivative $\boldsymbol{n}$ or is equal to the entire space $\Omega = \mathbb{R}^{\mathrm{d}}$. The sought-for solution $\rho_t : [0,\infty) \times \Omega \to [0,\infty]$ should be nonnegative and preserves the initial mass. Here, the by now standard framework of this equation is the L^2-Wasserstein space $(\mathcal{P}_2(\Omega), \mathbf{W}_2)$ with the free energy functional

$$\mathcal{E}_t(\mu) := \begin{cases} \int_\Omega \rho \log(\rho) + V_t \rho + \frac{1}{2}(W_t * \rho)\rho \,\mathrm{d}x & \text{if } m = 1, \\ \int_\Omega \frac{1}{m-1}\rho^m + V_t \rho + \frac{1}{2}(W_t * \rho)\rho \,\mathrm{d}x & \text{if } m > 1, \end{cases} \tag{3.2.2}$$

if $\mu = \rho \,\mathrm{d}\mathcal{L}^d$ and otherwise we set $\mathcal{E}_t(\mu) = \infty$, see section 2.4 for more details.

Method. Using the notation of the L^2-Wasserstein framework, the approximation via the time-dependent implicit Euler method reads than as:

Scheme. For a partition $\boldsymbol{\tau} := (\tau_1, \tau_2, \ldots)$ of step sizes $\tau_k \in (0, \tau_*)$ let an initial condition $\rho_0^{\boldsymbol{\tau}}$ be given that approximates ρ_0. Then define inductively a discrete solution $(\rho_k^{\boldsymbol{\tau}})_{k\in\mathbb{N}}$ such that each $\rho_k^{\boldsymbol{\tau}}$ with $k = 1, 2, \ldots$ is a minimizer of the Moreau-Yosida-penalized energy functional

$$\rho \mapsto \Phi(\tau, t_k^{\boldsymbol{\tau}}, \rho_{k\text{-}1}^{\boldsymbol{\tau}}; \rho) := \frac{1}{2\tau_k} \mathbf{W}_2^2(\rho_{k\text{-}1}^{\boldsymbol{\tau}}, \rho) + \mathcal{E}_{t_k^{\boldsymbol{\tau}}}(\rho), \tag{3.2.3}$$

where $t_k^{\boldsymbol{\tau}} = \sum_{l=1}^k \tau_l$ for $k \geq 1$.

Define the corresponding piecewise constant interpolation $\overline{\rho}_t^{\boldsymbol{\tau}} : [0,\infty) \to \mathcal{P}_2(\Omega)$ via

$$\overline{\rho}_0^{\boldsymbol{\tau}} = \rho_0^{\boldsymbol{\tau}}, \qquad \overline{\rho}_t^{\boldsymbol{\tau}} = \rho_k^{\boldsymbol{\tau}} \quad \text{for } t \in (t_{k\text{-}1}^{\boldsymbol{\tau}}, t_k^{\boldsymbol{\tau}}] \text{ and } k \in \mathbb{N}.$$

Strategy of the Proof. The aim of this section is to apply the variational formulation of time-dependent implicit Euler method to the non-autonomous and non-linear Fokker-Planck equation (3.2.1) and to strengthen the convergence results of the previous section 3.1. It is clear, that under mild assumptions on V_t and W_t, the functional $\mathcal{E}_t$ falls into the class of feasible free energy functionals of the previous section 3.1 and we recover the results therein. However, additional supplementary structural properties are derived, to mention in particular the time-discrete Euler-Lagrange equations (3.2.5) and the refined $\mathrm{BV}(\Omega)$-estimates on $(\rho_k^{\boldsymbol{\tau}})^m$ (3.2.10), see section 3.2.2. The convergence of the approximation $\overline{\rho}_t^{\boldsymbol{\tau}}$ is proven in section 3.2.3 by means of the extension of Aubin-Lions Compactness Theorem 2.5.4.

3.2.1 Setup and Assumptions

Throughout the rest of this section, $\Omega \subseteq \mathbb{R}^d$ is either equal to $\mathbb{R}^d$ or is equal to some open, bounded and connected domain with Lipschitz-continuous boundary $\partial\Omega$. In this case, the assumptions on the confinement potential V_t and the interaction potential W_t read as follows:

Assumption 3.2.1 (Regularity Assumptions on V_t and W_t)**.** Let the confinement potential $V_t \in \mathcal{C}^1([0,\infty) \times \Omega)$ and the symmetric interaction kernel $W_t \in \mathcal{C}^1([0,\infty) \times \mathbb{R}^d)$ be such that

(F1) There exists a non-negative constant d_1 such that

$$|V_t(x)|,\ |W_t(x)|,\ |\nabla V_t(x)|,\ |\nabla W_t(x)|,\ |\partial_t V_t(x)|,\ |\partial_t W_t(x)| \leq d_1(1 + \|x^2\|).$$

(F2) There exists a non-negative function $\alpha_t \in \mathrm{L}^1_{loc}(0,\infty)$ such that

$$|V_t(x) - V_s(x)|\,,\ |W_t(x) - W_s(x)| \leq (1 + \|x\|^2) \int_s^t \alpha_t \,\mathrm{d}t.$$

(F3) There exists a non-negative function $\widetilde{\alpha}_t \in \mathrm{L}^1_{loc}(0,\infty)$ such that

$$|\nabla V_t(x) - \nabla V_s(x)|\,,\ |\nabla W_t(x) - \nabla W_s(x)| \leq (1 + \|x\|^2) \int_s^t \widetilde{\alpha}_t \,\mathrm{d}t.$$

Remark 3.2.2. The regularity assumptions on V_t and W_t and the bounds (F1) guarantee, that $\mathcal{F}_t$ satisfies the *LSCC*-assumption 3.1.2. Assumption (E4) follows from (F2) and (E5) follows from the uniform bounds on the time-derivatives $\partial_t V_t$ and $\partial_t W_t$. Lastly, the condition (F3) is necessary to perform the discrete-to-continuous limit in the Euler-Lagrange equations. Hence, the existence of the discrete solution $(\rho_k^{\boldsymbol{\tau}})_{k\in\mathbb{N}}$ and the classical estimates are also valid in this case.

3.2.2 Properties of the Time-Dependent MMS

Given a partition $\boldsymbol{\tau} = (\tau_1, \tau_2, \ldots)$ with time step sizes $\tau_n \in (0, \tau_*)$ and a pair of initial data $(\rho_0^{\boldsymbol{\tau}}, \rho_1^{\boldsymbol{\tau}})$ which approximates the initial datum ρ_0, which satisfy (I1)&(I2). Then, the *discrete solution* $(\rho_k^{\boldsymbol{\tau}})_{k\in\mathbb{N}}$ for $\mathcal{E}$ on $(\mathcal{P}_2(\Omega), \mathbf{W}_2)$ defined in (3.1.2) and equivalently defined by the recursive formula

$$\rho_{k+1}^{\boldsymbol{\tau}} \in \underset{\rho\in\mathcal{P}_2(\Omega)}{\operatorname{argmin}}\ \Phi(\tau, t_k^{\boldsymbol{\tau}}, \rho_k^{\boldsymbol{\tau}}; \rho) \quad \text{for } k \in \mathbb{N}$$

is well-posed by theorem 3.1.9, since the energy $\mathcal{E}_t$ satisfies the abstract assumptions (E1)–(E5) due to (F1)–(F3). The rest of this section is devoted to deriving structural properties of the Minimizing Movement scheme, namely: *Step size independent estimates*, *discrete Euler-Lagrange equations*, *better a priori estimates*.

Step Size Independent Estimates. The following two estimate is a specialization of the classical estimate (3.1.13).

Lemma 3.2.3 (Classical Estimates II). *For fixed $T > 0$, there exists a constant C, independent of the partition $\boldsymbol{\tau}$, such that the corresponding discrete solutions $(\rho_k^{\boldsymbol{\tau}})_{k\in\mathbb{N}}$ satisfies for all k with $t_k^{\boldsymbol{\tau}} < T$:*

$$\boldsymbol{M}_2(\rho_k^{\boldsymbol{\tau}}) \le C(T, \tau_*, \rho_0), \qquad \mathcal{U}_m(\rho_k^{\boldsymbol{\tau}}) \le C(T, \tau_*, \rho_0). \tag{3.2.4}$$

Proof. The first bound of (3.2.4) follows from the fact that one can estimate the second moment in terms of the Wasserstein distance, i.e.,

$$\boldsymbol{M}_2(\rho_k^{\boldsymbol{\tau}}) \le 2\mathbf{W}_2^2(\rho_k^{\boldsymbol{\tau}}, \mu_*) + 2\boldsymbol{M}_2(\mu_*).$$

The first term is bounded by the classical estimate (3.1.13) and hence we established the first bound (3.2.4). The second bound follows than straight forward by the growth bounds (F1) of V_t and W_t, i.e.,

$$\mathcal{U}_m(\rho_k^{\boldsymbol{\tau}}) \le \mathcal{E}_t(\rho_k^{\boldsymbol{\tau}}) - \int_\Omega V_t\rho_k^{\boldsymbol{\tau}} + (W_t * \rho_k^{\boldsymbol{\tau}})\rho_k^{\boldsymbol{\tau}}\,\mathrm{d}x \le \mathcal{E}(\rho_k^{\boldsymbol{\tau}}) + C(1 + \boldsymbol{M}_2(\rho_k^{\boldsymbol{\tau}})).$$

Using the first result and the classical estimate (3.1.13) yields the desired bound. □

Discrete Euler-Lagrange Equation. In the next theorem, we derive approximate Euler-Lagrange equations for the weak formulation of the non-autonomous and non-linear Fokker-Planck equation (3.2.1). The key idea is the JKO-method introduced in [54] and recalled in section 2.4.

Theorem 3.2.4 (Discrete Euler-Lagrange Equations). *The discrete solution $(\rho_k^{\boldsymbol{\tau}})_{k\in\mathbb{N}}$ obtained by the time-dependent Minimizing Movement scheme* (3.1.2) *satisfies for each $k \in \mathbb{N}$ and for all vector fields $\xi \in C_c^\infty(\overline{\Omega}, \mathbb{R}^{\mathrm{d}})$ with $\xi \cdot \boldsymbol{n} = 0$ on $\partial\Omega$:*

$$\begin{aligned} 0 =& \frac{1}{\tau_k}\int_{\Omega^2} \langle \xi(x), x - y\rangle\,\mathrm{d}\boldsymbol{p}_k^{\boldsymbol{\tau}}(x,y) - \int_\Omega \mathrm{div}(\xi)\,(\rho_k^{\boldsymbol{\tau}})^m\,\mathrm{d}x \\ &+ \int_\Omega \langle \xi, \nabla V_{t_k^{\boldsymbol{\tau}}}\rangle\,\rho_k^{\boldsymbol{\tau}} + \langle \xi, \nabla W_{t_k^{\boldsymbol{\tau}}} * \rho_k^{\boldsymbol{\tau}}\rangle\,\rho_k^{\boldsymbol{\tau}}\,\mathrm{d}x, \end{aligned} \tag{3.2.5}$$

where $\boldsymbol{p}_k^{\boldsymbol{\tau}} \in \Gamma(\rho_k^{\boldsymbol{\tau}}, \rho_{k\text{-}1}^{\boldsymbol{\tau}})$ is the optimal transport plans.

Proof. Fix $\rho_k^{\boldsymbol{\tau}}, \rho_{k\text{-}1}^{\boldsymbol{\tau}}$ and $\xi \in C_c^\infty(\overline{\Omega}, \mathbb{R}^{\mathrm{d}})$ with $\xi\cdot\boldsymbol{n} = 0$ on $\partial\Omega$. We consider the perturbation ρ^s of $\rho_k^{\boldsymbol{\tau}}$ as the solution of the Transport equation with velocity field ξ starting at $\rho_k^{\boldsymbol{\tau}}$, i.e., ρ^s is the solution of (2.4.6) as in section 2.4. The first variation of the energy $\mathcal{E}_{t_k^{\boldsymbol{\tau}}}$ along the solution to the Transport equation amounts to

$$\frac{\mathrm{d}}{\mathrm{d}s}\left[\mathcal{E}_{t_k^{\boldsymbol{\tau}}}(\rho^s)\right]_{s=0} = \int_\Omega -\mathrm{div}(\xi)\,(\rho_k^{\boldsymbol{\tau}})^m + \langle \xi, \nabla V_{t_k^{\boldsymbol{\tau}}}\rangle\,\rho_k^{\boldsymbol{\tau}} + \langle \xi, \nabla W_{t_k^{\boldsymbol{\tau}}} * \rho_k^{\boldsymbol{\tau}}\rangle\,\rho_k^{\boldsymbol{\tau}}\,\mathrm{d}x. \tag{3.2.6}$$

The differentiability of the quadratic L^2-Wasserstein distance $\mathbf{W}_2$ along the solution ρ^s of the Transport equation is more technical, for the proof we refer to [4, 92]. Since $\rho^{\boldsymbol{\tau}}_{k\text{-}1}, \rho^{\boldsymbol{\tau}}_k, \rho^s$ are all absolutely continuous measures, Theorem 8.13 from [92] is applicable and we can conclude:

$$\frac{\mathrm{d}}{\mathrm{d}s}\Big[\frac{1}{2\tau_k}\mathbf{W}_2^2(\rho^{\boldsymbol{\tau}}_{k\text{-}1}, \rho^s)\Big]_{s=0} = \frac{1}{\tau_k}\int_{\Omega^2}\langle \xi(x), x-y\rangle \,\mathrm{d}\boldsymbol{p}^{\boldsymbol{\tau}}_k(x,y), \tag{3.2.7}$$

where $\boldsymbol{p}^k_{\boldsymbol{\tau}} \in \Gamma(\rho^{\boldsymbol{\tau}}_k, \rho^{\boldsymbol{\tau}}_{k\text{-}1})$ is the optimal transport plan. Since $\rho^{\boldsymbol{\tau}}_k$ is a minimizer of the time-dependent Moreau-Yosida penalization $\Phi(\tau, t^{\boldsymbol{\tau}}_k, \rho^{\boldsymbol{\tau}}_{k\text{-}1};\,\cdot)$ and since $s \mapsto \Phi(\tau, t^{\boldsymbol{\tau}}_k, \rho^{\boldsymbol{\tau}}_{k\text{-}1}; \rho^s)$ is differentiable at $s = 0$,

$$\begin{aligned} 0 = \frac{\mathrm{d}}{\mathrm{d}s}\left[\Phi(\tau, t^{\boldsymbol{\tau}}_k, \rho^{\boldsymbol{\tau}}_{k\text{-}1}; \rho^s)\right]_{s=0} =& \frac{1}{\tau_k}\int_{\Omega^2}\langle \xi(x), x-y\rangle \,\mathrm{d}\boldsymbol{p}^{\boldsymbol{\tau}}_k(x,y) - \int_\Omega \mathrm{div}(\xi)\,(\rho^{\boldsymbol{\tau}}_k)^m \,\mathrm{d}x \\ &+ \int_\Omega \langle \xi, \nabla V_{t^{\boldsymbol{\tau}}_k}\rangle\, \rho^{\boldsymbol{\tau}}_k + \langle \xi, \nabla W_{t^{\boldsymbol{\tau}}_k} * \rho^{\boldsymbol{\tau}}_k\rangle\, \rho^{\boldsymbol{\tau}}_k \,\mathrm{d}x. \end{aligned}$$

Indeed, we have the desired equality (3.2.5). □

Refined Regularity. The already obtained regularity results for the interpolated solution $\overline{\rho}^{\boldsymbol{\tau}}_t$ are not sufficient to pass to the limit in the first term of the discrete Euler-Lagrange equation (3.2.5). Nevertheless, the following bounds in the $\mathrm{BV}(\Omega)$-norm of $(\rho^{\boldsymbol{\tau}}_k)^m$ are sufficient to obtain the desired regularity results. These estimates can be derived from the discrete Euler-Lagrange equation quite naturally.

Proposition 3.2.5 (Step Size Independent Local $\mathrm{BV}(\Omega)$-estimate)**.** *Fix a time horizon $T > 0$. There exists a constant C, independent of the partition $\boldsymbol{\tau}$, such that the corresponding discrete solutions $(\rho^{\boldsymbol{\tau}}_k)_{k\in\mathbb{N}}$ satisfy for all $k \in \mathbb{N}$ with $t^{\boldsymbol{\tau}}_k \le T$:*

$$\|(\rho^{\boldsymbol{\tau}}_k)^m\|_{\mathrm{BV}(\Omega)} \le C\Big(1 + \frac{\mathbf{W}_2(\rho^{\boldsymbol{\tau}}_k, \rho^{\boldsymbol{\tau}}_{k\text{-}1})}{\tau_k}\Big). \tag{3.2.8}$$

Proof. The $\mathrm{L}^1(\Omega)$-norm of $(\rho^{\boldsymbol{\tau}}_k)^m$ is equal to $(m-1)\mathcal{U}_m$ evaluated at $\rho^{\boldsymbol{\tau}}_k$. Hence, we can bound the first term in the definition of the $\mathrm{BV}(\Omega)$-norm uniformly by (3.2.4). In order to estimate the variation of $(\rho^{\boldsymbol{\tau}}_k)^m$, we estimate the term inside the supremum of the definition of $V((\rho^{\boldsymbol{\tau}}_k)^m, \Omega)$. Thus let $\xi \in C^\infty_c(\Omega, \mathbb{R}^{\mathrm{d}})$ with $\|\xi\|_\infty \le 1$, then we can use the discrete Euler-Lagrange equations (3.2.5) to substitute

$$\begin{aligned} \int_\Omega (\rho^{\boldsymbol{\tau}}_k)^m \,\mathrm{div}(\xi)\,\mathrm{d}x =& \int_\Omega \langle \xi, \nabla V_{t^{\boldsymbol{\tau}}_k}\rangle \rho^{\boldsymbol{\tau}}_k + \langle \xi, \nabla W_{t^{\boldsymbol{\tau}}_k} * \rho^{\boldsymbol{\tau}}_k\rangle \rho^{\boldsymbol{\tau}}_k \,\mathrm{d}x \\ &+ \frac{2}{\tau_k}\int_{\Omega^2}\langle \xi(x), x-y\rangle \,\mathrm{d}\boldsymbol{p}^{\boldsymbol{\tau}}_k(x,y). \end{aligned} \tag{3.2.9}$$

By (F1) we have quadratic growth bounds for ∇V_t and ∇W_t, so using the step size independent bounds on the second moment (3.2.4), we can estimate the first terms in

(3.2.9) as follows:

$$\int_\Omega \langle \xi, \nabla V_{t_k^{\boldsymbol\tau}} + \nabla W_{t_k^{\boldsymbol\tau}} * \rho_k^{\boldsymbol\tau}\rangle \rho_k^{\boldsymbol\tau}\,\mathrm{d}x \le 2d_1 \left\|\xi\right\|_\infty (1 + \boldsymbol{M}_2(\rho_k^{\boldsymbol\tau})) \le 2d_1(1+C).$$

The second integral on the right-hand side of (3.2.9) can be estimated using Jensen's inequality

$$\Big|\int_{\Omega^2} \langle \xi(x), x-y\rangle \;\mathrm{d}\boldsymbol{p}_{\boldsymbol\tau}^k(x,y)\Big| \le \|\xi\|_\infty \Big(\int_{\Omega^2} \|x-y\|^2 \;\mathrm{d}\boldsymbol{p}_{\boldsymbol\tau}^k(x,y)\Big)^{1/2} \le \mathbf{W}_2(\rho_k^{\boldsymbol\tau}, \rho_{k\text{-}1}^{\boldsymbol\tau}).$$

Hence, we have the following upper bound for the variation of $(\rho_k^{\boldsymbol\tau})^m$:

$$V((\rho_k^{\boldsymbol\tau})^m, \Omega) \le C\Big(1 + \frac{\mathbf{W}_2(\rho_k^{\boldsymbol\tau}, \rho_{k\text{-}1}^{\boldsymbol\tau})}{\tau_k}\Big).$$

In conclusion, the discrete solution $(\rho_k^{\boldsymbol\tau})_{k\in\mathbb{N}}$ satisfies the desired bound (3.2.8). ☐

Theorem 3.2.6 (Step Size Independent Global $\mathrm{L}^2(0,T;\mathrm{BV}(\Omega))$-estimate)**.** *Fix a time horizon $T>0$. There exists a constant C, independent of the partition $\boldsymbol\tau$, such that the corresponding interpolated solution $\overline{\rho}_t^{\boldsymbol\tau}$ satisfies:*

$$\|(\overline{\rho}_t^{\boldsymbol\tau})^m\|_{\mathrm{L}^2(0,T;\mathrm{BV}(\Omega))} \le C. \tag{3.2.10}$$

Proof. We use the classical estimates on the kinetic energy (3.1.13) and the result from Proposition 3.2.5 to estimate the $\mathrm{L}^2(0,T;\mathrm{BV}(\Omega))$-norm of $(\overline{\rho}_t^{\boldsymbol\tau})^m$. Define as usual $N_T := \max\{k\in\mathbb{N} \mid t_k^{\boldsymbol\tau} \le T\}$, then we have

$$\|(\overline{\rho}_t^{\boldsymbol\tau})^m\|^2_{\mathrm{L}^2(0,T;\mathrm{BV}(\Omega))} \le \sum_{k=1}^{N_T+1} \int_{t_{k\text{-}1}^{\boldsymbol\tau}}^{t_k^{\boldsymbol\tau}} \|(\rho_k^{\boldsymbol\tau})^m\|^2_{\mathrm{BV}(\Omega)}\;\mathrm{d}t \le C\sum_{k=1}^{N_T+1} \tau_k\Big(1 + \frac{\mathbf{W}_2(\rho_k^{\boldsymbol\tau}, \rho_{k\text{-}1}^{\boldsymbol\tau})}{\tau_k}\Big)^2.$$

By the a Cauchy type inequality we obtain

$$\begin{aligned}\|(\overline{\rho}_t^{\boldsymbol\tau})^m\|^2_{\mathrm{L}^2(0,T;\mathrm{BV}(\Omega))} &\le C\sum_{k=1}^{N_T+1}\Big[\tau_k + \frac{\mathbf{W}_2^2(\rho_k^{\boldsymbol\tau}, \rho_{k\text{-}1}^{\boldsymbol\tau})}{\tau_k}\Big]\\ &\le C(T+\tau_*) + C\sum_{k=0}^{N_T+1}\frac{\mathbf{W}_2^2(\rho_k^{\boldsymbol\tau}, \rho_{k\text{-}1}^{\boldsymbol\tau})}{\tau_k}.\end{aligned}$$

Finally, we can conclude with the step size independent bounds on the kinetic energy (3.1.13) the desired estimate (3.2.10) for some universal constant C, independent of the partition $\boldsymbol\tau$. ☐

3.2.3 Convergence

In this section, we prove our main theorem concerning the strong convergence of the approximation $\overline{\rho}_t^{\boldsymbol{\tau}}$ to the solution of the non-linear Fokker-Planck equation (3.2.1). The convergence in the strong $\mathrm{L}^p(0,T;\mathrm{L}^m(\Omega))$-topology for any $p \in [1,\infty)$ follows by the improved $\mathrm{L}^2(0,T;\mathrm{BV}(\Omega))$-estimates (3.2.10) and by the general version of the Aubin-Lions Theorem 2.5.4, c.f. [85, Theorem 2].

Theorem 3.2.7 (Strong Convergence in $\mathrm{L}^p(0,T;\mathrm{L}^m(\Omega))$). *There exists a further (non-relabelled) subsequence $(\boldsymbol{\tau}_n)_{n\in\mathbb{N}}$ such that for all $T > 0$ and any $p \in [1,\infty)$ and any bounded $\Theta \subseteq \Omega$:*

$$\overline{\rho}_t^{\boldsymbol{\tau}_n} \to \rho_t^* \qquad \text{strongly in } \mathrm{L}^p(0,T;\mathrm{L}^m(\Theta)) \text{ as } n \to \infty.$$

Proof of Theorem 3.2.7 for $\Omega \subsetneq \mathbb{R}^{\mathrm{d}}$. Fix $T > 0$. In order to prove the strong convergence result we use the Aubin-Lions Theorem 2.5.4 with the underlying Banach space $\mathbf{X} = \mathrm{L}^m(\Omega)$. We consider the functional $\mathcal{A} : \mathrm{L}^m(\Omega) \to \mathbb{R}$, defined via

$$\mathcal{A}(\rho) := \begin{cases} \|\rho^m\|^2_{\mathrm{BV}(\Omega)} & \text{if } \rho \in \mathcal{P}_2(\Omega) \text{ and } \rho^m \in \mathrm{BV}(\Omega), \\ +\infty & \text{else.} \end{cases}$$

Using the lemma 2.5.2 in the introductory section it follows that the functional $\mathcal{A}$ is measurable, lower semi-continuous with respect to the $\mathrm{L}^m(\Omega)$-topology, and has compact sublevels. Next, we choose as pseudo-distance $g = \mathbf{W}_2$ on $\mathrm{L}^m(\Omega)$, i.e.,

$$g(f,h) := \begin{cases} \mathbf{W}_2(f\mathcal{L}^d(\Omega), h\mathcal{L}^d(\Omega)) & \text{if } f,h \in \mathrm{L}^m(\Omega), \\ +\infty & \text{else.} \end{cases}$$

Note, we have $g(\rho,\nu) = \mathbf{W}_2(\rho,\nu)$ for absolutely continuous measures $\rho,\nu \in \mathcal{P}_2(\Omega) \cap \mathrm{L}^m(\Omega)$. The L^2-Wasserstein distance is lower semi-continuous with respect to the $\mathrm{L}^m(\Omega)$-topology and clearly compatible with $\mathcal{A}$, see lemma 2.5.2.

Next, we verify the assumption (2.5.1) on $(\overline{\rho}_t^{\boldsymbol{\tau}_n})_{n\in\mathbb{N}}$ of Theorem 2.5.4. By the refined $\mathrm{L}^2(0,T;\mathrm{BV}(\Omega))$-estimates of Theorem 3.2.6 it is clear, that the sequence $(\overline{\rho}_t^{\boldsymbol{\tau}_n})_{n\in\mathbb{N}}$ is tight with respect to $\mathcal{A}$, since we have:

$$\sup_{n\in\mathbb{N}} \int_0^T \|(\overline{\rho}_t^{\boldsymbol{\tau}_n})^m\|^2_{\mathrm{BV}(\Omega)} \,\mathrm{d}t = \sup_{n\in\mathbb{N}} \|(\overline{\rho}_t^{\boldsymbol{\tau}_n})^m\|^2_{\mathrm{L}^2(0,T;\mathrm{BV}(\Omega))} \le C < \infty.$$

For the proof of the relaxed averaged weak integral equicontinuity condition of $(\overline{\rho}_t^{\boldsymbol{\tau}_n})_{n\in\mathbb{N}}$ with respect to $\mathbf{W}_2$, we use the auxiliary inequality (3.1.17) to obtain

$$\limsup_{n\to\infty} \mathbf{W}_2(\overline{\rho}_{s+t}^{\boldsymbol{\tau}_n}, \overline{\rho}_s^{\boldsymbol{\tau}_n}) \le \int_s^{s+t} A_r \,\mathrm{d}r.$$

Indeed, Fatou's Lemma and Fubini's Theorem yields

$$\liminf_{h\searrow 0}\limsup_{n\to\infty}\frac{1}{h}\int_0^h\int_0^{T-t}\mathbf{W}_2(\overline{\rho}_{s+t}^{\boldsymbol{\tau}_n},\overline{\rho}_s^{\boldsymbol{\tau}_n})\,\mathrm{d}s\,\mathrm{d}t\le\liminf_{h\searrow 0}h\int_0^T A_t\,\mathrm{d}t=0.$$

Therefore, we can conclude that there exists a (non-relabeled) subsequence such that $(\overline{\rho}_t^{\boldsymbol{\tau}_n})_{n\in\mathbb{N}}$ converges in $\mathcal{M}(0,T;\mathrm{L}^m(\Omega))$ to some curve ρ_t^+. Due to the uniform bounds in $\mathrm{L}^\infty(0,T;\mathrm{L}^m(\Omega))$, the sequence $\overline{\rho}_t^{\boldsymbol{\tau}_n}$ is also uniformly bounded in $\mathrm{L}^p(0,T;\mathrm{L}^m(\Omega))$ and we get the desired convergence result with Remark 2.1.1. Moreover, the limit curves ρ_t^+ and ρ_t^* have to coincide, since $\overline{\rho}_t^{\boldsymbol{\tau}_n}$ converges also in measure on Ω to ρ_t^+ and ρ_t^*, so both limits have to be equal. □

In the case of $\Omega=\mathbb{R}^\mathrm{d}$, we have to alter the proof given above since the embedding of $\mathrm{BV}(\mathbb{R}^\mathrm{d})$ into $\mathrm{L}^1(\mathbb{R}^\mathrm{d})$ is not compact anymore. So we restrict ourself to the compact domains $\Theta=\mathbb{B}_R(0)$. The set $\overline{\Theta}$ is clearly compact with Lipschitz-continuous boundary $\partial\overline{\Theta}$, so the embedding of $BV(\Theta)$ into $L^1(\Theta)$ is compact again.

Proof of Theorem 3.2.7 for $\Omega=\mathbb{R}^\mathrm{d}$. Fix $T>0$. Without loss of generality we can assume $\Theta=\mathbb{B}_R(0)$, since every compact subset $\widetilde{\Theta}\Subset\mathbb{R}^\mathrm{d}$ is contained in the closure of a ball with radius R and convergence in $\mathrm{L}^m(0,T;\mathrm{L}^m(\mathbb{B}_R(0)))$ implies convergence in $\mathrm{L}^m(0,T;\mathrm{L}^m(\widetilde{\Theta}))$.

As before, we want to use the Aubin-Lions Theorem 2.5.4 for the Banach space $\mathrm{L}^m(\Theta)$ equipped with the natural topology induced by the $\mathrm{L}^m(\Theta)$-norm applied to $(\overline{\rho}_t^{\boldsymbol{\tau}_n}|_\Theta)_{n\in\mathbb{N}}$, the restriction of the density $\overline{\rho}_t^{\boldsymbol{\tau}_n}$ to the subspace Θ. In this case we consider the functional $\widetilde{\mathcal{A}}:\mathrm{L}^m(\Theta)\to\mathbb{R}$, defined via

$$\widetilde{\mathcal{A}}(\rho):=\begin{cases}\|\rho^m\|^2_{\mathrm{BV}(\Theta)} & \text{if } \rho\in\mathcal{M}_f(\Theta) \text{ and } \rho^m\in\mathrm{BV}(\Theta),\\ +\infty & \text{else.}\end{cases}$$

Now, the functional $\widetilde{\mathcal{A}}$ is measurable, lower semi-continuous with respect to the $\mathrm{L}^m(\Theta)$ topology, and has compact sublevels. Since $\widetilde{\mathcal{A}}(\rho|_\Theta)\le\mathcal{A}(\rho)$, we obtain by the same calculations as above the tightness of $(\overline{\rho}_t^{\boldsymbol{\tau}_n}|_\Theta)_{n\in\mathbb{N}}$ with respect to $\widetilde{\mathcal{A}}$.

Since the measure $\rho|_\Theta$ does not have unit mass anymore, we cannot consider the L^2-Wasserstein distance $\mathbf{W}_2$ as pseudo-distance anymore. However, we can use the following pseudo-distance $\widetilde{g}$:

$$\widetilde{g}(\rho,\nu):=\inf\left\{\mathbf{W}_2(\widetilde{\rho},\widetilde{\nu})\mid\widetilde{\rho}\in\Sigma(\rho),\ \widetilde{\nu}\in\Sigma(\nu)\right\},$$
$$\Sigma(\rho):=\left\{\widetilde{\rho}\in\mathscr{P}(\mathbb{R}^\mathrm{d})\mid\widetilde{\rho}|_\Theta=\rho\mathcal{L}^d,\ \boldsymbol{M}_2(\widetilde{\rho})\le C\right\},$$

where C is the constant from the classical estimates (3.2.4) for the specific T. Since $\Sigma(\rho)$ and $\Sigma(\nu)$ are compact sets with respect to the narrow topology, the infimum is attained at some pair $\widetilde{\rho}_*,\widetilde{\nu}_*$. The pseudo-distance $\widetilde{g}$ is compatible with $\widetilde{\mathcal{A}}$, i.e., if $\rho^m,\nu^m\in\mathrm{BV}(\Theta)$ and $\widetilde{g}(\rho,\nu)=0$ then $\rho=\nu$ a.e. on Θ. The lower semi-continuity of the pseudo-distance

$\widetilde{g}$ with respect to the $\mathrm{L}^m(\Theta)$-topology can be proven as follows. Choose to convergent sequences $\rho_n \to \rho$ and $\nu_n \to \nu$ in $\mathrm{L}^m(\Theta)$ with $\sup_n \widetilde{g}(\rho_n, \nu_n) < \infty$. By the remark from above, there exists $\widetilde{\rho}_n, \widetilde{\nu}_n$ such that $\widetilde{g}(\rho_n, \nu_n) = \mathbf{W}_2(\widetilde{\rho}_n, \widetilde{\nu}_n)$. Since the second moments are by definition of $\Sigma(\rho)$ uniformly bounded, we can extract a non-relabeled convergent subsequence which converges narrowly to $\widetilde{\rho} \in \Sigma(\rho), \widetilde{\nu} \in \Sigma(\nu)$. By the lower semi-continuity of $\mathbf{W}_2$ with respect to narrow convergence, we get in the end

$$\widetilde{g}(\rho, \nu) \leq \mathbf{W}_2(\widetilde{\rho}, \widetilde{\nu}) \leq \liminf_{n\to\infty} \mathbf{W}_2(\widetilde{\rho}_n, \widetilde{\nu}_n) = \liminf_{n\to\infty} \mathbf{W}_2(\rho_n, \nu_n).$$

Therefore, the pseudo-distance $\widetilde{g}$ is lower semi-continuous with respect to the $\mathrm{L}^m(\Theta)$-topology. Thus, $\widetilde{g}$ satisfies the assumptions of theorem 2.5.4. Further, by definition one has $\widetilde{g}(\rho|_\Theta, \nu|_\Theta) \leq \mathbf{W}_2(\rho, \nu)$. Thus we derive, using the same proof as above, the equicontinuity of $(\overline{\rho}_t^{\boldsymbol{\tau}_n}|_\Theta)_{n\in\mathbb{N}}$ with respect to the pseudo-distance $\widetilde{g}$.

Hence, we can conclude that there exists a non-relabeled subsequence of $\overline{\rho}_t^{\boldsymbol{\tau}_n}|_\Theta$ which converges in $\mathcal{M}(0, T; \mathrm{L}^m(\Theta))$ to some limit ρ_t^+. As before, we use the uniform bounds in $\mathrm{L}^\infty(0, T; \mathrm{L}^m(\Theta))$, to obtain the strong convergence in $\mathrm{L}^p(0, T; \mathrm{L}^m(\Theta))$. Moreover, the limit curves ρ_t^+ and $\rho_t^*|_\Theta$ have to coincide on Θ, since $\overline{\rho}_t^{\boldsymbol{\tau}_n}|_{\overline{\Theta}}$ converges also in measure on Θ to ρ_t^+ and $\rho_t^*|_\Theta$, so both limits have to be equal on Θ. Two diagonal arguments in $T \to \infty$ and $R \to \infty$ yield the desired convergence result. □

To complete the proof of the main theorem 3.0.2, we have to validate that ρ_t^* is indeed a solution to (3.2.1) in the sense of distributions.

Theorem 3.2.8 (Solution of the Non-autonomous and Non-linear Fokker-Planck Equation)**.** *The limit curve ρ_t^* of theorem 3.2.7 is a solution to the non-autonomous and non-linear Fokker-Planck equation with no-flux boundary condition* (3.2.1) *in the weak sense of* (3.2.11).

Proof. Fix $\varphi_t \in C_c^\infty([0,\infty) \times \overline{\Omega})$ with $\nabla\varphi_t \cdot \boldsymbol{n} = 0$ on $\partial\Omega$ and let be $T > 0$ and $\Theta \Subset \Omega$ be compact such that $\operatorname{supp} \varphi_t \subset [0, T] \times \Theta$. Further, define $N_T^n := \{k \mid t_k^{\boldsymbol{\tau}_n} \leq T\}$ and the piecewise constant interpolation $\overline{\varphi}_t^{\boldsymbol{\tau}_n}$ of φ_t by

$$\overline{\varphi}_0^{\boldsymbol{\tau}_n} = \varphi_0, \quad \overline{\varphi}_t^{\boldsymbol{\tau}_n} = \varphi_{t_k^{\boldsymbol{\tau}_n}} \quad \text{for } t \in (t_{k\text{-}1}^{\boldsymbol{\tau}_n}, t_k^{\boldsymbol{\tau}_n}] \quad \text{and } k \in \mathbb{N}.$$

Similarly, we define $\overline{V}_t^{\boldsymbol{\tau}_n}$ and $\overline{W}_t^{\boldsymbol{\tau}_n}$ as the piecewise constant interpolation of V_t and W_t, respectively.

For each $k \in \mathbb{N}$ insert the smooth function $\varphi_{t_k^{\boldsymbol{\tau}_n}}$ in the discrete Euler-Lagrange equation (3.2.5) for the vector field $\xi \in \mathcal{C}_c^\infty(\Theta)$. Summing the resulting equations from $k = 1$ to $N_T^n + 1$ and multiplying with $\tau_{k,n}$ yields:

$$\begin{aligned} 0 = &\int_0^T \int_\Theta -\Delta(\overline{\varphi}_t^{\boldsymbol{\tau}_n})\, (\overline{\rho}_t^{\boldsymbol{\tau}_n})^m + \langle \nabla(\overline{\varphi}_t^{\boldsymbol{\tau}_n}), \nabla\overline{V}_t^{\boldsymbol{\tau}_n} \rangle\, \overline{\rho}_t^{\boldsymbol{\tau}_n} + \langle \nabla(\overline{\varphi}_t^{\boldsymbol{\tau}_n}), \nabla\overline{W}_t^{\boldsymbol{\tau}_n} * \overline{\rho}_t^{\boldsymbol{\tau}_n} \rangle\, \overline{\rho}_t^{\boldsymbol{\tau}_n} \,\mathrm{d}x\,\mathrm{d}t \\ &+ \sum_{k=1}^{N_T^n} \int_{\Theta^2} \langle \nabla\varphi_{t_k^{\boldsymbol{\tau}_n}}(x), x - y \rangle\, \mathrm{d}\boldsymbol{p}_k^{\boldsymbol{\tau}_n}(x, y) \qquad =: I_1 + I_2 + I_3 + I_4. \end{aligned}$$

Due to the strong convergence in $\mathrm{L}^m(0,T;\mathrm{L}^m(\Theta))$ of $\overline{\rho}_t^{\boldsymbol{\tau}_n}$ to ρ_* and due to the uniform convergence of $\Delta\overline{\varphi}_t^{\boldsymbol{\tau}_n}$ to $\Delta\varphi_t$

$$\lim_{n\to\infty} I_1 = \int_0^T \int_\Theta -\Delta\varphi_t\,(\rho_t^*)^m \,\mathrm{d}x\,\mathrm{d}t.$$

The second and third integrals $I_2\&I_3$ converge to

$$\lim_{n\to\infty} I_2 + I_3 = \int_0^T \int_\Theta \langle\nabla(\overline{\varphi}_t^{\boldsymbol{\tau}_n}), \nabla\overline{V}_t^{\boldsymbol{\tau}_n}\rangle\,\overline{\rho}_t^{\boldsymbol{\tau}_n} + \langle\nabla(\overline{\varphi}_t^{\boldsymbol{\tau}_n}), \nabla\overline{W}_t^{\boldsymbol{\tau}_n} * \overline{\rho}_t^{\boldsymbol{\tau}_n}\rangle\,\overline{\rho}_t^{\boldsymbol{\tau}_n}\,\mathrm{d}x\,\mathrm{d}t$$

thanks to the convergence of $\overline{\rho}_t^{\boldsymbol{\tau}_n}$ to ρ_t^* in the strong $\mathrm{L}^1(0,T;\mathrm{L}^m(\Theta))$-topology, due to (F3) and (3.2.4) (and therefore the uniform convergence of $\nabla\overline{V}_t^{\boldsymbol{\tau}_n}$ to ∇V_t and $\nabla\overline{W}_t^{\boldsymbol{\tau}_n}$ to ∇W_t, respectively). In order to calculate the limit of I_4, we can expand by Taylor's formula the integrand as follows

$$\begin{aligned}
I_4 &= \sum_{k=1}^{N_T^n} \int_{\Theta^2} \langle\nabla\varphi_{t_k^{\boldsymbol{\tau}_n}}(x), x-y\rangle\,\mathrm{d}\boldsymbol{p}_k^{\boldsymbol{\tau}_n}(x,y) \\
&= \sum_{k=1}^{N_T^n} \int_{\Theta^2} \varphi_{t_k^{\boldsymbol{\tau}_n}}(x) - \varphi_{t_k^{\boldsymbol{\tau}_n}}(y) + \mathcal{O}(\|x-y\|^2)\,\mathrm{d}\boldsymbol{p}_k^{\boldsymbol{\tau}_n}(x,y) \\
&= \sum_{k=1}^{N_T^n} \int_\Theta \left(\rho_k^{\boldsymbol{\tau}_n}(x) - \rho_{k\text{-}1}^{\boldsymbol{\tau}_n}(x)\right)\varphi_{t_k}^{\boldsymbol{\tau}_n}(x)\,\mathrm{d}x + \sum_{k=1}^{N_T^n} \mathcal{O}(\mathbf{W}_2^2(\rho_k^{\boldsymbol{\tau}_n}, \rho_{k\text{-}1}^{\boldsymbol{\tau}_n})).
\end{aligned}$$

Rearrange the first term and use (3.1.13) to bound the second term, to obtain

$$I_4 = -\int_0^T \int_\Theta \partial_t\varphi_{t+\tau_{k,n}}\overline{\rho}_t^{\boldsymbol{\tau}_n}\,\mathrm{d}x\,\mathrm{d}t - \int_\Theta \varphi_0\rho_0^{\boldsymbol{\tau}_n}\,\mathrm{d}x + \mathcal{O}(\sup_k \tau_{k,n}).$$

In combination with the narrow convergence of $\overline{\rho}_t^{\boldsymbol{\tau}_n}$, the uniform convergence of $\partial_t\varphi_{t+\tau_{k,n}}$ to $\partial_t\varphi_t$ and the narrow convergence of $\rho_0^{\boldsymbol{\tau}_n}$ to ρ^0 the limit I_4 and is given by

$$\lim_{n\to\infty} I_4 = -\int_0^T \int_\Theta \partial_t\varphi_t\rho_t^*\,\mathrm{d}x\,\mathrm{d}t - \int_\Theta \varphi_0\rho^0\,\mathrm{d}x.$$

Finally, we can conclude that for an arbitrary test function φ_t the limit curve ρ_t^* satisfies:

$$\begin{aligned}
&\int_0^\infty \int_\Omega -\Delta\varphi_t(\rho_t^*)^m + \langle\nabla\varphi_t, \nabla V_t\rangle\,\rho_t^* + \langle\nabla\varphi_t, \nabla W_t * \rho_t^*\rangle\,\rho_t^*\,\mathrm{d}x\,\mathrm{d}t \\
&= \int_0^\infty \int_\Omega \partial_t\varphi_t\,\rho_t^*\,\mathrm{d}x\,\mathrm{d}t + \int_\Omega \varphi_0\,\rho^0\,\mathrm{d}x.
\end{aligned} \tag{3.2.11}$$

This yields that ρ_t^* is a solution to the non-autonomous and non-linear Fokker-Planck equation (3.2.1) in the weak sense. □

4 Time-Homogenization of Gradient Flows

This chapter is based on the second part of the joint work with Jonathan Zinsl [82]. It is devoted to the time-homogenization of non-autonomous evolution systems in the high-frequency limit. I.e., we study the limit $\omega \to \infty$ of the family $(u_t^\omega)_\omega$ of curves of steepest descent with respect to the time-dependent and periodic free energy functionals $\mathcal{E}_{\omega t}$ in the separable, complete metric space $(\mathbf{X}, \boldsymbol{d})$ when the driving free energy functional $\mathcal{E}_t$ is periodic and λ-convex and the oscillatory part $\mathcal{P}_t$ of the free energy functional $\mathcal{E}_t$ is Lipschitz-continuous in space uniformly in time.

Further, we are interested in the high-frequency limit of the non-autonomous Fokker-Planck equation even in situations where the confinement potential V_t and the interaction kernel W_t are not λ-convex.

Main Idea in Short. In the simple Euclidean setting, when $\mathbf{X} = \mathbb{R}^{\mathrm{d}}$, $\boldsymbol{d}$ is induced by the Euclidean metric, and $\mathcal{E}_t \in \mathcal{C}^\infty([0,\infty) \times \mathbb{R}^{\mathrm{d}})$ is convex, we decompose the free energy functional $\mathcal{E}$ into the time-averaged part $\overline{\mathcal{E}}$, defined by $\overline{\mathcal{E}}(u) = \fint \mathcal{E}_t(u)\,\mathrm{d}r$, and the remaining oscillatory party $\mathcal{P}_t$, defined by $\mathcal{P}_t := \mathcal{E}_t - \overline{\mathcal{E}}$. The aim of this proof is to derive a comparison principle or u_t^ω and u_t^∞, the solution of the gradient flow equation with respect to $\overline{\mathcal{E}}$. We can resort on the implicit representations of the two solutions u_t^ω and u_t^∞, which are given by

$$u_t^\omega = u_0 - \int_0^t \nabla \mathcal{E}_{\omega r}(u_r^\omega)\,\mathrm{d}r \qquad \text{and} \qquad u_t^\infty = u_0 - \int_0^t \nabla \overline{\mathcal{E}}(u_r^\infty)\,\mathrm{d}r.$$

Inserting these representation into the the squared distance of u_t^ω and u_t^∞ yields

$$\begin{aligned} \left\|u_t^\omega - u_t^\infty\right\|^2 &= \int_0^t \langle \nabla \mathcal{E}_{\omega r}(u_r^\omega), u_t^\infty - u_t^\omega \rangle + \langle \nabla \overline{\mathcal{E}}(u_r^\infty), u_t^\omega - u_t^\infty \rangle\,\mathrm{d}r \\ &\leq \int_0^t \mathcal{E}_{\omega r}(u_r^\infty) - \mathcal{E}_{\omega r}(u_r^\omega) + \overline{\mathcal{E}}(u_r^\omega) - \overline{\mathcal{E}}(u_r^\infty)\,\mathrm{d}r. \end{aligned}$$

Note, the last inequality is due to the convexity of the free energy functional $\mathcal{E}_t$. Using $\mathcal{E}_t - \overline{\mathcal{E}} = \mathcal{P}_t$, we get the fundamental inequality of the this chapter

$$\left\|u_t^\omega - u_t^\infty\right\|^2 \leq \int_0^t \mathcal{P}_{\omega r}(u_r^\infty) - \mathcal{P}_{\omega r}(u_r^\omega)\,\mathrm{d}r. \tag{4.0.1}$$

Now with this inequality at hand, one is able to prove – after a tedious and technical calculation, where one exploits the periodicity and the uniform Lipschitz-continuity in space of $\mathcal{P}_t$ – the convergence of the family $(u_t^\omega)_\omega$ to u_t^∞ as $\omega \to \infty$ with rate $\mathcal{O}(1/\sqrt{\omega})$.

Contribution&Method. The comparatively weak solution concept of curves of steepest descent when solutions solve the EDE is not suitable to pass to the high-frequency limit. Hence, we can not extend our results from the previous chapter 3 in the case that the free energy functional $\mathcal{E}_t$ satisfies solely (E0)–(E5).

As opposed to this the stronger notion of solutions in the sense of the EVI when the free energy functional $\mathcal{E}_t$ is additional λ-convex is eligible. Here, it is remarkable that the fundamental inequality (4.0.1) can be derived from the EVI (2.3.2), see (4.1.7). With this inequality and the assumption on uniform Lipschitz-continuity of the oscillatory part $\mathcal{P}_t$ we can derive the desired convergence result. The necessary ω-uniform stability results to pass to the limit $\omega \to \infty$ are retrieved by transferring the classical estimates (3.1.16) and the better a priori bounds (3.2.10) from the discrete level of the time-dependent Minimizing Movement scheme to the continuous level.

In the case of the non-autonomous Fokker-Planck equation, we relinquish the assumption on the λ-convexity of the free energy functional $\mathcal{E}_t$. It is still possible to pass to the high-frequency limit and prove that the solutions ρ_t^ω from (4.2.1) converge in a strong topology to a limit curve ρ_t^∞ which is the solution to the autonomous Fokker-Planck equation with time-averaged confinement potential $\overline{V}$ and interaction kernel $\overline{W}$.

Main Results. Our main result of this part concerning the limit behaviour of the family $(u_t^\omega)_\omega$ with respect to the semi-convex free energy functional $\mathcal{E}_{\omega t}$ reads as follows:

Theorem 4.0.1 (Abstract Metric Space)**.** *Assume (E1)–(E3), (E4'), (E5), and (E6) from Assumptions 3.1.2&3.1.3&4.1.1holds. Then, the family $(u_t^\omega)_\omega$ of curves of steepest descent with respect to the free energy $\mathcal{E}_{\omega t}$ converges to a solution u_t^∞ of the gradient flow with respect to the averaged energy $\overline{\mathcal{E}}$, where one defines $\overline{\mathcal{E}}(u) := \fint \mathcal{E}_t(u)\,\mathrm{d}t$.*

Furthermore, there exists a constant C, depending only on u_0 and T such that we have the uniform convergence

$$\boldsymbol{d}(u_t^\omega, u_t^\infty) \leq \frac{C}{\sqrt{\omega}} \qquad \forall\, t \in [0,T].$$

Our main result concerning the limit behaviour in the high-frequency limit as $\omega \to \infty$ of the family $(\rho_t^\omega)_\omega$ of solutions to (4.2.1) reads as follows.

Theorem 4.0.2 (Fokker-Planck Equation)**.** *Let $\Omega \subset \mathbb{R}^\mathrm{d}$ be either an open, bounded, and connected domain with Lipschitz continuous boundary $\partial\Omega$ or let $\Omega = \mathbb{R}^\mathrm{d}$. Further, assume $m \geq 1$ and that V_t, W_t satisfy additionally to (F1)–(F3) as in Assumption 3.2.1 also (F4) specified in Assumption 4.2.1. Consider a sequence $(\omega_n)_{n\in\mathbb{N}}$ with $\omega_n \to \infty$.*

For every $T > 0$, there exists a (non-relabelled) subsequence of $(\omega_n)_{n\in\mathbb{N}}$ and a curve $\rho_t^\infty : [0,\infty) \times \Omega \to [0,\infty)$ such that the family $(\rho_t^{\omega_n})_{n\in\mathbb{N}}$ of weak solutions to (4.2.1) obtained from (3.0.2) converges to ρ_t^∞,

$$\begin{aligned} \rho_t^{\omega_n} &\rightharpoonup^* \rho_t^\infty && \text{narrowly for every } t \in [0,\infty), \\ \rho_t^{\omega_n} &\to \rho_t^\infty && \text{strongly in } \mathrm{L}^p(0,T;\mathrm{L}^m(\Theta)) \text{ as } n \to \infty, \end{aligned}$$

with $\Theta = \Omega$ if the latter is bounded or $\Theta \subset \mathbb{R}^\mathrm{d}$ any bounded subset of $\mathbb{R}^\mathrm{d}$ and any $p \in [1,\infty)$.

4.1 Application to Gradient Flows in Abstract Metric Space

This section is devoted to study the high-frequency limit $\omega \to \infty$ of solutions u_t^ω to the non-autonomous evolution problem of the form

$$\dot{u}_t^\omega = -\nabla_{\mathbf{X}} \mathcal{E}_{\omega t}(u_t^\omega), \qquad u_0^\omega = u_0, \tag{4.1.1}$$

where the driving free energy functional $\mathcal{E}_t$ obeys a sort of λ-convexity. In this case, there exists a solution u_t^ω to (4.1.1) in the sense of the evolution variational equation which solves the evolution variational equation

$$\frac{1}{2}\boldsymbol{d}^2(u_t^\omega, w) - \frac{1}{2}\boldsymbol{d}^2(u_s^\omega, w) \leq \int_s^t \left[\mathcal{E}_{\omega r}(w) - \mathcal{E}_{\omega r}(u_r^\omega) - \frac{\lambda}{2}\boldsymbol{d}^2(u_r^\omega, w)\right] \mathrm{d}r \tag{4.1.2}$$

for each $0 \leq s \leq t$ and any $w \in \mathcal{D}(\mathcal{E}_0)$. The aim is to prove that the family of solutions $(u_t^\omega)_\omega$ converges with convergence rate one-half to the solution u_t^∞ to the gradient flow with respect to the time-averaged free energy functional $\overline{\mathcal{E}}$, defined by $\overline{\mathcal{E}}(u) = ⨍ \mathcal{E}_t(u)\,\mathrm{d}t$ for each $u \in \mathbf{X}$.

Convexity. In the context of abstract metric spaces one notion of convexity of the free energy functional $\mathcal{E}$, which is well-adapted to the gradient flow theory, is the famous Assumption 4.0.1 in [4]. In their recent work [39], Ferreira and Valencia-Guevara extended this notion of convexity for time-dependent free energy functionals $\mathcal{E}_t$: There exists a function λ_t such that for every triple $u, v_0, v_1 \in \mathbf{X}$, there exists a curve $\gamma_t : [0,1] \to \mathbf{X}$ with $\gamma_0 = v_0,\ \gamma_1 = v_1$ and

$$\Phi(\tau, t, u; \gamma_s) \leq (1-s)\Phi(\tau, t, u; v_0) + s\Phi(\tau, t, u; v_1) - \frac{1}{2}\Big(\frac{1}{\tau} + \lambda_t\Big)s(1-s)\boldsymbol{d}^2(v_0, v_1)$$

where Φ is the Moreau-Yosida functional of the free energy functional $\mathcal{E}_t$. With this assumption on the free energy functional $\mathcal{E}_t$ it is possible to construct solutions u_t^* of the gradient flow (4.1.1) in the sense of the EVI (4.1.2) by means of the time-dependent Minimizing Movement scheme (3.0.2).

Strategy of Proof. The strategy of the proof for the high-frequency limit is divided into two parts. Firstly, we derive the necessary compactness estimates with the help of the time-dependent Minimizing Movement scheme. The key ingredient in the derivation of these ω-independent estimates is the uniform Lipschitz-continuity in space of the oscillatory part $\mathcal{P}_t$ of the free energy functional $\mathcal{E}_t$, defined by $\mathcal{P}_t(u) := \mathcal{E}_t(u) - \overline{\mathcal{E}}(u)$. This rather restrictive assumption uncouples the dependency on the frequency ω from the classical estimates (3.1.13), see section 4.1.2. Secondly, we exploit in section 4.1.3 the evolution variational inequalities for $\mathcal{E}_{\omega t}$ and $\overline{\mathcal{E}}$ to establish the metric surrogate of the comparison principle (4.0.1). Now, combining the ω-independent estimates and the comparison principle yields after a technical and tedious calculation the desired convergence result of $(u_t^\omega)_\omega$ to u_t^∞ with explicit (sub-optimal) convergence rate one half.

4.1.1 Setup and Assumption

Given a separable, complete metric space $(\mathbf{X}, \boldsymbol{d})$. Additionally to the assumptions (E1)–(E3), and (E5) on the proper free energy functional $\mathcal{E}_t$ from the previous section 3.1, we shall also assume the following periodicity assumptions.

Assumption 4.1.1. Decompose the free energy functional $\mathcal{E}_t = \overline{\mathcal{E}} + \mathcal{P}_t$, where $\overline{\mathcal{E}}$ denotes the time-mean of $\mathcal{E}_t$, i.e., $\overline{\mathcal{E}}(u) := \fint \mathcal{E}_t(u)\,\mathrm{d}t$, and $\mathcal{P}_t$ is the oscillatory part with zero mean and period p. The oscillatory part $\mathcal{P}_t : [0, \infty) \times \mathbf{X} \to \mathbb{R}$ satisfies additionally:

(E4') **Lipschitz-continuity in space:** There exists $L \geq 0$ such that for all $u, v \in \mathbf{X}$ and $t \in [0, \infty)$

$$|\mathcal{P}_t(u) - \mathcal{P}_t(v)| \leq L\boldsymbol{d}(u, v).$$

(E6) **Semi-Convexity:** There exists a function λ_t such that for every $u, \gamma_0, \gamma_1 \in \mathcal{D}(\mathcal{E}_t)$ and every $\tau \in [0, \tau_*)$, there exists a continuous curve $\gamma_t : [0,1] \to \mathbf{X}$ joining the given end points γ_0 and γ_1, along which the penalized energy Φ satisfies

$$\Phi(\tau, t, u; \gamma_s) \leq (1-s)\Phi(\tau, t, u; \gamma_0) + s\Phi(\tau, t, u; \gamma_1) - \frac{1}{2}\Big(\frac{1}{\tau} + \lambda_t\Big) s(1-s)\boldsymbol{d}^2(\gamma_0, \gamma_1).$$

4.1.2 Classical Estimates Revisited

In this section we prove the stability of the classical estimate (3.1.13) in the discrete-to-continuous limit. Note, these estimates are a priori not stable in the high-frequency limit. However, the additional spatial Lipschitz-continuity (E4') of the oscillatory part $\mathcal{P}_t$ yields the classical bounds *independent* of ω.

Lemma 4.1.2 (Classical Estimates Revisited)**.** *Let $u_0 \in \mathcal{D}(\mathcal{E}_0)$ and let u_t^ω be the curve of steepest descent with respect to $\mathcal{E}_{\omega t}$. For fixed $T > 0$, there exists a constant C, independent of ω, such that for all $t \leq T$ there holds:*

$$\big\| |(u_t^\omega)'| \big\|_{\mathrm{L}^2(0,T)} \leq C, \qquad \mathcal{E}_0(u_t^\omega) \leq C, \qquad \boldsymbol{d}^2(u_*, u_t^\omega) \leq C. \tag{4.1.3}$$

Proof. We will prove the estimates on a discrete level and then we use the lower semi-$\boldsymbol{\sigma}$-continuity of each bound to obtain the desired result. For this purpose fix a partition $\boldsymbol{\tau}$, with $\sup_k \tau_k < \frac{\tau_*}{2}$ and (I1) from Assumption 3.1.6, and let $\big(u_k^{\boldsymbol{\tau},\omega}\big)_{k\in\mathbb{N}}$ be the corresponding discrete solution with $u_0^{\boldsymbol{\tau},\omega} = u_0$. As in the proof of Theorem 3.1.16 we derive for the discrete solutions $(u_k^{\boldsymbol{\tau},\omega})_{k\in\mathbb{N}}$ the inequality

$$\sum_{k=1}^{N} \frac{1}{2\tau_k}\boldsymbol{d}^2(u_{k-1}^{\boldsymbol{\tau},\omega}, u_k^{\boldsymbol{\tau},\omega}) \leq \overline{\mathcal{E}}(u^0) - \overline{\mathcal{E}}(u_N^{\boldsymbol{\tau},\omega}) + \sum_{k=1}^{N-1} \left[\mathcal{P}_{\omega t_k^{\boldsymbol{\tau}}}(u_{k-1}^{\boldsymbol{\tau},\omega}) - \mathcal{P}_{\omega t_k^{\boldsymbol{\tau}}}(u_k^{\boldsymbol{\tau},\omega})\right].$$

Exploit the Lipschitz-continuity of $\mathcal{P}_t$ (E4') and use Young's inequality to further estimate the right-hand-side to obtain

$$\sum_{k=1}^{N-1}\left[\mathcal{P}_{\omega t_k^{\boldsymbol{\tau}}}(u_{k\text{-}1}^{\boldsymbol{\tau},\omega})-\mathcal{P}_{\omega t_k^{\boldsymbol{\tau}}}(u_k^{\boldsymbol{\tau},\omega})\right]\le\sum_{k=1}^{N-1}L\boldsymbol{d}(u_{k\text{-}1}^{\boldsymbol{\tau},\omega},u_k^{\boldsymbol{\tau},\omega})\le\sum_{k=1}^{N-1}\Big[L^2\tau_k+\frac{1}{4\tau_k}\boldsymbol{d}^2(u_{k\text{-}1}^{\boldsymbol{\tau},\omega},u_k^{\boldsymbol{\tau},\omega})\Big].$$

A kick-back argument and the coercivity of $\overline{\mathcal{E}}$ yields now

$$\sum_{k=1}^{N}\frac{1}{4\tau_k}\boldsymbol{d}^2(u_{k\text{-}1}^{\boldsymbol{\tau},\omega},u_k^{\boldsymbol{\tau},\omega})\le\overline{\mathcal{E}}(u^0)+\frac{1}{2\tau_*}\boldsymbol{d}^2(u_*,u_N^{\boldsymbol{\tau},\omega})-c_*+L^2T.$$

Perform a calculation which is similar to the proof of theorem 3.1.16 to get

$$\boldsymbol{d}^2(u_*,u_N^{\boldsymbol{\tau},\omega})\le 2\tau_*\left(\overline{\mathcal{E}}(u_0)-c_*+TL^2\right)+2\boldsymbol{d}^2(u_*,u_0)+\frac{2}{\tau_*}\sum_{k=1}^{N}\tau_k\boldsymbol{d}^2(u_*,u_k^{\boldsymbol{\tau},\omega}).$$

Notice that every constant appearing in this equation is independent of ω. Now repeat the remaining part of the proof with the discrete Gronwall's lemma [4, Lemma 3.2.4] to get the desired estimates on the discrete level, which are independent of ω, i.e., we have for all N with $t_{\boldsymbol{\tau}}^N<T$:

$$\sum_{k=1}^{N}\frac{1}{2\tau_k}\boldsymbol{d}^2(u_{k\text{-}1}^{\boldsymbol{\tau},\omega},u_k^{\boldsymbol{\tau},\omega})\le C,\qquad\mathcal{E}_{\omega t_k^{\boldsymbol{\tau}}}(u_N^{\boldsymbol{\tau},\omega})\le C,\qquad\boldsymbol{d}^2(u_*,u_N^{\boldsymbol{\tau},\omega})\le C.\tag{4.1.4}$$

Due to periodicity of $\mathcal{E}_t$ we have also

$$\mathcal{E}_0(u_N^{\boldsymbol{\tau},\omega})\le\mathcal{E}_{\omega t_k^{\boldsymbol{\tau}}}(u_N^{\boldsymbol{\tau},\omega})+(1+\boldsymbol{d}^2(u_*,u_N^{\boldsymbol{\tau},\omega}))\int_0^p\alpha_t\,\mathrm{d}t\le C.$$

Hence, we have proven the estimates (4.1.3) on a discrete level.

Now consider a family $\boldsymbol{\tau}_n$ of admissible partitions satisfying (I1), then by theorem 3.1.18 the corresponding piecewise constant interpolation $\overline{u}_t^{\boldsymbol{\tau}_n,\omega}$ converges with respect to the weak $\boldsymbol{\sigma}$-topology to a limit curve u_t^ω. Since the $\mathrm{L}^2(0,T)$-norm is lower semi-continuous with respect to the weak topology, we have for the metric slope $|(u_t^\omega)'|$:

$$\int_0^T|(u_t^\omega)'|^2\,\mathrm{d}t\le\int_0^T(A_t)^2\,\mathrm{d}t\le\liminf_{n\to\infty}\int_0^T(A_t^n)^2\,\mathrm{d}t$$

where A_t is the weak limit of the discrete derivative A_t^n from the proof of theorem 3.1.18. Hence, with (4.1.4) and the definition of A_t^n we get

$$\int_0^T|(u_t^\omega)'|^2\,\mathrm{d}t\le\liminf_{n\to\infty}\int_0^T(A_t^n)^2\,\mathrm{d}t=\liminf_{n\to\infty}\sum_{k=1}^{k_n(T)+1}\frac{1}{2\tau_k}\boldsymbol{d}^2(u_{k\text{-}1}^{\boldsymbol{\tau},\omega},u_k^{\boldsymbol{\tau},\omega})\le C.$$

Since $\mathcal{E}_0$, and $\boldsymbol{d}$ are lower semi-$\boldsymbol{\sigma}$-continuous, we have

$$\mathcal{E}_0(u_t^\omega)\le\liminf_{n\to\infty}\mathcal{E}_0(\boldsymbol{\tau}_t^{\boldsymbol{\tau}_n,\omega})\le C,\quad\text{and}\quad\boldsymbol{d}^2(u_*,u_t^\omega)\le\liminf_{n\to\infty}\boldsymbol{d}^2(u_*,\overline{u}_t^{\boldsymbol{\tau}_n,\omega})\le C.$$

Which yields the desired uniform estimates (4.1.3) which are independent of the frequency parameter ω. □

4.1.3 High-Frequency Limit

Theorem 4.1.3 (Convergence). *The family $(u_t^\omega)_\omega$ of curves of steepest descent with respect to the free energy $\mathcal{E}_{\omega t}$ converges to the solution u_t^∞ of the gradient flow with respect to the averaged energy $\overline{\mathcal{E}}$.*

Furthermore, there exists a Constant C, depending only on u_0, τ_* and T such that we have the uniform convergence

$$\boldsymbol{d}(u_t^\omega, u_t^\infty) \leq \frac{C}{\sqrt{\omega}} \qquad \forall\, t \in [0,T].$$

Proof. Without loss of generality we assume that the period p of $\mathcal{P}_t$ is equal to one and that λ_t from (E6) can be constant and negative, i.e., $\lambda_t = \lambda \leq 0$. By [39, Equation (5.15)] and [4, Thm. 4.0.4], u_t^ω and u_t^∞ satisfy the following two evolution variational inequalities, respectively for all $w \in \mathbf{X}$:

$$\frac{1}{2}\boldsymbol{d}^2(u_t^\omega, w) - \frac{1}{2}\boldsymbol{d}^2(u_s^\omega, w) + \int_s^t \frac{\lambda}{2}\boldsymbol{d}^2(u_r^\omega, w) + \mathcal{E}_{\omega r}(u_r^\omega)\,\mathrm{d}r \leq \int_s^t \mathcal{E}_{\omega r}(w)\,\mathrm{d}r, \tag{4.1.5}$$

$$\frac{1}{2}\boldsymbol{d}^2(u_t^\infty, w) - \frac{1}{2}\boldsymbol{d}^2(u_s^\infty, w) + \int_s^t \frac{\lambda}{2}\boldsymbol{d}^2(u_r^\infty, w) + \overline{\mathcal{E}}(u_r^\infty)\,\mathrm{d}r \leq \int_s^t \overline{\mathcal{E}}(w)\,\mathrm{d}r. \tag{4.1.6}$$

In order to prove the statement, we apply a Gronwall type argument, i.e., we differentiate the square of the distance of u_t^ω and u_t^∞. Since the solution curves are absolutely continuous, this step is valid and we can apply [4, Lemma 4.3.4] to obtain

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}s}\frac{1}{2}\boldsymbol{d}^2(u_s^\omega, u_s^\infty)\Big|_{s=t} \leq &\limsup_{h\searrow 0} \frac{\frac{1}{2}\boldsymbol{d}^2(u_t^\omega, u_t^\infty) - \frac{1}{2}\boldsymbol{d}^2(u_{t-h}^\omega, u_t^\infty)}{h} \\ &+ \limsup_{h\searrow 0} \frac{\frac{1}{2}\boldsymbol{d}^2(u_t^\omega, u_{t+h}^\infty) - \frac{1}{2}\boldsymbol{d}^2(u_t^\omega, u_t^\infty)}{h}.\end{aligned}$$

The first term on the right-hand-side can be estimated using the evolution variational equation (4.1.5), the lower semi-$\boldsymbol{\sigma}$-continuity of $\mathcal{E}_t$ and $\boldsymbol{d}$ to get

$$\begin{aligned}&\limsup_{h\searrow 0} \frac{\frac{1}{2}\boldsymbol{d}^2(u_t^\omega, u_t^\infty) - \frac{1}{2}\boldsymbol{d}^2(u_{t-h}^\omega, u_t^\infty)}{h} \\ \leq &\limsup_{h\searrow 0} \frac{1}{h}\int_{t-h}^t \left[\mathcal{E}_{\omega r}(u_t^\infty) - \frac{\lambda}{2}\boldsymbol{d}^2(u_t^\omega, u_t^\infty) - \mathcal{E}_{\omega r}(u_r^\omega)\right]\mathrm{d}r \\ \leq &\mathcal{E}_{\omega t}(u_t^\infty) - \frac{\lambda}{2}\boldsymbol{d}^2(u_t^\omega, u_t^\infty) - \mathcal{E}_{\omega t}(u_t^\omega).\end{aligned}$$

Analogously, using the EVI (4.1.6) we obtain for the limit of the second term

$$\limsup_{h\searrow 0} \frac{\frac{1}{2}\boldsymbol{d}^2(u_t^\omega, u_{t+h}^\infty) - \frac{1}{2}\boldsymbol{d}^2(u_t^\omega, u_t^\infty)}{h} \leq \overline{\mathcal{E}}(u_t^\omega) - \frac{\lambda}{2}\boldsymbol{d}^2(u_t^\omega, u_t^\infty) - \overline{\mathcal{E}}(u_t^\infty).$$

Hence, by adding these results we get the following estimate

$$\frac{\mathrm{d}}{\mathrm{d}s}\frac{1}{2}\boldsymbol{d}^2(u_s^\omega, u_s^\infty)\Big|_{s=t} \leq \mathcal{P}_{\omega t}(u_t^\infty) - \mathcal{P}_{\omega t}(u_t^\omega) - \lambda \boldsymbol{d}^2(u_t^\omega, u_t^\infty)$$

from which we conclude with the differential form of Gronwall's inequality that

$$e^{2\lambda t}\boldsymbol{d}^2(u_t^\omega, u_t^\infty) \leq \int_0^t e^{2\lambda r}\left(\mathcal{P}_{\omega r}(u_r^\infty) - \mathcal{P}_{\omega r}(u_r^\omega)\right)\,\mathrm{d}r. \tag{4.1.7}$$

After a rescaling of the time variable the r.h.s. can be decomposed, i.e.,

$$\begin{aligned}\int_0^t e^{2\lambda r}\left(\mathcal{P}_{\omega r}(u_r^\infty) - \mathcal{P}_{\omega r}(u_r^\omega)\right)\,\mathrm{d}r =& \frac{1}{\omega}\int_0^{\omega t} e^{2\lambda r/\omega}\big(\mathcal{P}_r(u_{r/\omega}^\infty) - \mathcal{P}_r(u_{r/\omega}^\omega)\big)\,\mathrm{d}r \\ =& \sum_{k=0}^{\lfloor \omega t\rfloor - 1}\frac{1}{\omega}\int_k^{k+1} e^{2\lambda r/\omega}\big(\mathcal{P}_r(u_{r/\omega}^\infty) - \mathcal{P}_r(u_{r/\omega}^\omega)\big)\,\mathrm{d}r \\ &+ \frac{1}{\omega}\int_{\lfloor \omega t\rfloor}^{\omega t} e^{2\lambda r/\omega}\big(\mathcal{P}_r(u_{r/\omega}^\infty) - \mathcal{P}_r(u_{r/\omega}^\omega)\big)\,\mathrm{d}r.\end{aligned}$$

The first term in this equation can be further expanded inserting two productive zeros, such that we have

$$\begin{aligned}&\frac{1}{\omega}\sum_{k=0}^{\lfloor \omega t\rfloor - 1}\int_k^{k+1} e^{2\lambda r/\omega}\big(\mathcal{P}_r(u_{r/\omega}^\infty) - \mathcal{P}_r(u_{r/\omega}^\omega)\big)\,\mathrm{d}r \\ &= \frac{1}{\omega}\sum_{k=0}^{\lfloor \omega t\rfloor - 1}\int_k^{k+1}\Big[e^{2\lambda r/\omega}\big(\mathcal{P}_r(u_{r/\omega}^\infty) - \mathcal{P}_r(u_{r/\omega}^\omega)\big) - e^{2\lambda k/\omega}\big(\mathcal{P}_r(u_{k/\omega}^\infty) - \mathcal{P}_r(u_{k/\omega}^\omega)\big) \\ &\qquad + \big[e^{2\lambda r/\omega} - e^{2\lambda r/\omega}\big]\big(\mathcal{P}_r(u_{k/\omega}^\infty) - \mathcal{P}_r(u_{k/\omega}^\omega)\big)\Big]\,\mathrm{d}r.\end{aligned}$$

Subsequently, use Taylor's expansion for the function $r \mapsto e^{2\lambda r/\omega}$ to get

$$\begin{aligned}&\frac{1}{\omega}\sum_{k=0}^{\lfloor \omega t\rfloor - 1}\int_k^{k+1} e^{2\lambda r/\omega}\big(\mathcal{P}_r(u_{r/\omega}^\infty) - \mathcal{P}_r(u_{r/\omega}^\omega)\big)\,\mathrm{d}r \\ &= \frac{1}{\omega}\sum_{k=0}^{\lfloor \omega t\rfloor - 1}\int_k^{k+1}\Big[e^{2\lambda r/\omega}\big(\mathcal{P}_r(u_{r/\omega}^\infty) - \mathcal{P}_r(u_{r/\omega}^\omega) - \mathcal{P}_r(u_{k/\omega}^\infty) + \mathcal{P}_r(u_{k/\omega}^\omega)\big) \\ &\qquad + e^{2\lambda\zeta_r/\omega}\,\frac{2\lambda(r-k)}{\omega}\big(\mathcal{P}_r(u_{k/\omega}^\infty) - \mathcal{P}_r(u_{k/\omega}^\omega)\big)\Big]\,\mathrm{d}r.\end{aligned}$$

for some $\zeta_r \in [k, r]$. Exploit the Lipschitz continuity of $\mathcal{P}_t$, the absolute continuity of u_t^ω, respectively of u_t^∞, and use the estimates (4.1.3) to obtain the following upper bound for the modulus of the previous equation

$$\begin{aligned}
&\Big|\frac{1}{\omega}\sum_{k=0}^{\lfloor\omega t\rfloor-1}\int_k^{k+1} e^{2\lambda r/\omega}\big(\mathcal{P}_r(u^\infty_{r/\omega})-\mathcal{P}_r(u^\omega_{r/\omega})\big)\,\mathrm{d}r\Big| \\
\leq&\frac{1}{\omega}\sum_{k=0}^{\lfloor\omega t\rfloor-1}\int_k^{k+1}\Big[e^{2\lambda r/\omega}\big(L\boldsymbol{d}(u^\infty_{r/\omega},u^\infty_{k/\omega})+L\boldsymbol{d}(u^\omega_{r/\omega},u^\omega_{k/\omega})\big) \\
&\qquad\qquad + e^{2\lambda\zeta_r/\omega}\frac{2\,|\lambda|\,(r-k)}{\omega}L\boldsymbol{d}(u^\infty_{k/\omega},u^\omega_{k/\omega})\Big]\,\mathrm{d}r \\
\leq&\frac{L}{\omega}\sum_{k=0}^{\lfloor\omega t\rfloor-1}\int_k^{k+1}\Big[\int_{\frac{k}{\omega}}^{\frac{r}{\omega}}|(u^\infty_s)'|\,\mathrm{d}s+\int_{\frac{k}{\omega}}^{\frac{r}{\omega}}|(u^\omega_s)'|\,\mathrm{d}s+\frac{2\,|\lambda|}{\omega}\boldsymbol{d}(u^\infty_{k/\omega},u^\omega_{k/\omega})\Big]\,\mathrm{d}r \\
\leq&\frac{L}{\omega}\int_0^{\frac{\lfloor\omega t\rfloor}{\omega}}|(u^\infty_s)'|+|(u^\omega_s)'|\,\mathrm{d}s+\frac{L}{\omega}\sum_{k=0}^{\lfloor\omega t\rfloor-1}\frac{2\,|\lambda|}{\omega}\boldsymbol{d}(u^\infty_{k/\omega},u^\omega_{k/\omega}) \\
\leq&\frac{L\sqrt{T}}{\omega}\big\||(u^\infty_t)'|\big\|_{\mathrm{L}^2(0,T)}+\frac{L\sqrt{T}}{\omega}\big\||(u^\omega_t)'|\big\|_{\mathrm{L}^2(0,T)}+\frac{2L\,|\lambda|\,T}{\omega}C \\
\leq&\frac{C}{\omega}.
\end{aligned}$$

We estimate the remainder term of the starting equation accordingly with a combination of the Lipschitz continuity of $\mathcal{P}_t$ and estimate (4.1.3) such that we obtain

$$\Big|\frac{1}{\omega}\int_{\lfloor\omega t\rfloor}^{\omega t} e^{2\lambda r/\omega}\big(\mathcal{P}_r(u^\infty_{r/\omega})-\mathcal{P}_r(u^\omega_{r/\omega})\big)\,\mathrm{d}r\Big| \leq \frac{\omega t-\lfloor\omega t\rfloor}{\omega}C\leq\frac{C}{\omega}.$$

Thus, combining these results we obtain

$$e^{2\lambda t}\frac{1}{2}\boldsymbol{d}^2(u^\omega_t,u^\infty_t)\leq\frac{C}{\omega}$$

yielding the desired uniform convergence of u^ω_t to u^∞_t for every finite horizon T. □

4.2 Application to Non-autonomous Fokker-Planck Equation

We analyze in the second part of this chapter the high-frequency limit in the special case of the L^2-Wasserstein formalism when the free energy functional $\mathcal{E}_t$ is not λ-convex. In particular, we investigate the high-frequency limit of the family $(\rho_t^\omega)_\omega$ of solutions to the non-autonomous and non-linear Fokker-Planck equation

$$\partial_t \rho_t^\omega = \Delta(\rho_t^\omega)^m + \mathrm{div}(\rho_t^\omega \nabla V_{\omega t}) + \mathrm{div}(\rho_t^\omega(\nabla W_{\omega t} * \rho_t^\omega)), \qquad \rho_0^\omega = \rho_0, \tag{4.2.1}$$

with non-flux boundary condition in a domain $\Omega \subseteq \mathbb{R}^\mathrm{d}$, which is either an open, bounded and connected domain Ω with Lipschitz continuous boundary $\partial\Omega$ and normal derivative $\boldsymbol{n}$ or equal to the entire space $\Omega = \mathbb{R}^\mathrm{d}$. As before, the driving free energy functional $\mathcal{E}_{\omega t}$ of the L^2-Wasserstein formalism is given by

$$\mathcal{E}_{\omega t}(\mu) := \begin{cases} \int_\Omega \rho\log(\rho) + V_{\omega t}\rho + \frac{1}{2}(W_{\omega t} * \rho)\rho \,\mathrm{d}x & \text{if } m = 1, \\ \int_\Omega \frac{1}{m-1}\rho^m + V_{\omega t}\rho + \frac{1}{2}(W_{\omega t} * \rho)\rho \,\mathrm{d}x & \text{if } m > 1, \end{cases} \tag{4.2.2}$$

if $\mu = \rho\,\mathrm{d}\mathcal{L}^d$ and otherwise we set $\mathcal{E}_{\omega t}(\mu) = \infty$. Here, we focus on the case when the confinement potential V_t and the interaction kernel W_t are not λ-convex. Thus we are in the L^2-Wasserstein framework where the driving free energy functional $\mathcal{E}_t$ is not convex along generalized geodesics and the theory of high-frequency limits in abstract metric spaces is not applicable.

Strategy of Proof. Due to the lack of convexity of the free energy functional $\mathcal{E}_t$ we are not able to apply the theory developed in the previous section 4.1. Neither we can establish a comparison principle for the family $(\rho_t^\omega)_\omega$ and the limit function ρ_t^∞, but this goes in line with the L^2-Wasserstein theory of the Fokker-Planck equation without having additional regularity like the λ-convexity.

Still, your ideology behind the proof of the high-frequency limit is to derive the necessary compactness estimates by means of the time-dependent Minimizing Movement scheme and then pass in the weak formulations of the non-autonomous Fokker-Planck equations (4.2.1) to the limit $\omega \to \infty$. Note, the classical estimates (4.1.3) are intrinsic properties of the time-dependent Minimizing Movement scheme and solely rely on the uniform Lipschitz-regularity in space of the oscillatory party $\mathcal{P}_t$ of free energy functional $\mathcal{E}_t$. Hence, we recover in section 4.2.2 the classical estimates. Even for the refined regularity $\mathrm{L}^2(0,T;\mathrm{BV}(\Omega))$-estimates for $(\overline{\rho}_t^{\boldsymbol{\tau},\omega})^m$ it can be proven that these bounds are stable with respect to the discrete-to-continuous limit $\boldsymbol{\tau} \to 0$ and are independent of the frequency ω. These estimates are sufficient to pass to the high-frequency limit $\omega \to \infty$ of the family $(\rho_t^\omega)_\omega$ and prove the convergence in the narrow-topology and in the strong $\mathrm{L}^p(0,T;\mathrm{L}^m(\Omega))$-topology to a solution ρ_t^∞ to the time-averaged Fokker-Planck equation, see section 4.2.3. Note, to pass to the limit $\omega \to \infty$ in the weak formulation of the non-autonomous Fokker-Planck equations (4.2.1) we utilize the theory of Γ-convergence and use the fact that the families $(V_{\omega t})_\omega$ and $(W_{\omega t})_\omega$ converges in a weak sense to the time-averaged functions $\overline{V}$ and $\overline{W}$, respectively.

4.2.1 Setup and Assumptions

Compared to the previous section, we solely assume the Lipschitz-continuity in space uniformly in time and not the semi-convexity of the free-energy $\mathcal{E}_t$.

Assumption 4.2.1. Decompose the confinement potential $V_t := \overline{V} + \mathcal{P}_t$ and the interaction kernel $W_t : \overline{W} + \mathcal{R}_t$ into the time independent parts $\overline{V}$ and $\overline{W}$, respectively, and the oscillatory parts $\mathcal{P}_t$ and $\mathcal{R}_t$, respectively, with zero mean and period p. The oscillatory parts $\mathcal{P}_t$ and $\mathcal{R}_t$ satisfy additionally:

(F4) **Lipschitz-continuity in space:** There exists $L \geq 0$ such that for all $t \in [0, \infty)$

$$|\mathcal{P}_t(x) - \mathcal{P}_t(y)|\,,\, |\mathcal{R}_t(x) - \mathcal{R}_t(y)| \leq L\,\|x - y\|\,.$$

4.2.2 Classical Estimates Revisited

It is clear that (E4') follows from (F4) and therefore, we can derive from this in the same manner the classical estimates from Lemma 4.1.2, i.e.:

Lemma 4.2.2 (Classical Estimates)**.** *For fixed $T > 0$, there exists a constant C, independent of ω, such that for all $t \leq T$ there holds:*

$$\big\||(\rho_t^\omega)'|\big\|_{\mathrm{L}^2(0,T)} \leq C, \qquad \mathcal{E}_0(\rho_t^\omega) \leq C, \qquad \boldsymbol{d}^2(\rho_*, \rho_t^\omega) \leq C. \tag{4.2.3}$$

Subsequently, also the estimate from Lemma 3.2.3 is independent of ω and therefore for fixed $T > 0$ there holds for the same constant C and for all $t \leq T$:

$$\boldsymbol{M}_2(\rho_t^\omega) \leq C, \qquad\qquad \mathcal{U}_m(\rho_t^\omega) \leq C.$$

Lastly, the better a priori bounds 3.2.6 are independent of ω, too. Since the $\mathrm{L}^2(0,T;\mathrm{BV}(\Omega))$ is lower semi-continuous with respect to the $\mathrm{L}^2(0,T;\mathrm{L}^m(\Omega))$-topology, this estimate is preserved in the discrete-to-continuous limit, i.e.:

Lemma 4.2.3 (Step Size Independent Global $\mathrm{L}^2(0,T;\mathrm{BV}(\Omega))$-estimates)**.** *For fixed $T > 0$, there exits a constant C, independent of ω, such that:*

$$\|(\rho_t^\omega)^m\|_{\mathrm{L}^2(0,T;\mathrm{BV}(\Omega))} \leq C. \tag{4.2.4}$$

4.2.3 High-Frequency Limit

In this section, we finally prove the high-frequency limit $\omega_n \to \infty$ for the family of weak solutions $(\rho_t^{\omega_n})_{n\in\mathbb{N}}$ to (4.2.1) obtained by the time dependent Minimizing Movement scheme (3.0.2). In particular, we prove that this family $(\rho_t^{\omega_n})_{n\in\mathbb{N}}$ converges in the narrow-topology and in the strong $\mathrm{L}^p(0,T;\mathrm{L}^m(\Theta))$-topology to the solution ρ_t^∞ to the time-averaged Fokker-Planck equation.

Theorem 4.2.4 (Narrow Convergence). *Define for a sequence $(\omega_n)_{n\in\mathbb{N}}$ with $\omega_n \to \infty$ the family of weak solutions $(\rho_t^{\omega_n})_{n\in\mathbb{N}}$ obtained by the time-dependent Minimizing Movement scheme* (3.0.2) *with respect to the free energy functional $\mathcal{E}_{\omega_n t}$. Then, there exists a (non-relabelled) subsequence of $(\omega_n)_{n\in\mathbb{N}}$ such that*

$$\rho_t^{\omega_n} \rightharpoonup^* \rho_t^\infty \qquad \textit{narrowly for every } t \in [0,\infty).$$

Proof. At first, we prove the existence of a narrow convergent subsequence, which converges to an absolutely continuous curve. By the uniform $\mathrm{L}^2(0,T)$-estimate (4.2.3) on the metric velocity $|(\rho_t^{\omega_n})'|$ we can extract a (non-relaballed) subsequence such that $|(\rho_t^{\omega_n})'|$ converges weakly in $\mathrm{L}^2(0,T)$ to $A_t \in \mathrm{L}^2(0,T)$. To apply the Arzelá-Ascoli theorem we estimate now

$$\limsup_{n\to\infty} \mathbf{W}_2(\rho_t^{\omega_n}, \rho_s^{\omega_n}) \le \limsup_{n\to\infty} \int_s^t |(\rho_r^{\omega_n})'| \,\mathrm{d}r = \int_s^t A_r \,\mathrm{d}r.$$

Since the entropy of ρ_t^ω is uniformly bounded by (4.2.3), the sequence $(\rho_t^{\omega_n})_{n\in\mathbb{N}}$ is contained in a weak*-compact set. Therefore, by the refined Arzelá-Ascoli theorem [4, Proposition 3.3.1] we obtain the existence of a limit curve $\rho_t^\infty \in \mathrm{AC}^2\left(0,\infty;(\mathcal{P}_2(\Omega),\mathbf{W}_2)\right)$ such that $\rho_t^{\omega_n}$ converges pointwise with respect to the narrow convergence. □

Theorem 4.2.5 (Strong Convergence in $\mathrm{L}^p(0,T;\mathrm{L}^m(\Omega))$). *Given a sequence $(\omega_n)_{n\in\mathbb{N}}$ with $\omega_n \to \infty$, and given the limit curve ρ_t^∞ in theorem 4.2.4. Then, there exists a further (non-relabelled) subsequence of $(\omega_n)_{n\in\mathbb{N}}$ such that for any $T > 0$, any $p \ge 1$ and any bounded subset $\Theta \subseteq \Omega$*

$$\rho_t^{\omega_n} \to \rho_t^\infty \qquad \textit{strongly in } \mathrm{L}^p(0,T;\mathrm{L}^m(\Theta)) \textit{ as } n \to \infty.$$

Proof. With our los of generality we fix $\Theta = \Omega$ or $\Theta = \mathbb{B}_R(0)$. To obtain the strong $\mathrm{L}^p(0,T;\mathrm{L}^m(\Theta))$-convergence result, we proceed as in the convergence proofs of Theorem 3.2.7 and apply the extension of the Aubin-Lions Theorem 2.5.4 to the sequence $(\rho_t^{\omega_n}|_\Theta)_{n\in\mathbb{N}}$ with the auxiliary functionals $\widetilde{\mathcal{A}}$ and $\widetilde{g}$ as in the proof of Theorem 3.2.7. As before, due to the better a priori estimates (4.2.4) the sequence is tight with respect to $\widetilde{\mathcal{A}}$. To verify the relaxed averaged weak integral equi-continuity with respect to $\widetilde{g}$ use that $|(\rho_t^{\omega_n})'|$ converges weakly in $\mathrm{L}^2_{\mathrm{loc}}(0,\infty)$ to A. This yields with Fatou's Lemma

$$\begin{aligned}
&\liminf_{h\searrow 0} \limsup_{n\to\infty} \frac{1}{h}\int_0^h \int_0^{T-t} \widetilde{g}(\rho_{s+t}^{\omega_n}|_\Theta, \rho_s^{\omega_n}|_\Theta) \mathrm{d}s \,\mathrm{d}t \\
&\le \liminf_{h\searrow 0} \limsup_{n\to\infty} \frac{1}{h}\int_0^h \int_0^{T-t} \int_s^{s+t} |(\rho_r^{\omega_n})'| \,\mathrm{d}r \,\mathrm{d}s \,\mathrm{d}t \\
&\le \liminf_{h\searrow 0} \frac{1}{h}\int_0^h \int_0^{T-t} \int_s^{s+t} A_r \,\mathrm{d}r \,\mathrm{d}s \,\mathrm{d}t = 0
\end{aligned}$$

By the extension of the Aubin-Lions Theorem 2.5.4 and the Remark 2.1.1 we get the desired convergence result for every compact set $\Theta \subseteq \Omega$ and for every finite time horizon T. □

Theorem 4.2.6 (Solution to the Time-average Fokker-Planck Equation)**.** *Given a sequence $(\omega_n)_{n\in\mathbb{N}}$ with $\omega_n \to \infty$, and given the limit curve ρ_t^∞ in theorem 4.2.4. Then, ρ_t^∞ is a weak solution to the time-averaged Fokker-Planck equation in the sense of* (4.2.6).

Proof. At last we prove that the limit of $\rho_t^{\omega_n}$ solves the time averaged Fokker-Planck equation in a weak sense. Therefore, we calculate the limit, as $n \to \infty$, in the weak formulations of each $\rho_t^{\omega_n}$, i.e., for each test function $\varphi_t \in \mathcal{C}_c^\infty([0,\infty)\times\Omega)$ we have

$$\begin{aligned}&\int_0^\infty \int_\Omega -\Delta\varphi_t(\rho_t^{\omega_n})^m + \langle\nabla\varphi_t, \nabla V_t\rangle\, \rho_t^{\omega_n} + \langle\nabla\varphi_t, \nabla W_t * \rho_t^{\omega_n}\rangle\, \rho_t^{\omega_n}\,\mathrm{d}x\,\mathrm{d}t \\ &= \int_0^\infty \int_\Omega \partial_t\varphi_t\, \rho_t^{\omega_n}\,\mathrm{d}x\,\mathrm{d}t + \int_\Omega \varphi_0\,\rho^0\,\mathrm{d}x.\end{aligned} \tag{4.2.5}$$

Let $T > 0$ and $\Theta \subset \Omega$ be compact such that $\operatorname{supp}\varphi_t \subset [0,T]\times\Theta$. The limit of the first integral in (4.2.5) follows from the $\mathrm{L}^m(0,T;\mathrm{L}^m(\Theta))$-convergence of $\rho_t^{\omega_n}$ and we get

$$\lim_{n\to\infty}\int_0^\infty \int_\Omega -\Delta\varphi_t(\rho_t^{\omega_n})^m\,\mathrm{d}x\,\mathrm{d}t = \int_0^\infty \int_\Omega \Delta\varphi_t(\rho_t^\infty)^m\,\mathrm{d}x\,\mathrm{d}t.$$

To deduce the limit of the second integral in (4.2.5), we utilize $\nabla V_{\omega t} \rightharpoonup^* \nabla\overline{V}$ in $\mathrm{L}^\infty(0,T)$ for every $x \in \Theta$ (see, for instance [12]). Thus, also $\nabla V_{\omega_n t} \rightharpoonup^* \nabla\overline{V}$ in $L^\infty(0,T;L^\infty(\Theta))$. Since, $\rho_t^{\omega_n}$ converges to ρ_t^∞ in $\mathrm{L}^1(0,T;\mathrm{L}^1(\Theta))$ we have

$$\nabla V_{\omega_n t}\,\rho_t^{\omega_n} \rightharpoonup^* \nabla\overline{V}\,\rho_t^\infty \quad \text{in } L^\infty(0,T;L^\infty(\Theta)).$$

Since the effective domain of integration is compact we finally have

$$\lim_{n\to\infty}\int_0^\infty \int_\Omega \langle\nabla\varphi_t, \nabla V_t\rangle\,\rho_t^{\omega_n}\,\mathrm{d}x\,\mathrm{d}t = \int_0^\infty \int_\Omega \langle\nabla\varphi_t, \nabla\overline{V}\rangle\,\rho_t^{\omega_n}\,\mathrm{d}x\,\mathrm{d}t.$$

To compute the limit of the third integral, we proceed similar. As before, $\nabla W_{\omega_n t} \rightharpoonup^* \nabla\overline{W}$ for every $x, y \in \mathbb{R}^\mathrm{d}$ and also $\nabla W_{\omega_n t}(x-y) \rightharpoonup^* \nabla\overline{W}(x-y)$ in $L^\infty(0,T;L^\infty(\Theta\times\Theta))$. So by the strong $\mathrm{L}^1(0,T;\mathrm{L}^1(\Theta))$-convergence of $\rho_t^{\omega_n}$ to ρ_t^∞ we have

$$\nabla W_{\omega_n t}(x-y)\,\rho_t^{\omega_n}(x)\rho_t^{\omega_n}(y) \rightharpoonup^* \nabla\overline{W}(x-y)\,\rho_t^\infty(x)\rho_t^\infty(y) \quad \text{in } L^\infty(0,T;L^\infty(\Theta\times\Theta)).$$

Since the effective domain of integration is compact we finally have

$$\lim_{n\to\infty}\int_0^\infty \int_\Omega \langle\nabla\varphi_t, \nabla W_t * \rho_t^{\omega_n}\rangle\,\rho_t^{\omega_n}\,\mathrm{d}x\,\mathrm{d}t = \int_0^\infty \int_\Omega \langle\nabla\varphi_t, \nabla\overline{W} * \rho_t^\infty\rangle\,\rho_t^\infty\,\mathrm{d}x\,\mathrm{d}t.$$

Lastly, by the narrow convergence of $\rho_t^{\omega_n}$ the limit of the left-hand-side of (4.2.5) equals

$$\lim_{n\to\infty}\int_0^\infty \int_\Omega \partial_t\varphi_t\,\rho_t^{\omega_n}\,\mathrm{d}x\,\mathrm{d}t + \int_\Omega \varphi_0\,\rho^0\,\mathrm{d}x = \int_0^\infty \int_\Omega \partial_t\varphi_t\,\rho_t^\infty\,\mathrm{d}x\,\mathrm{d}t + \int_\Omega \varphi_0\,\rho^0\,\mathrm{d}x.$$

Summarized, we obtain that ρ_t^∞ solves for each test function $\varphi_t \in \mathcal{C}_c^\infty([0,\infty)\times\Omega)$:

$$\begin{aligned}&\int_0^\infty \int_\Omega -\Delta\varphi_t(\rho_t^\infty)^m + \langle\nabla\varphi_t, \nabla\overline{V}\rangle\,\rho_t^\infty + \langle\nabla\varphi_t, \nabla\overline{W} * \rho_t^\infty\rangle\,\rho_t^\infty\,\mathrm{d}x\,\mathrm{d}t \\ &= \int_0^\infty \int_\Omega \partial_t\varphi_t\,\rho_t^\infty\,\mathrm{d}x\,\mathrm{d}t + \int_\Omega \varphi_0\,\rho^0\,\mathrm{d}x\end{aligned} \tag{4.2.6}$$

yielding that ρ_t^∞ is a weak solution of the time-averaged Fokker-Planck equation. □

5 Backward Differentiation Formula 2

This chapter is concerned with the novel temporal discretization by means of the second order Backward Differentiation Formula for gradient flows in metric spaces and in particular of drift-diffusion equations like the non-linear Fokker-Planck equation (5.2.1) or the Derrida-Lebowitz-Speer-Spohn equation (5.3.1) cast in the L^2-Wasserstein formalism. The first part is based on the joint work with D. Matthes [70], the second part is based on my own work [81], and the third part was developed during the preparation of this thesis.

Main Idea in Short. Generally speaking, we study the approximation of curves of steepest descent in the energy landscape of a functional $\mathcal{E} : \mathbf{X} \to \mathbb{R} \cup \{\infty\}$ with respect to a metric $\boldsymbol{d}$ on $\mathbf{X}$. Before we elaborate on our motivation and results, we briefly outline the concept in the simplest setting, namely when $\mathbf{X} = \mathbb{R}^{\mathrm{d}}$, $\boldsymbol{d}$ is the Euclidean metric, and $\mathcal{E} \in \mathcal{C}_c^\infty(\mathbb{R}^{\mathrm{d}})$, in which case the problem amounts to approximate solutions to

$$\dot{u} = -\nabla\mathcal{E}(u). \tag{5.0.1}$$

With these strong assumptions on $\mathcal{E}$, it follows that the second order Backward Differentiation Formula (BDF2) method with any sufficiently small uniform time step $\tau > 0$,

$$\frac{3u_k^\tau - 4u_{k\text{-}1}^\tau + u_{k\text{-}2}^\tau}{2\tau} = -\nabla\mathcal{E}(u_k^\tau), \tag{5.0.2}$$

is well-defined and convergent for well-prepared initial data (u_0^τ, u_1^τ). It is further well-known that this is a second order approximation of the true solution u_t to (5.0.1), i.e., $u_k^\tau = u(k\tau) + O(\tau^2)$ as $\tau \to 0$, see e.g. [11]. Hence, the strength of the BDF2 method in comparison to the implicit Euler scheme is that the former — at least in the smooth setting at hand — converges to second order in τ.

The strategy for our own convergence analysis in the abstract metric space case is the variational formulation of the BDF2 method (5.0.2) which is inspired by the Minimizing Movement scheme from [4, 29, 30]. Therefore, we proposed the following scheme to construct a discrete approximation $(u_k^\tau)_{k\in\mathbb{N}}$ of a curve of steepest descent emerging from the initial datum $u_0 \in \mathbf{X}$ for metric gradient flows: For a given pair of initial conditions (u_0^τ, u_1^τ) that approximate u_0 define inductively a discrete solution $(u_k^\tau)_{k\in\mathbb{N}}$ such that

$$u_{k+1}^\tau \in \operatorname*{argmin}_{w\in\mathbf{X}} \frac{1}{\tau}\boldsymbol{d}^2(u_k^\tau, w) - \frac{1}{4\tau}\boldsymbol{d}^2(u_{k\text{-}1}^\tau, w) + \mathcal{E}(w). \tag{5.0.3}$$

Note, in the euclidean setting the minimizer u_k^τ satisfies the BDF2 recursion (5.0.2).

Contribution. Studies on the BDF2 scheme in the ODE setting, when $\mathbf{X} = \mathbb{R}^d$ and $\boldsymbol{d}$ is the Euclidean metric have been an active topic in the 1950's and 1960's [26, 27]. Subsequently, BDF2-based problem adapted methods for the integration of flows on (finite-dimensional) submanifolds $\mathbf{X}$, e.g., for ODEs with constraints, have been proposed and analyzed, see e.g. [25, 44]. More recently, the method in the Hilbertian setup, where $\mathbf{X}$ is a Hilbert space and $\boldsymbol{d}$ is induced by the norm, has attracted a lot of attention. There is now a rich literature on convergence results, particularly for very general nonlinear right-hand sides, see e.g. [1, 2, 5, 13, 35, 36, 37, 52, 59, 62, 90]. The analysis appears to be more or less complete now, at least under reasonable conditions on the nonlinearity.

We are apparently the first to analyze (a variational formulation of) the BDF2 method *for approximation of gradient flows in abstract metric spaces*, and to prove its convergence just under the hypothesis of semi-convexity in the abstract metric space case and just under mild assumptions on the confinement potential V and the interaction kernel W in the case of displacement-λ-convex flows in the L^2-Wasserstein space. Our proof is different from the one in ODE textbooks [11, 43, 51], also from the ones typically given in the Hilbertian setting, like in [35]. The key difference is that due to the possible "roughness" of the metric space $\mathbf{X}$, there is no appropriate notion of smooth solution for the gradient flow (in general, there does not even exist a definition for the differentiability of a curve). Hence, we cannot invoke error estimates that rely on Taylor expansions around the limiting solution.

Method in the Abstract Metric Space Case. These difficulties of roughness are already present in the convergence analysis of the implicit Euler method in metric spaces, and have been overcome in [4] by formulating all essential estimates in a robust way that requires no smoothness. Naturally, the strategy for our own convergence analysis of the variational BDF2 method is inspired by that from [4], and there are various similarities also on the technical level. For instance, being unable to estimate the error between the genuine and the approximating discrete solutions directly, we resort to a Cauchy-type argument that compares discrete solutions with different time steps as in [4, section 4.1]. Further, the basis for the control of the local error is a convexity inequality for the variational functional, which estimates the change of distance to a fixed "observer point" during one iteration: this is [4, Corollary 4.1.3] for the implicit Euler method, and (5.1.17) for the BDF2 method. While the accumulation of the global error is relatively easy to control for the one-step Euler discretization, see [4, section 4.4], this is an extremely tedious piece of work for our two-step method.

Our approach yields a control on the global approximation error, which is not of order two but only of order one-half. We also provide an example to show that indeed, even for specific, seemingly harmless choices of $(\mathbf{X}, \boldsymbol{d})$ and $\mathcal{E}$, convergence takes place at first order only. In view of the results in [4, section 4.4] on the implicit Euler method, it seems likely that our variational BDF2 converges to first order in general. Currently, we are not able to close the apparent gap between order one-half and order one. There is little hope to adapt the methods leading to improved convergence in [4, section 4.4], since there, profound properties of the Yosida regularization play a pivotal role in estimating the local error. No comparable estimates are known for our BDF2 functional, and it seems unlikely that an appropriate surrogate exists e.g. for the duality formula for the slope [4, Lemma 3.1.5]. And according to our general philosophy, that we describe below, any further investigations in the direction of improving the rate beyond one-half appear rather pointless.

We emphasize that the proven slow convergence order one-half does not contradict our initial intention of providing a method of faster convergence than the implicit Euler one. Indeed, *if* the approximated solution is smooth enough (which, in specific situations, can often be verified a posteriori by considering it in a different setting or in an ambient space), *then* the classical convergence proofs from textbooks apply and yield the desired rate of order two. That philosophy is justified by a series of numerical experiments that all show second order convergence. *Our contribution is that — regardless of the regularity of the limiting solution under consideration — convergence of the method is guaranteed, even with an explicit rate. And our proof utilizes solely the variational structure of the scheme* (5.0.2) *and the semi-convexity hypothesis on* $\mathcal{E}$.

Method in the L^2-Wasserstein Case. Our main contribution of the second and third section of this chapter is to improve the convergence result of [70] from weak to strong convergence of the discrete solution $(\rho_k^{\tau})_{k\in\mathbb{N}}$ when one wants to approximate solutions to the non-linear Fokker-Planck equation (5.2.1) or to the Derrida-Lebowitz-Speer-Spohn equation (5.3.1) by means of the variational formulation of the BDF2 method. Also in contrast to [70], our approach is independent of the uniform semi-convexity of the augmented energy functional on the right-hand side of (5.0.3). More in the spirit of the original works on the linear Fokker-Planck equation of Kinderlehrer et al. [54], we solely utilize the differential structure of both the L^2-Wasserstein space and of the augmented energy functional.

Note, the BDF2 method and the techniques presented here have two further possible applications. Firstly, the formally higher-order approximation in time is expected to improve the performance of numerical simulations due to the better resolution of the solution with respect to a coarser time grid. Secondly, PDEs with gradient flow structure such that the energy function $\mathcal{E}$ do not possess any uniform semi-convexity property – like the Hele-Shaw equation seen as L^2-Wasserstein gradient flows – are not covered in [70] nor in this chapter. However, as long as the subdifferential calculus in the L^2-Wasserstein space is applicable to $\mathcal{E}$ our method is feasible. With this technique at hand one can compute from (5.0.3) the discrete Euler-Lagrange equations for the discrete approximation by variations along solutions to the continuity equation (likewise theorem 5.2.9). Hence, having sufficiently good regularity estimates for the discrete solution, passing to the limit as $\boldsymbol{\tau}$ tends to zero could yield directly a distributional solution for the aforementioned class of PDEs without using the abstract theory of curves of steepest descent for λ-contractive gradient flows.

In conclusion, the BDF2 method provides a structure-preserving numerical scheme of formally higher-order approximation in time with a strong notion of convergence for drift-diffusion equations like (5.2.1) *or* (5.3.1).

Main Results in the Abstract Metric Space Case. Our main result concerning the well-posedness and the limit behavior is given as follows. In the abstract metric space case define the *interpolated solution* $\overline{u}_t^{\boldsymbol{\tau}} : [0, \infty) \to \mathbf{X}$, given by

$$\overline{u}_0^{\boldsymbol{\tau}} = u_0^{\boldsymbol{\tau}}, \quad \overline{u}_t^{\boldsymbol{\tau}} = u_k^{\boldsymbol{\tau}} \quad \text{for } t \in ((k-1)\tau, k\tau] \quad \text{and } k \in \mathbb{N}.$$

The limit-behavior as the time step size $\boldsymbol{\tau} \to 0$ of the equidistant partition $\boldsymbol{\tau} = (\tau, 2\tau, \ldots)$ is stated in the following theorem.

Theorem 5.0.1. *Assume $(\mathbf{X}, \boldsymbol{d})$ is a complete, separable metric space, the free energy functional $\mathcal{E}$ satisfies (E1)–(E3), specified in Assumption 5.1.1, and given the equidistant partition $\boldsymbol{\tau} = (\tau, 2\tau, 3\tau, \ldots)$ with step size $\tau \in (0, \tau_*)$. Then, the following statements holds:*

a) ***Existence of Discrete Solutions.*** *For each approximation $(u_0^{\boldsymbol{\tau}}, u_1^{\boldsymbol{\tau}})$ of the initial datum $u_0 \in \mathcal{D}(\mathcal{E})$ satisfying (I1) as defined in Assumption 5.1.3 one obtains a unique discrete solution $(u_k^{\boldsymbol{\tau}})_{k\in\mathbb{N}}$.*

b) ***Step Size Independent Estimates.*** *For fixed time horizon $T > 0$, there is a constant C, depending only on d_1, d_2 and T, such that the corresponding discrete solutions $(u_k^{\boldsymbol{\tau}})_{k\in\mathbb{N}}$ satisfy for all $N \in \mathbb{N}$ with $N\tau \leq T$:*

$$\sum_{k=1}^{N} \frac{1}{2\tau} \boldsymbol{d}^2(u_{k\text{-}1}^{\boldsymbol{\tau}}, u_k^{\boldsymbol{\tau}}) \leq C, \qquad |\mathcal{E}(u_N^{\boldsymbol{\tau}})| \leq C, \qquad \boldsymbol{d}^2(u_*, u_N^{\boldsymbol{\tau}}) \leq C.$$

Furthermore, consider a sequence of equidistant partitions $\boldsymbol{\tau}_n = (\tau_n, 2\tau_n, 3\tau_n, \ldots)$ with vanishing step sizes $\tau_n \in (0, \tau_*)$ which are strictly decreasing, and which are such that the quotients τ_n/τ_{n+1} are natural numbers. Let a sequence of initial data $(u_0^{\boldsymbol{\tau}_n}, u_1^{\boldsymbol{\tau}_n})_{n\in\mathbb{N}}$ be given that satisfy (I1) as defined in Assumption 5.1.3, and such that $u_0^{\boldsymbol{\tau}_n} \xrightarrow{\boldsymbol{d}} u_0$. Then,

c) ***Convergence.*** *There exists $u_t^* \in \mathrm{AC}^2(0, \infty; (\mathbf{X}, \boldsymbol{d}))$ such that the sequence of piecewise constant interpolations $(\overline{u}_t^{\boldsymbol{\tau}_n})_{n\in\mathbb{N}}$ converges locally uniformly with respect to time to u_t^*.*

d) ***Convergence Rate.*** *More precisely, for every time horizon $T > 0$, there is a constant C that can be expressed in terms of d_1, d_2 and T alone, such that for all $t \in [0, T]$:*

$$\boldsymbol{d}\big(\overline{u}_t^{\boldsymbol{\tau}_n}, u_t^*\big) \leq C \left(\boldsymbol{d}(u_0^{\boldsymbol{\tau}_n}, u_0) + \sqrt{\tau_n}\right). \tag{5.0.4}$$

e) ***Solution of the Gradient Flow.*** *The limit curve u_t^* from c) is a solution of the gradient flow for $\mathcal{E}$ in the sense of the definition the EVI* (2.3.2).

Main Result L^2-Wasserstein case. In the L^2-Wasserstein case we denote by $(\rho_k^{\boldsymbol{\tau}})_{k\in\mathbb{N}}$ the *discrete solution* and by $\overline{\rho}_t^{\boldsymbol{\tau}}$ the *interpolated solution*. Our first result on the approximation of the solution to the non-linear Fokker-Planck equation (5.2.1) reads than as. Note, this case the free energy functional $\mathcal{E}$ is given by (5.2.2).

Theorem 5.0.2 (Non-linear Fokker-Planck Equation)**.** *Let $\Omega \subset \mathbb{R}^{\mathrm{d}}$ be either an open, bounded, and connected domain with Lipschitz continuous boundary $\partial\Omega$ or let $\Omega = \mathbb{R}^{\mathrm{d}}$. Further, assume $m \geq 1$ and that V and W satisfy (F1)&(F2) as specified in Assumption 5.2.1. Given an equidistant partition $\boldsymbol{\tau} = (\tau, 2\tau, 3\tau, \ldots)$ with step size $\tau \in (0, \tau_*)$ and an approximation $(\rho_0^{\boldsymbol{\tau}}, \rho_1^{\boldsymbol{\tau}})$ of the initial datum ρ_0 satisfying (I1)&(I2) defined in Assumption 5.2.2. Then, the following statements hold:*

a) ***Existence of the Discrete Solutions.*** *There exists a sequence $(\rho_k^{\boldsymbol{\tau}})_{k\in\mathbb{N}}$ satisfying the BDF2 scheme* (5.0.3)*, which satisfies the step size independent bounds* (5.2.10) *on the kinetic energy, on the internal energy, and on the second moments.*

b) ***Step Size Independent $\mathrm{L}^2(0,T;\mathrm{BV}(\Omega))$-estimate.*** *For each fixed time horizon $T > 0$ there exists a non-negative constant C, depending only on m, V, W, and T such that for each $\tau \in (0, \tau_*)$:*

$$\|(\overline{\rho}_t^{\boldsymbol{\tau}})^m\|_{\mathrm{L}^2(0,T;\mathrm{BV}(\Omega))} \leq C.$$

Given a vanishing sequence $(\tau_n)_{n\in\mathbb{N}}$ of step sizes $\tau_n \in (0, \tau_*)$ and initial data $(\rho_0^{\boldsymbol{\tau}_n}, \rho_1^{\boldsymbol{\tau}_n})$ satisfying Assumption 5.2.2, then:

c) ***Narrow Convergence in $\mathcal{P}_2(\Omega)$.*** *There exists a (non-relabelled) subsequence $(\tau_n)_{n\in\mathbb{N}}$ and a limit curve $\rho_* \in \mathrm{AC}^2(0,\infty;(\mathcal{P}_2(\Omega),\mathbf{W}_2))$ such that for any $t \geq 0$:*

$$\overline{\rho}_t^{\boldsymbol{\tau}_n} \rightharpoonup \rho_*(t) \qquad \text{narrowly in the space } \mathcal{P}_2(\Omega) \text{ as } n \to \infty.$$

d) ***Strong Convergence in $\mathrm{L}^m(0,T;\mathrm{L}^m(\Omega))$.*** *With the notations from c), there exists a further (non-relabelled) subsequence $(\tau_n)_{n\in\mathbb{N}}$ such that for all $T > 0$ and any bounded subset $\Theta \subseteq \Omega$:*

$$\overline{\rho}_t^{\boldsymbol{\tau}_n} \to \rho_* \qquad \text{strongly in } \mathrm{L}^m(0,T;\mathrm{L}^m(\Theta)) \text{ as } n \to \infty.$$

e) ***Solution of the Non-linear Fokker-Planck Equation.*** *The limit curve ρ_t^* from c) satisfies the non-linear Fokker-Planck equation with no-flux boundary condition* (5.2.1) *in the following weak sense: For each test function $\varphi_t \in C_c^\infty([0,\infty) \times \overline{\Omega})$ with $\nabla\varphi_t \cdot \boldsymbol{n} = 0$ on $\partial\Omega$ the limit function ρ_t^* satisfies:*

$$\begin{aligned}&\int_0^\infty \int_\Omega -\Delta\varphi_t\,(\rho_t^*)^m + \langle\nabla\varphi_t, \nabla V\rangle\,\rho_t^* + \langle\nabla\varphi_t, \nabla W * \rho_t^*\rangle\,\rho_t^*\,\mathrm{d}x\,\mathrm{d}t\\ =&\int_0^\infty \int_\Omega \partial_t\varphi_t\,\rho_*\,\mathrm{d}x\,\mathrm{d}t + \int_\Omega \varphi_0\,\rho_0\,\mathrm{d}x.\end{aligned}$$

Our second results on the approximation of solutions to the Derrida-Lebowitz-Speer-Spohn equation (5.3.1) reads than as. Here, the free energy functional $\mathcal{E}$ is given by the Fisher information $\mathcal{I}$ as defined in (2.4.5).

Theorem 5.0.3 (Derrida-Lebowitz-Speer-Spohn Equation). *Let $\Omega \subset \mathbb{R}^{\mathrm{d}}$ be an open, bounded, and convex domain with Lipschitz continuous boundary $\partial\Omega$ or let $\Omega = \mathbb{R}^{\mathrm{d}}$. Given an equidistant partition $\boldsymbol{\tau} = (\tau, 2\tau, 3\tau, \ldots)$ with step size $\tau \in (0, \tau_*)$ and an approximation $(\rho_0^{\boldsymbol{\tau}}, \rho_1^{\boldsymbol{\tau}})$ of the initial datum ρ_0 satisfying (I1)&(I2) defined in assumption 5.3.1. Then, the following statements hold:*

a) ***Existence of the Discrete Solutions.*** *There exists a sequence $(\rho_k^{\boldsymbol{\tau}})_{k\in\mathbb{N}}$ satisfying the BDF2 scheme (5.0.3), which satisfies the step size independent bounds (5.3.3) on the kinetic energy, on the Fisher information, and on the second moments.*

b) ***Step Size Independent $\mathrm{L}^2(0,T;\mathrm{H}^2(\Omega))$-estimate.*** *For each fixed time horizon $T > 0$ there exists a non-negative constant C, depending only on Ω, and T such that for each $\tau \in (0, \tau_*)$:*

$$\|\sqrt{\rho_t^{\boldsymbol{\tau}}}\|_{\mathrm{L}^2(0,T;\mathrm{H}^2(\Omega))} \leq C.$$

Given a vanishing sequence $(\tau_n)_{n\in\mathbb{N}}$ of step sizes $\tau_n \in (0, \tau_*)$ and initial data $(\rho_0^{\boldsymbol{\tau}_n}, \rho_1^{\boldsymbol{\tau}_n})$ satisfying Assumption 5.3.1, then:

c) ***Narrow Convergence in $\mathcal{P}_2(\Omega)$.*** *There exists a (non-relabelled) subsequence $(\tau_n)_{n\in\mathbb{N}}$ and a limit curve $\rho_* \in \mathrm{AC}^2(0,\infty;(\mathcal{P}_2(\Omega), \mathbf{W}_2))$ such that for any $t \geq 0$:*

$$\overline{\rho}_t^{\boldsymbol{\tau}_n} \rightharpoonup \rho_*(t) \qquad \text{narrowly in the space } \mathcal{P}_2(\Omega) \text{ as } n \to \infty.$$

d) ***Strong Convergence.*** *With the notations from c), there exists a further (non-relabelled) subsequence $(\tau_n)_{n\in\mathbb{N}}$ such that for all $T > 0$ and any $p \geq 1$:*

$$\begin{aligned} \overline{\rho}_t^{\boldsymbol{\tau}_n} &\to \rho_t^* && \text{strongly in } \mathrm{L}^p(0,T;\mathrm{L}^1(\Omega)) \text{ as } n \to \infty, \\ \sqrt{\overline{\rho}_t^{\boldsymbol{\tau}_n}} &\to \sqrt{\rho_t^*} && \text{strongly in } \mathrm{L}^2(0,T;\mathrm{H}^1(\Omega)) \text{ as } n \to \infty, \\ \sqrt{\overline{\rho}_t^{\boldsymbol{\tau}_n}} &\rightharpoonup \sqrt{\rho_t^*} && \text{weakly in } \mathrm{L}^2(0,T;\mathrm{H}^2(\Omega)) \text{ as } n \to \infty. \end{aligned}$$

e) ***Solution of the Derrida-Lebowitz-Speer-Spohn Equation.*** *The limit curve ρ_t^* from c) satisfies the Derrida-Lebowitz-Speer-Spohn equation with no-flux boundary condition (5.2.1) in the following weak sense: For each test function $\varphi_t \in C_c^\infty([0,\infty) \times \overline{\Omega})$ with $\nabla\varphi_t \cdot \boldsymbol{n} = 0$ on $\partial\Omega$ the limit curve ρ_t^* satisfies:*

$$\begin{aligned} &-\int_0^\infty \int_\Omega \frac{1}{2}\langle \nabla(\Delta\varphi_t), \nabla\rho_t^*\rangle + 2\langle \operatorname{Hess}\varphi_t \nabla\sqrt{\rho_t^*}, \nabla\sqrt{\rho_t^*}\rangle \,\mathrm{d}x \\ &= \int_0^\infty \int_\Omega \partial_t\varphi_t \,\rho_* \,\mathrm{d}x\,\mathrm{d}t + \int_\Omega \varphi_0\,\rho_0\,\mathrm{d}x. \end{aligned}$$

Analytic (Counter-)Example. We give a simple example showing that under the given assumptions, in general one cannot expect second order convergence of the BDF2 method, i.e., $u_k^{\boldsymbol{\tau}} = u(k\tau) + \mathcal{O}(\tau^2)$ in place of $u_k^{\boldsymbol{\tau}} = u(k\tau) + \mathcal{O}(\sqrt{\tau})$ in (5.1.18). Our example is placed on the (very regular) metric space $\mathbf{X} = \mathbb{R}$ with the usual distance, with the convex but not globally differentiable potential $\mathcal{E}$ defined by $\mathcal{E}(u) := u\,\mathbf{1}_{\{u\geq 0\}}$. The associated gradient flow with initial condition $u_0 = 1$ is the continuous curve $u_t^* := (1-t)\,\mathbf{1}_{\{0\leq t\leq 1\}}$ that fails to be differentiable at $t = 1$.

The solution $u_{k+1}^{\boldsymbol{\tau}}$ to the kth minimization problem in (5.0.3) is elementary to compute — making a case distinction whether the minimizer is positive, negative or zero — and explicitly given by

$$u_{k+1}^{\boldsymbol{\tau}} = \begin{cases} \frac{4}{3}u_k^{\boldsymbol{\tau}} - \frac{1}{3}u_{k-1}^{\boldsymbol{\tau}} - \frac{2}{3}\tau & \text{if that expression is positive,} \\ \frac{4}{3}u_k^{\boldsymbol{\tau}} - \frac{1}{3}u_{k-1}^{\boldsymbol{\tau}} & \text{if that expression is negative,} \\ 0 & \text{otherwise, i.e., if } 0 \leq \frac{4}{3}u_k^{\boldsymbol{\tau}} - \frac{1}{3}u_{k-1}^{\boldsymbol{\tau}} \leq \frac{2}{3}\tau. \end{cases} \tag{5.0.5}$$

One easily concludes that for the initial conditions $u_0^{\boldsymbol{\tau}} = 1$ and $u_1^{\boldsymbol{\tau}} = 1-\tau$, the kth approximation equals $u_{k+1}^{\boldsymbol{\tau}} = 1-(k+1)\tau$ as long as that expression is positive. Indeed, one has

$$\frac{4}{3}u_k^{\boldsymbol{\tau}} - \frac{1}{3}u_{k-1}^{\boldsymbol{\tau}} - \frac{2}{3}\tau = \frac{4}{3}\big(1-k\tau\big) - \frac{1}{3}\big(1-(k-1)\tau\big) - \frac{2}{3}\tau = 1-(k+1)\tau > 0,$$

so the first case in the recursion (5.0.5) applies. Accordingly, let $N_{\boldsymbol{\tau}}$ be the smallest index $k \geq 1$ for which $k\tau \geq 1$. For simplicity, we assume that $u_{N_{\boldsymbol{\tau}}}^{\boldsymbol{\tau}} = 0$, i.e., that the third case in (5.0.5) applies:

$$1 - N_{\boldsymbol{\tau}}\tau = \frac{4}{3}u_\tau^{N_{\boldsymbol{\tau}}-1} - \frac{1}{3}u_\tau^{N_{\boldsymbol{\tau}}-2} - \frac{2}{3}\tau \in \big[-\frac{2}{3}\tau, 0\big]. \tag{5.0.6}$$

The other case, in which $-\tau < 1 - N_\tau\tau < -\frac{2}{3}\tau$, leads to a similar result, but with more complicated formulae. Recalling that the two-step recursion $a_{k+1} = \frac{4}{3}a_k - \frac{1}{3}a_{k-1}$ has the general solution $a_k = p + 3^{-k}q$ with real parameters p and q, one easily deduces from (5.0.5) in combination with $u_{N_{\boldsymbol{\tau}}}^{\boldsymbol{\tau}} = 0$ and $u_{N_{\boldsymbol{\tau}}-1}^{\boldsymbol{\tau}} = 1-(N_{\boldsymbol{\tau}}-1)\tau \in [\frac{1}{3}\tau, \tau]$ because of (5.0.6) that

$$u_k^{\boldsymbol{\tau}} = \frac{4}{3}u_{k-1}^{\boldsymbol{\tau}} - \frac{1}{3}u_{k-2}^{\boldsymbol{\tau}} = -\frac{1}{2}\big(1-3^{-(k-N_{\boldsymbol{\tau}})}\big)u_\tau^{N_{\boldsymbol{\tau}}-1} \leq -\frac{1}{6}\big(1-3^{-(k-N_{\boldsymbol{\tau}})}\big)\tau < 0$$

for each index $k > N_\tau$. In conclusion, we have exact approximation for $t < 1$, i.e., $u_k^{\boldsymbol{\tau}} = u_{k\tau}^*$ for every k with $k\tau < 1$, but a residual of order τ at every point $t > 1$: with indices $k_t^{\boldsymbol{\tau}}$ chosen such that one has $k_t^{\boldsymbol{\tau}}\tau \to t > 1$ as $\tau \to 0$, it follows that

$$\lim_{\tau\to 0} \frac{u_t^* - u_{k_t^{\boldsymbol{\tau}}}^{\boldsymbol{\tau}}}{\tau} \geq \frac{1}{6}\lim_{\tau\to 0}\big(1-3^{-(k_t^{\boldsymbol{\tau}}-N_{\boldsymbol{\tau}})}\big) = \frac{1}{6}.$$

This clearly excludes the possibility of second order convergence.

5.1 Application to Gradient Flows in Abstract Metric Space

This section is based on the joint-work with D. Matthes [70]. For definiteness, we are working in this section inside the abstract framework developed in the first part of the book [4]; given a separable, complete metric space $(\mathbf{X}, \boldsymbol{d})$ we are seeking to approximate solutions to the gradient flow equation

$$\dot{u}_t = -\nabla_{\mathbf{X}}\mathcal{E}(u_t), \qquad u_0 = u_0$$

by means of the second order Backward Differentiation Formula

$$\frac{3u_k^{\boldsymbol{\tau}} - 4u_{k\text{-}1}^{\boldsymbol{\tau}} + u_{k\text{-}2}^{\boldsymbol{\tau}}}{2\tau} = -\nabla\mathcal{E}(u_k^{\boldsymbol{\tau}}).$$

Although our considerations are very general, we have three specific settings in mind. The first is that of gradient flows for smooth functions on a finite dimensional compact manifold, the second concerns uniformly semi-convex functionals on Hilbert spaces, and in the third, we consider flows for uniformly displacement semi-convex functionals on the L^2-Wasserstein space $(\mathcal{P}_2(\Omega), \mathbf{W}_2)$.

Method. We propose the following construction of a discrete approximation $(u_\tau^k)_{k\in\mathbb{N}}$ via a variational formulation of the BDF2 method:

Scheme. For each equidistant partition $\boldsymbol{\tau} = (\tau, 2\tau, 3\tau, \ldots)$ with sufficiently small time step $\tau > 0$, let a pair of initial conditions $(u_0^{\boldsymbol{\tau}}, u_1^{\boldsymbol{\tau}})$ be given that approximate u_0. Then define inductively a discrete solution $(u_k^{\boldsymbol{\tau}})_{k\in\mathbb{N}}$ such that each $u_{k+1}^{\boldsymbol{\tau}}$ with $k \in \mathbb{N}$ is a minimizer of the following functional,

$$w \mapsto \Psi(\tau, u_{k\text{-}1}^{\boldsymbol{\tau}}, u_k^{\boldsymbol{\tau}}; w) := \frac{1}{\tau}\boldsymbol{d}^2(u_k^{\boldsymbol{\tau}}, w) - \frac{1}{4\tau}\boldsymbol{d}^2(u_{k\text{-}1}^{\boldsymbol{\tau}}, w) + \mathcal{E}(w).$$

Define the corresponding piecewise constant interpolation in time $\overline{u}_t^{\boldsymbol{\tau}} : [0, \infty) \to \mathbf{X}$ of the discrete solution $u_k^{\boldsymbol{\tau}}$ in time via

$$\overline{u}_0^{\boldsymbol{\tau}} = u_0^{\boldsymbol{\tau}}, \quad \overline{u}_t^{\boldsymbol{\tau}} = u_k^{\boldsymbol{\tau}} \quad \text{for } t \in ((k-1)\tau, k\tau] \quad \text{and } k \in \mathbb{N}.$$

Strategy of the Proof. The main idea of our convergence analysis is to exploit the λ-convexity of the free energy functional $\mathcal{E}$ and of the BDF2 penalization Ψ. The precise definition and some examples satisfying this assumption are contained in section 5.1.1. This specific convexity assumption of the BDF2 penalization Ψ and the variational formulation of the scheme allows us to derive existence of the discrete solution $(\rho_k^{\boldsymbol{\tau}})_{k\in\mathbb{N}}$ and further intrinsic properties like the almost energy diminishing property (5.1.11), the classical stability estimates (5.1.14), and the time-discrete version of the EVI (5.1.17), see section 5.1.2. The latter is the surrogate of a time-discrete Euler-Lagrange equation and comprises a comparison principle which is the essential ingredient in the proof of convergence of the approximation $\overline{\rho}_t^{\boldsymbol{\tau}}$, see section 5.1.4. This comparison principle allows us to derive the sought-for (sub-optimal) convergence rate of order one-half as $\boldsymbol{\tau} \to 0$.

5.1.1 Setup and Assumptions

Fix a complete, separable metric space $(\mathbf{X}, \boldsymbol{d})$ and define the *BDF2 penalization* $\Psi : (0, \tau_*) \times \mathbf{X} \times \mathbf{X} \times \mathbf{X} \to \mathbb{R} \cup \{\infty\}$ of $\mathcal{E}$ by

$$\Psi(\tau, u, v; \cdot) : \mathbf{X} \to \mathbb{R} \cup \{\infty\}; \quad \Psi(\tau, u, v; w) := \frac{1}{\tau}\boldsymbol{d}^2(v, w) - \frac{1}{4\tau}\boldsymbol{d}^2(u, w) + \mathcal{E}(w).$$

The *discrete solution* (for $\mathcal{E}$ on $(\mathbf{X}, \boldsymbol{d})$) corresponding to a time step size $\tau \in (0, \tau_*)$ and pair of initial data $(u_0^\tau, u_1^\tau) \in \mathbf{X} \times \mathbf{X}$ is the sequence $(u_k^\tau)_{k\in\mathbb{N}}$, which is inductively obtained via

$$u_{k+1}^\tau \in \underset{w\in\mathbf{X}}{\operatorname{argmin}}\ \Psi(\tau, u_{k-1}^\tau, u_k^\tau; w) \tag{5.1.1}$$

for $k \in \mathbb{N}$. From now on, we shall work with the following assumptions on the free energy functional $\mathcal{E}$.

Assumption 5.1.1. The free energy functional $\mathcal{E} : \mathbf{X} \to \mathbb{R} \cup \{\infty\}$ is proper and satisfies the following regularity conditions:

(E1) **Semi-continuity.** $\mathcal{E}$ is sequentially lower semi-continuous on $(\mathbf{X}, \boldsymbol{d})$:

$$u_n \stackrel{\boldsymbol{d}}{\to} u \qquad \Longrightarrow \qquad \mathcal{E}(u) \le \liminf_{n\to\infty} \mathcal{E}(u_n).$$

(E2) **Coercivity.** There exist $\tau_* > 0$ and $u_* \in \mathbf{X}$, such that

$$c_* := \inf_{v\in\mathbf{X}} \frac{1}{2\tau_*}\boldsymbol{d}^2(u_*, v) + \mathcal{E}(v) > -\infty.$$

(E3) **Semi-convexity.** There exists a constant λ such that for every $u, v, \gamma_0, \gamma_1 \in \mathcal{D}(\mathcal{E})$ and every $\tau \in [0, \tau_*)$, there exists a continuous curve $\gamma_s : [0, 1] \to \mathbf{X}$ joining the given end points γ_0 and γ_1, along which the penalized energy Ψ satisfies

$$\begin{aligned}\Psi(\tau, u, v; \gamma_s) \le &(1 - s)\Psi(\tau, u, v; \gamma_0) + s\Psi(\tau, u, v; \gamma_1) \\ &- \frac{1}{2}\Big(\frac{3}{2\tau} + \lambda\Big)s(1 - s)\boldsymbol{d}^2(\gamma_0, \gamma_1).\end{aligned} \tag{5.1.2}$$

Moreover, without loss of generality, we assume that

$$\lambda \le 0 \quad \text{and} \quad (-\lambda)\tau_* \le \frac{1}{2} \quad \text{and} \quad \tau_* \le 1. \tag{5.1.3}$$

Note that for $0 < \tau < \tau_*$, the last term on the right hand side of (5.1.2) is positive for $\gamma_0 \neq \gamma_1$ and $0 < s < 1$, implying that $s \mapsto \Psi(\tau, u, v; \gamma_s)$ is strictly convex.

Remark 5.1.2. Assumptions (E1)&(E2) are standard minimal hypotheses on the energy in the context of metric gradient flows. (E3) plays an analogous role for the BDF2 discretization as Assumption 4.0.1 in [4] plays for the minimizing movement scheme.

Lastly, we have to specify our assumptions on the approximation (u_0^τ, u_1^τ) of the initial datum u_0, which will be of interest in derivation of the classical estimates.

Assumption 5.1.3. The approximation (u_0^τ, u_1^τ) of the initial datum $u_0 \in \mathcal{D}(\mathcal{E})$ satisfy:

(I1) There are constants d_1 and d_2, such that for all $\tau \in (0, \tau_*)$:

$$\mathcal{E}(u_0^\tau) \leq d_1, \qquad \mathcal{E}(u_1^\tau) \leq d_1, \qquad \text{and} \qquad \boldsymbol{d}(u_0^\tau, u_1^\tau) \leq d_2\tau.$$

Examples

In this part we discuss three general situations in which the convexity assumption (E3) is satisfied, namely that of uniformly semi-convex functionals $\mathcal{E}$ on a Hilbert space H, that of semi-convex C^1-functions $\mathcal{E}$ on Riemannian manifolds of non-negative cross-curvature, and that of functionals $\mathcal{E}$ on the L^2-Wasserstein space $(\mathcal{P}_2(\Omega), \mathbf{W}_2)$ that are uniformly displacement semi-convex.

Hilbert Spaces. Our first results concerns uniformly semi-convex functionals on Hilbert spaces. This class provides fairly easy examples for the validity of assumption (E3), thanks to the linear structure of the space.

Theorem 5.1.4. *Assume that the metric space $(\mathbf{X}, \boldsymbol{d})$ is a Hilbert space $\mathbf{X} = H$, with the distance $\boldsymbol{d}$ induced by the norm $\|\cdot\|$. Assume further that $\mathcal{E}$ is uniformly semi-convex with modulus λ. Then (E3) is satisfied, with γ_s the straight line between γ_0, γ_1 and with the same λ.*

Proof. Let $\gamma_0, \gamma_1 \in \mathcal{D}(\mathcal{E})$ as well as $u, v \in \mathcal{D}(\mathcal{E})$ and $\tau > 0$ be given. We verify (5.1.2) for the particular curve $\gamma_s := (1-s)\gamma_0 + s\gamma_1$. On the one hand, by the convexity hypothesis on $\mathcal{E}$, we know that

$$\mathcal{E}(\gamma_s) \leq (1-s)\mathcal{E}(\gamma_0) + (1-s)\mathcal{E}(\gamma_1) - \frac{\lambda}{2}s(1-s)\|\gamma_0 - \gamma_1\|^2. \tag{5.1.4}$$

On the other hand, a direct calculation using the property of the scalar product yields

$$\begin{aligned} &\|\gamma_s - v\|^2 - \frac{1}{4}\|\gamma_s - u\|^2 \\ =&(1-s)\big(\|\gamma_0 - v\|^2 - \frac{1}{4}\|\gamma_0 - u\|^2\big) + s\big(\|\gamma_1 - v\|^2 - \frac{1}{4}\|\gamma_1 - u\|^2\big) \\ &- \frac{3}{4}s(1-s)\|\gamma_0 - \gamma_1\|^2. \end{aligned} \tag{5.1.5}$$

Adding $\frac{1}{\tau}$ times (5.1.5) to (5.1.4) yields (5.1.2). □

Riemannian Manifolds. Another situation of interest is that of the gradient flow on a compact smooth Riemannian manifold $(\mathcal{M}, \mathbf{g})$, which is induced by a semi-convex function $\mathcal{E} \in C^1(\mathcal{M})$. Here, our very general approach is clearly not optimal: in that finite-dimensional setting, gradient flows can be characterized in a direct way instead of using the EVI (2.3.2). Further, there are explicit and local variants of the BDF2 method (avoiding the global minimization of Ψ in each time step), see e.g. [50], which are simpler to implement, and whose convergence is expected under more easily verifiable hypotheses than (E3). Still, for the sake of completeness, we shall detail a sufficient criterion for the applicability of our results in that situation.

To indicate why the verification of (E3) indeed poses a (surprisingly hard) problem, observe that it is in general *not* possible to use the geodesic $\widetilde{\gamma}_s$ for the curve connecting γ_0 to γ_1 in (5.1.2). Indeed, for $s \mapsto \Psi(\tau, u, v; \widetilde{\gamma}_s)$ to be uniformly convex of modulus $\frac{3}{2\tau} + \lambda$, independently of u and v, one would essentially need that both $s \mapsto \boldsymbol{d}^2(v, \gamma_s)$ and $s \mapsto -\boldsymbol{d}^2(u, \gamma_s)$ are uniformly convex of modulus $\boldsymbol{d}^2(\gamma_0, \gamma_1)$. By Toponogov's theorem, the first condition would imply that $\mathcal{M}$ has non-negative sectional curvature, and the latter would imply that $\mathcal{M}$ has non-positive sectional curvature; hence, $\mathcal{M}$ would need to be flat.

A more appropriate class of connecting curves are *segments*, which are defined with the help of the exponential map $\exp_{(\cdot)}$ as follows. Fix $v \in \mathcal{M}$, and let $\gamma_0, \gamma_1 \in \mathcal{M}$ lie outside of v's cut locus $\mathrm{cut}(v)$. Then, there are unique ξ_0, ξ_1 in the injectivity domain $I(v) \subset T_v\mathcal{M}$ of the exponential map $\exp_v : T_v\mathcal{M} \to \mathcal{M}$ such that $\exp_v(\xi_0) = \gamma_0$ and $\exp_v(\xi_1) = \gamma_1$. Further, assume that the straight line from ξ_0 to ξ_1 lies in $I(v)$. The segment $[\gamma_0, \gamma_1; v]_s : [0,1] \to \mathcal{M}$ with base v connecting γ_0 to γ_1 is then defined by $[\gamma_0, \gamma_1; v]_s = \exp_v((1-s)\xi_0 + s\xi_1)$.

Kim and McCann [57, Corollary 2.11] have established a sufficient criterion for the convexity of

$$[0,1] \ni s \mapsto \boldsymbol{d}^2\big(v, [\gamma_0, \gamma_1; v]_s\big) - \boldsymbol{d}^2\big(u, [\gamma_0, \gamma_1; v]_s\big), \tag{5.1.6}$$

independently of $u \in \mathcal{M}$ in terms of the *cross curvature*. Their hypotheses are as follows.

(KM0) The squared metric $\boldsymbol{d}^2(\cdot,\cdot)$, induced on $\mathcal{M}$ via $\mathbf{g}$, is C^4-regular outside of the cut locus.

(KM1) For each $v \in \mathcal{M}$, its injectivity domain $I(v)$ is convex, so segments $[\gamma_0, \gamma_1; v]$ can be defined for arbitrary $\gamma_0, \gamma_1 \notin \mathrm{cut}(v)$.

(KM2) For each segment $[\gamma_0, \gamma_1; v]$, there is a dense subset $U \subset \mathcal{M}$, such that there is no $u \in U$ and no $s \in [0,1]$ with $u \in \mathrm{cut}([\gamma_0, \gamma_1; v]_s)$.

(KM3) $(\mathcal{M}, \mathbf{g})$ has non-negative cross curvature.

Note, a Riemannian manifold $(\mathcal{M}, \mathbf{g})$ is said to have non-negative cross curvature if and only if for each $(x, y) \in \mathcal{N}$ and $v \in T_x\mathcal{M}$, $w \in T_y\mathcal{M}$,

$$-\frac{d^2}{dt^2}\Big|_{t=0} \frac{d^2}{ds^2}\Big|_{s=0} \boldsymbol{d}^2\big([x_0, x_1; y_0]_t, y_s\big) \geq 0. \tag{5.1.7}$$

for a given curve $y : (-\varepsilon, \varepsilon) \to \mathcal{M}$ and points $x_0, x_1 \notin \mathrm{cut}(y_0)$ where $w := \dot{y}_0 \in T_{y_0}\mathcal{M}$, $v \in T_{x_0}\mathcal{M}$, and the derivative of $t \mapsto [x_0, x_1; y_0]_t$ at $t = 0$.

Apart from (KM0), each of these conditions is rather demanding. A class of examples satisfying (KM0)–(KM3) are the round spheres $\mathbb{S}^d$. For these, (KM0)–(KM2) are easily verified since $\mathrm{cut}(v) = \{-v\}$ only contains the antipodal point, and $I(v)$ is the open d-dimensional ball of radius π, for each $v \in \mathbb{S}^d$. In contrast, the proof of (KM3) has been a challenge even for spheres, that has been mastered in [57, Theorem 6.2]. It seems that — apart from products and quotients of spheres — no further explicit examples satisfying (KM0)–(KM3) are currently known.

Theorem 5.1.5. *Assume that $(\mathbf{X}, \boldsymbol{d})$ is a compact Riemannian manifold $(\mathcal{M}, \mathbf{g})$ that satisfies (KM0)–(KM3) above. Assume further that $\mathcal{E} \in C^1(\mathcal{M})$ is semi-convex. Then (E3) is satisfied with some $\lambda \in \mathbb{R}$, with $\gamma_s := [\gamma_0, \gamma_1; v]_s$.*

Proof. For given $u, v \in \mathcal{M}$ and $\gamma_0, \gamma_1 \in \mathcal{M} \setminus \mathrm{cut}(v)$, let $\gamma_s := [\gamma_0, \gamma_1; v]_s$; the result for general $\gamma_0, \gamma_1 \in \mathcal{M}$ follows by continuity a forteriori. Further, we shall assume that $\mathcal{E} \in C^2(\mathcal{M})$ during the computations. Since $\mathcal{E}$ is semi-convex, and $\mathcal{M}$ is compact, there is a global modulus $\lambda' \leq 0$ of convexity, i.e., $\mathrm{Hess}\,\mathcal{E}(v) \geq \lambda'$ as a quadratic form on each $T_v\mathcal{M}$. The final estimate (5.1.8) depends only on λ', so (E3) follows for general semi-convex $\mathcal{E} \in C^1(\mathcal{M})$ by approximation.

We split
$$\Psi(\tau, u, v; \gamma_s) = h_1(s) + h_2(s) + h_3(s),$$
with $h_1, h_2, h_3 : [0,1] \to \mathbb{R}$ given by
$$h_1(s) = \frac{3}{4\tau}\boldsymbol{d}^2(v, \gamma_s), \quad h_2(s) = \frac{1}{4\tau}\big(\boldsymbol{d}^2(v, \gamma_s) - \boldsymbol{d}^2(u, \gamma_s)\big), \quad h_3(s) = \mathcal{E}(\gamma_s).$$
First, by definition of the segment γ_s via the exponential map, $s \mapsto \boldsymbol{d}^2(v, \gamma_s)$ is twice differentiable with
$$\frac{3}{4\tau}\frac{d^2}{ds^2}\boldsymbol{d}^2(v, \gamma_s) \equiv \frac{3}{2\tau}\|\xi_1 - \xi_0\|_v^2,$$
where $\|\xi\|_v^2 = \mathbf{g}_v(\xi, \xi)$. Second, by the hypotheses (KM0)–(KM3), the result from [57, Corollary 2.11] applies, so h_2 is convex. Finally, concerning h_3: in the normal coordinates induced by $\exp_v : I(v) \to \mathcal{M}$, the segment γ_s is the straight line connecting ξ_0 to ξ_1, hence (recalling the definition of the Hessian, and that $\exp_v$ is a 1-Lipschitz map):
$$\begin{aligned} h_3''(s) = \frac{d^2}{ds^2}\mathcal{E}(\gamma_s) &= \mathrm{Hess}\,\mathcal{E}(\gamma_s)[\dot\gamma_s] + d\mathcal{E}(\gamma_s)\big[\nabla_{\dot\gamma_s}\dot\gamma_s\big] \\ &\geq \lambda'\|\dot\gamma_s\|_{\gamma_s}^2 - \|\mathcal{E}\|_{C^1}\Big\|\sum_{i,j,k}\Gamma_{i,j}^k(\xi_1^i - \xi_0^i)(\xi_1^j - \xi_0^j)\frac{\partial}{\partial x_k}\Big\|_{\gamma_s} \\ &\geq \big(\lambda' - K\|\mathcal{E}\|_{C^1}\big)\|\xi_1 - \xi_0\|_v^2. \end{aligned}$$
Here K is a bound on the Christoffel symbols Γ_{ij}^k on the smooth and compact manifold $(\mathcal{M}, \mathbf{g})$, and for the estimate $\|\dot\gamma_s\| \leq \|\xi_1 - \xi_0\|_v$, we have used that (KM3) implies that $(\mathcal{M}, \mathbf{g})$ is of non-negative sectional curvature.

In summary, we have shown that $s \mapsto \Phi(\tau, u, v; \gamma_s)$ is uniformly convex of modulus

$$\Big(\frac{3}{4\tau} + \lambda\Big)\|\xi_1 - \xi_0\|_v^2 \quad \text{with} \quad \lambda := \lambda' - K\|\mathcal{E}\|_{C^1}. \tag{5.1.8}$$

Recalling that (KM3) implies non-negative sectional curvature on $(\mathcal{M}, \mathbf{g})$, we conclude that $\boldsymbol{d}^2(\gamma_0, \gamma_1) \leq \|\xi_1 - \xi_0\|_v^2$, so the claim (E3) follows. □

L^2-Wasserstein Space. In our last example, we consider the classical L^2-Wasserstein space $(\mathcal{P}_2(\Omega), \mathbf{W}_2)$ of the probability measures of finite second moment over a convex, possibly unbounded domain $\Omega \subset \mathbb{R}^{\mathrm{d}}$. And we assume that $\mathcal{E}$ is uniformly displacement semi-convex; the definition is recalled below. We remark that the class of gradient flows generated in this setting encompasses nonlinear drift-diffusion-aggregation equations of the form

$$\partial_t \rho_t = \Delta(\rho_t^m) + \nabla \cdot (\rho_t \nabla V) + \nabla \cdot (\rho_t * \nabla W \, \rho_t),$$

under the restrictions that $m \geq (d-1)/d$, and that $V, W \in C^2(\Omega)$ are uniformly semi-convex.

$(\mathcal{P}_2(\Omega), \mathbf{W}_2)$ is a complete geodesic space, which has non-negative curvature in the sense of Alexandrov. Similarly, as in the case of (non-negatively cross-curved) Riemannian manifolds discussed above, one cannot expect that hypothesis (E3) is satisfied for the geodesic $\widetilde{\gamma}_s$ connecting the two given measures $\gamma_0, \gamma_1 \in \mathcal{P}_2(\Omega)$. Indeed, $s \mapsto \mathbf{W}_2^2(\widetilde{\gamma}_s, u)$ is typically *not* uniformly convex of modulus $\mathbf{W}_2^2(\gamma_0, \gamma_1)$, see [4, Example 7.3.3]. Again, segments with a prescribed base point are more appropriate.

We need to recall some basic notations from the theory of optimal mass transport. $\mathcal{P}_2(\Omega_j \times \Omega_k)$ is the space of probability measures with finite second moment on the cross product $\Omega \times \Omega$, and the indices j and k indicate that we use coordinates $x_j \in \Omega$ and $x_k \in \Omega$ on the components, i.e., we write $x = (x_j, x_k) \in \Omega \times \Omega$. We introduce the canonical projections $\pi_j : (x_j, x_k) \mapsto x_j$, and for $s \in [0,1]$, we define $\pi_s : \Omega_0 \times \Omega_1 \to \Omega$ by $\pi_s := (1-s)\pi_0 + s\pi_1$ for brevity. We write $(\pi_j)_\# \boldsymbol{p}$ for the j-marginal of $\boldsymbol{p} \in \mathcal{P}_2(\Omega_j \times \Omega_k)$, and analogously, for $\boldsymbol{p} \in \mathcal{P}_2(\Omega_0 \times \Omega_1)$ and $s \in [0,1]$, the interpolating measure $(\pi_s)_\# \boldsymbol{p} \in \mathcal{P}_2(\Omega)$ is characterized by

$$\int_\Omega \varphi(y) \, \mathrm{d}(\pi_s)_\# \boldsymbol{p}(y) = \int_\Omega \varphi\big((1-s)x_0 + s x_1\big) \, \mathrm{d}\boldsymbol{p}(x), \quad \text{for all } \varphi \in C_b^0(\Omega).$$

A transport plan from $\mu_0 \in \mathcal{P}_2(\Omega_0)$ to $\mu_1 \in \mathcal{P}_2(\Omega_1)$ is any $\boldsymbol{p} \in \mathcal{P}_2(\Omega_0 \times \Omega_1)$ satisfying the marginal constraints $\pi_0 \# \boldsymbol{p} = \mu_0$ and $(\pi_1)_\# \boldsymbol{p} = \mu_1$. Such a plan $\boldsymbol{p}$ is called optimal if it is a minimizer in the Kantorovich problem

$$\boldsymbol{p} \mapsto \int_{\Omega^2} |x_0 - x_1|^2 \, \mathrm{d}\boldsymbol{p}(x). \tag{5.1.9}$$

For any given $\mu_0, \mu_1 \in \mathcal{P}_2(\Omega)$, there exists at least one optimal plan and if one of the measures μ_i is absolutely continuous, the optimal plan $\boldsymbol{p}$ is unique.

We are going to use the following two facts, which are essentially [4, Lemma 5.3.2] and [4, Proposition 7.3.1]:

1. *Glueing lemma:* Given $\alpha \in \mathcal{P}_2(\Omega_0 \times \Omega_2)$ and $\beta \in \mathcal{P}_2(\Omega_1 \times \Omega_2)$ with $(\pi_2)_\#\alpha = (\pi_2)_\#\beta$, there exists a $\mu \in \mathcal{P}_2(\Omega_0 \times \Omega_1 \times \Omega_2)$ such that $(\pi_0, \pi_2)_\#\mu = \alpha$ and $(\pi_1, \pi_2)_\#\mu = \beta$.
2. *Curve lemma:* Given $\alpha \in \mathcal{P}_2(\Omega_0 \times \Omega_1)$, $\beta \in \mathcal{P}_2(\Omega_3)$ and $t \in [0,1]$, there exists a $\mu \in \mathcal{P}_2(\Omega_0 \times \Omega_1 \times \Omega_3)$ such that $(\pi_0, \pi_1)_\#\mu = \alpha$, and $(\pi_t, \pi_3)_\#\mu$ is an optimal transport plan from $(\pi_t)_\#\alpha$ to β.

Segments — which are referred to as generalized geodesics in [4] — are defined as follows. Let $\boldsymbol{p}_{02} \in \mathcal{P}_2(\Omega_0 \times \Omega_2)$ and $\boldsymbol{p}_{12} \in \mathcal{P}_2(\Omega_1 \times \Omega_2)$ be optimal plans for the transport of γ_0 and γ_1, respectively, to $v \in \mathcal{P}_2(\Omega_2)$. By the glueing lemma, there exists a $\boldsymbol{p}_{012}$ such that $(\pi_0, \pi_2)_\#\boldsymbol{p}_{012} = \boldsymbol{p}_{01}$ and $(\pi_1, \pi_2)_\#\boldsymbol{p}_{012} = \boldsymbol{p}_{12}$. Then $[\gamma_0, \gamma_1; v]_s := (\pi_s)_\#\boldsymbol{p}_{012}$. Finally, we recall that $\mathcal{E}$ being uniformly displacement semi-convex of modulus λ means that

$$\mathcal{E}\big([\gamma_0, \gamma_1; v]_s\big) \leq (1-s)\mathcal{E}(\gamma_0) + s\mathcal{E}(\gamma_1) - \frac{\lambda}{2}\int_\Omega |x_0 - x_1|^2 \, \mathrm{d}\boldsymbol{p}_{012}(x).$$

In the language of [4], this property is referred to as λ-convexity along generalized geodesics.

Theorem 5.1.6. *Let $\Omega \subseteq \mathbb{R}^\mathrm{d}$ and assume that the metric space $(\mathbf{X}, \boldsymbol{d})$ is the L^2-Wasserstein space $(\mathcal{P}_2(\Omega), \mathbf{W}_2)$. Assume further that $\mathcal{E}$ is uniformly displacement semi-convex of modulus λ. Then (E3) is satisfied, with the same λ, for $\gamma_s = [\gamma_0, \gamma_1; v]_s$.*

Proof. Let $u, v, \gamma_0, \gamma_1 \in \mathcal{P}_2(\Omega)$ be given, and let $\boldsymbol{p}_{012}$ be as above. We are going to prove the inequality (5.1.2) directly for a fixed value $s \in (0,1)$. Since $(\pi_s, \pi_2)_\#\boldsymbol{p}_{012}$ is *some* transport from γ_s to v, and $(\pi_0, \pi_2)_\#\boldsymbol{p}_{012}$, $(\pi_1, \pi_2)_\#\boldsymbol{p}_{012}$ are both optimal,

$$\begin{aligned}
\mathbf{W}_2^2(\gamma_s, v) &\leq \int_\Omega |(1-s)x_0 + sx_1 - x_2|^2 \, \mathrm{d}\boldsymbol{p}_{012}(x) \\
&= \int_\Omega \big[(1-s)|x_0 - x_2|^2 + s|x_1 - x_2|^2 - s(1-s)|x_0 - x_1|^2\big] \, \mathrm{d}\boldsymbol{p}_{012}(x) \\
&= (1-s)\mathbf{W}_2^2(\gamma_0, v) + s\mathbf{W}_2^2(\gamma_1, v) - s(1-s)\int_\Omega |x_0 - x_1|^2 \, \mathrm{d}\boldsymbol{p}_{012}(x).
\end{aligned}$$

By the curve extension lemma, there exists a $\boldsymbol{p}_{013} \in \mathcal{P}_2(\Omega_0 \times \Omega_1 \times \Omega_3)$, such that $(\pi_0, \pi_1)_\#\boldsymbol{p}_{013} = (\pi_0, \pi_1)_\#\boldsymbol{p}_{012}$, and $(\pi_s, \pi_3)_\#\boldsymbol{p}_{013}$ is an optimal plan from γ_s to u. It follows that $(\pi_0, \pi_3)_\#\boldsymbol{p}_{013}$ and $(\pi_1, \pi_3)_\#\boldsymbol{p}_{013}$ are some transport plans from γ_0 and γ_1, respectively, to u, and so

$$\begin{aligned}
\mathbf{W}_2^2(\gamma_s, u) &= \int_\Omega |(1-s)x_0 + sx_1 - x_3|^2 \, \mathrm{d}\boldsymbol{p}_{013}(x) \\
&= \int_\Omega \big[(1-s)|x_0 - x_3|^2 + s|x_1 - x_3|^2 - s(1-s)|x_0 - x_1|^2\big] \, \mathrm{d}\boldsymbol{p}_{013}(x).
\end{aligned}$$

Since $(\pi_0, \pi_3)_{\#}\boldsymbol{p}_{013}$ and $(\pi_1, \pi_3)_{\#}\boldsymbol{p}_{013}$ are not optimal we can conclude

$$\mathbf{W}_2^2(\gamma_s, u) \geq (1-s)\mathbf{W}_2^2(\gamma_0, u) + s\mathbf{W}_2^2(\gamma_1, u) - s(1-s)\int_\Omega |x_0 - x_1|^2 \, \mathrm{d}\boldsymbol{p}_{012}(x).$$

In combination with the definition of λ-uniform displacement convexity of $\mathcal{E}$, we arrive at

$$\begin{aligned}\Psi(\tau, u, v; \gamma_s) \leq &(1-s)\Psi(\tau, u, v; \gamma_0) + s\Psi(\tau, u, v; \gamma_1) \\ &- \frac{1}{2}\Big(\frac{3}{2\tau} + \lambda\Big)s(1-s)\int_\Omega |x_1 - x_0|^2 \, \mathrm{d}\boldsymbol{p}_{012}(x).\end{aligned}$$

Clearly, the integral above is larger or equal to $\mathbf{W}_2^2(\gamma_0, \gamma_1)$, hence (5.1.2) for any $\tau > 0$ so small that $\frac{3}{2\tau} + \lambda > 0$. □

5.1.2 Basic Properties of the BDF2 Penalization Ψ

In this section, we study the basic properties of the BDF2 scheme. First, we prove well-posedness in the sense that for all sufficiently small $\tau > 0$, and arbitrary data $u, v \in \mathcal{D}(\mathcal{E})$, the functional $\Psi(\tau, u, v; \cdot)$ possesses a unique minimizer in $\mathcal{D}(\mathcal{E})$. Consequently, for an arbitrary pair $(u_0^{\boldsymbol{\tau}}, u_1^{\boldsymbol{\tau}})$ of initial conditions, one obtains inductively a unique global discrete solution $(u_k^{\boldsymbol{\tau}})_{k\in\mathbb{N}}$ by solving the corresponding sequence of minimization problems in (5.3.2). Subsequently, we derive some fundamental estimates that are needed for the convergence proof later.

Recall that Assumptions (E1)–(E3) are supposed to hold, with (5.1.3).

Theorem 5.1.7 (Existence of a minimizer)**.** *For all $\tau \in (0, \tau_*)$ and for all $u, v \in \mathbf{X}$, there exists a unique minimizer $w_* \in \mathcal{D}(\mathcal{E})$ of $w \mapsto \Psi(\tau, u, v; w)$.*

Proof. Let $\tau \in (0, \tau_*)$ and $u, v \in \mathbf{X}$ be fixed. For brevity, we write $\psi(w) := \Psi(\tau, u, v; w)$.

First, we show that ψ is bounded from below. By the triangle inequality and the binomial formula, we have that

$$\boldsymbol{d}^2(u, w) \leq 2\boldsymbol{d}^2(u, v) + 2\boldsymbol{d}^2(v, w), \quad \boldsymbol{d}^2(u_*, w) \leq \frac{2\tau_*}{\tau_* + \tau}\boldsymbol{d}^2(w, v) + \frac{2\tau_*}{\tau_* - \tau}\boldsymbol{d}^2(v, u_*).$$

Substituting these estimates into the definition of $\psi(w) = \Psi(\tau, u, v; w)$ and using Assumption (E2), we obtain for each $w \in \mathcal{D}(\mathcal{E})$:

$$\begin{aligned}\psi(w) \geq &\ \frac{1}{\tau_* + \tau}\boldsymbol{d}^2(v, w) + \frac{1}{2\tau}\boldsymbol{d}^2(v, w) - \frac{1}{4\tau}\boldsymbol{d}^2(u, w) + \mathcal{E}(w) \\ \geq &- \frac{1}{\tau_* - \tau}\boldsymbol{d}^2(v, u_*) + \frac{1}{2\tau_*}\boldsymbol{d}^2(u_*, w) - \frac{1}{2\tau}\boldsymbol{d}^2(v, u) + \mathcal{E}(w) \\ \geq &- \frac{1}{\tau_* - \tau}\boldsymbol{d}^2(v, u_*) - \frac{1}{2\tau}\boldsymbol{d}^2(v, u) + c_*.\end{aligned}$$

The last expression, which only depends on the given quantities u and v, constitutes the sought for lower bound on ψ. Consequently,

$$\underline{\psi} := \inf_{w \in \mathcal{D}(\mathcal{E})} \psi(w) > -\infty.$$

Now, choose a minimizing sequence $(w_n)_{n\in\mathbb{N}}$ in $\mathcal{D}(\mathcal{E})$, i.e.,

$$\psi(w_n) \searrow \underline{\psi}. \tag{5.1.10}$$

We are going to prove that this is a Cauchy sequence. Towards that goal, we invoke Assumption (E3): specifically, for given indices m and n, we choose $\gamma_0 = w_m$, $\gamma_1 = w_n$, and we define $w_{m,n} := \gamma_{\frac{1}{2}}$, the midpoint of the respective curve joining w_m to w_n. Then, by (5.1.2),

$$\psi(w_{m,n}) \le \frac{1}{2}\psi(w_m) + \frac{1}{2}\psi(w_n) - \frac{1}{8}\Big(\frac{3}{2\tau} + \lambda\Big)\boldsymbol{d}^2(w_m, w_n).$$

Since $\tau < \tau_*$ by hypothesis, and $3 + 2\lambda\tau_* \ge 2$ thanks to (5.1.3), this yields an estimate on the distance from w_m to w_n:

$$\boldsymbol{d}^2(w_m, w_n) \le \frac{8\tau}{3+2\tau\lambda}\left(\psi(w_m) + \psi(w_n) - 2\psi(w_{m,n})\right) \le \frac{8\tau}{3+2\tau\lambda}\left(\psi(w_m) + \psi(w_n) - 2\underline{\psi}\right).$$

In view of (5.1.10), this verifies the Cauchy property of $(w_n)_{n\in\mathbb{N}}$. Consequently, and by completeness of $(\mathbf{X}, \boldsymbol{d})$, that sequence converges to a limit $w_* \in \mathbf{X}$.

According to Assumption (E1), $\mathcal{E}$ is lower $\boldsymbol{d}$-semi-continuous. Since the distance to a given point is clearly a continuous function, also ψ is lower $\boldsymbol{d}$-semi-continuous. By the usual argument

$$\underline{\psi} \le \psi(w_*) \le \liminf_{n\to\infty} \psi(w_n) = \underline{\psi},$$

we conclude that ψ attains its infimum $\underline{\psi}$ at w_*, i.e., w_* is a minimizer.

The uniqueness of the minimizer follows by Assumption (E3) as well: by the remarks following (5.1.3), ψ is *strictly* convex along some curve that connects two potentially different minimizers. But that would mean that ψ attains a value lower than that at the minimizers, a contradiction. □

5.1.3 Properties of the BDF2 Scheme

In the following, we assume that discrete initial data (u_0^τ, u_1^τ) are given for each $\tau \in (0, \tau_*)$, and we consider the — according to Theorem 5.1.7 above — well-defined family of discrete solutions $(u_k^\tau)_{k\in\mathbb{N}}$. We recall that one of the key features of the implicit Euler method is that the energy values $\mathcal{E}(u_{k+1}^\tau)$ are monotonically decreasing with k. This is not quite the case for the BDF2 scheme at hand, but we can prove a slightly weaker property.

Lemma 5.1.8 (Almost Energy Diminishing). *Each discrete solution $(u_k^\tau)_{k\in\mathbb{N}}$ satisfies*

$$\mathcal{E}(u_{k+1}^\tau) + \frac{1}{2\tau}\boldsymbol{d}^2(u_k^\tau, u_{k+1}^\tau) \leq \mathcal{E}(u_k^\tau) + \frac{1}{4\tau}\boldsymbol{d}^2(u_{k-1}^\tau, u_k^\tau) \tag{5.1.11}$$

at each step $k = 1, 2, \ldots$.

Proof. Since u_{k+1}^τ is a minimizer of $w \mapsto \Psi(\tau, u_{k-1}^\tau, u_k^\tau; w)$, it satisfies

$$\Psi(\tau, u_{k-1}^\tau, u_k^\tau; u_{k+1}^\tau) \leq \Psi(\tau, u_{k-1}^\tau, u_k^\tau; w)$$

for all $w \in \mathbf{X}$. For the choice $w = u_k^\tau$, we obtain

$$\frac{1}{\tau}\boldsymbol{d}^2(u_k^\tau, u_{k+1}^\tau) - \frac{1}{4\tau}\boldsymbol{d}^2(u_{k-1}^\tau, u_{k+1}^\tau) + \mathcal{E}(u_{k+1}^\tau) \leq -\frac{1}{4\tau}\boldsymbol{d}^2(u_{k-1}^\tau, u_k^\tau) + \mathcal{E}(u_k^\tau). \tag{5.1.12}$$

By the triangle inequality and the binomial formula,

$$\boldsymbol{d}^2(u_{k-1}^\tau, u_{k+1}^\tau) \leq 2\boldsymbol{d}^2(u_{k-1}^\tau, u_k^\tau) + 2\boldsymbol{d}^2(u_k^\tau, u_{k+1}^\tau). \tag{5.1.13}$$

Substitute this in the left-hand side of (5.2.9). This yields (5.1.11) □

Next, we derive the classical estimates on energy and distance. These require (I1) to hold for the approximation (u_0^τ, u_1^τ) of the initial datum u_0.

Theorem 5.1.9 (Classical Estimates). *Fix a time horizon $T > 0$. Then there is a constant C, depending only on d_1, d_2 and T, such that the corresponding discrete solutions $(u_k^\tau)_{k\in\mathbb{N}}$ satisfy*

$$\sum_{k=1}^{N} \frac{1}{2\tau}\boldsymbol{d}^2(u_{k-1}^\tau, u_k^\tau) \leq C, \qquad |\mathcal{E}(u_N^\tau)| \leq C, \qquad \boldsymbol{d}^2(u_*, u_N^\tau) \leq C \tag{5.1.14}$$

for all $N \in \mathbb{N}$ with $N\tau \leq T$.

Proof. The main estimate is easy to obtain: sum up the inequalities (5.1.11) for $k = 1$ to $k = N - 1$. After the cancellation of corresponding terms on both sides, we remain with

$$\mathcal{E}(u_N^\tau) + \frac{1}{4\tau}\sum_{k=1}^{N} \boldsymbol{d}^2(u_{k-1}^\tau, u_k^\tau) \leq \mathcal{E}(u_1^\tau) + \frac{1}{4\tau}\boldsymbol{d}^2(u_0^\tau, u_1^\tau) \leq d_1 + \frac{1}{4}d_2^2\tau_*, \tag{5.1.15}$$

where (I1) has been used in the last inequality. Clearly, if $\mathcal{E}$ would be bounded below, then (5.1.14) would follow immediately.

Since we only assume the weaker lower bound (E2), more work is required. First, we show that

$$\boldsymbol{d}^2(u_*, u_k^\tau) - \boldsymbol{d}^2(u_*, u_{k-1}^\tau) \leq 2\boldsymbol{d}(u_{k-1}^\tau, u_k^\tau)\boldsymbol{d}(u_*, u_k^\tau). \tag{5.1.16}$$

We only need to consider the case that $\boldsymbol{d}(u_*, u_k^\tau) \geq \boldsymbol{d}(u_*, u_{k\text{-}1}^\tau)$, since otherwise the inequality is trivially true. But then an application of the triangle inequality yields:

$$\begin{aligned}
&\boldsymbol{d}^2(u_*, u_k^\tau) - \boldsymbol{d}^2(u_*, u_{k\text{-}1}^\tau) \\
=&\big(\boldsymbol{d}(u_*, u_k^\tau) + \boldsymbol{d}(u_*, u_{k\text{-}1}^\tau)\big)\big(\boldsymbol{d}(u_*, u_k^\tau) - \boldsymbol{d}(u_*, u_{k\text{-}1}^\tau)\big) \\
\leq&\big(\boldsymbol{d}(u_*, u_k^\tau) + \boldsymbol{d}(u_*, u_k^\tau)\big)\big(\boldsymbol{d}(u_*, u_{k\text{-}1}^\tau) + \boldsymbol{d}(u_{k\text{-}1}^\tau, u_k^\tau) - \boldsymbol{d}(u_*, u_{k\text{-}1}^\tau)\big) \\
=&2\boldsymbol{d}(u_{k\text{-}1}^\tau, u_k^\tau)\boldsymbol{d}(u_*, u_k^\tau),
\end{aligned}$$

which is (5.1.16). We use inequality (5.1.16) and the binomial formula to estimate

$$\begin{aligned}
\frac{1}{2}\boldsymbol{d}^2(u_*, u_N^\tau) - \frac{1}{2}\boldsymbol{d}^2(u_*, u_1^\tau) &= \frac{1}{2}\sum_{k=2}^{N}\left[\boldsymbol{d}^2(u_*, u_k^\tau) - \boldsymbol{d}^2(u_*, u_{k\text{-}1}^\tau)\right] \\
&\leq \sum_{k=2}^{N} \boldsymbol{d}(u_{k\text{-}1}^\tau, u_k^\tau)\boldsymbol{d}(u_*, u_k^\tau) \\
&\leq \sum_{k=2}^{N} \frac{\tau_*}{8\tau}\boldsymbol{d}^2(u_{k\text{-}1}^\tau, u_k^\tau) + \sum_{k=2}^{N} \frac{2\tau}{\tau_*}\boldsymbol{d}^2(u_*, u_k^\tau).
\end{aligned}$$

At this point, we substitute estimate (5.1.16) and use Assumption (E2) to obtain

$$\begin{aligned}
&\frac{1}{2}\boldsymbol{d}^2(u_*, u_N^\tau) - \frac{1}{2}\boldsymbol{d}^2(u_*, u_1^\tau) \\
\leq&\frac{\tau_*}{2}\Big(\mathcal{E}(u_1^\tau) - \mathcal{E}(u_N^\tau) + \frac{1}{4\tau}\boldsymbol{d}^2(u_0^\tau, u_1^\tau)\Big) + \frac{2\tau}{\tau_*}\sum_{k=2}^{N}\boldsymbol{d}^2(u_*, u_k^\tau) \\
\leq&\frac{\tau_*}{2}\Big(\mathcal{E}(u_1^\tau) - c_* + \frac{1}{2\tau_*}\boldsymbol{d}^2(u_*, u_N^\tau) + \frac{1}{4\tau}\boldsymbol{d}^2(u_0^\tau, u_1^\tau)\Big) + \frac{2\tau}{\tau_*}\sum_{k=2}^{N}\boldsymbol{d}^2(u_*, u_k^\tau).
\end{aligned}$$

We rearrange terms and use (I1) to arrive at the following discrete Gronwall inequality:

$$\boldsymbol{d}^2(u_*, u_N^\tau) \leq 2K_0^2 + 2\tau_*\,(d_1 - c_*) + \frac{\tau_*}{2}d_2^2 + \frac{8\tau}{\tau_*}\sum_{k=2}^{N}\boldsymbol{d}^2(u_*, u_k^\tau).$$

One verifies by induction on N that

$$\boldsymbol{d}^2(u_*, u_N^\tau) \leq \Big[2K_0^2 + 2\tau_*\,(d_1 - c_*) + \frac{\tau_*}{2}d_2^2\Big]\Big(1 + \frac{8\tau}{\tau_*}\Big)^N \leq \widehat{C}\exp\Big(\frac{8N\tau}{\tau_*}\Big) \leq \widehat{C}\exp\Big(\frac{8T}{\tau_*}\Big),$$

proving the third estimate of (5.1.14).

From here, we conclude the second bound from (5.1.14): the bound on $\mathcal{E}(u_N^\tau)$ from above follows immediately from (5.1.15), for the bound from below, we combine the third estimate of (5.1.14) with Assumption (E2). Having the second estimate at hand, the bound first one from (5.1.14) follows again from (5.1.15). □

As a final preparation for the convergence proof, we derive a time-discrete version of the differential EVI (2.3.2). That estimate does not require any further assumptions on the discrete initial data.

Lemma 5.1.10 (Discrete EVI). *The discrete solution* $(u_k^{\boldsymbol{\tau}})_{k\in\mathbb{N}}$ *satisfies*

$$\begin{aligned}&\Big(\frac{3}{4\tau}+\frac{\lambda}{2}\Big)\boldsymbol{d}^2(u_{k+1}^{\boldsymbol{\tau}},w)-\frac{1}{\tau}\boldsymbol{d}^2(u_k^{\boldsymbol{\tau}},w)+\frac{1}{4\tau}\boldsymbol{d}^2(u_{k-1}^{\boldsymbol{\tau}},w)\\&\leq\mathcal{E}(w)-\mathcal{E}(u_{k+1}^{\boldsymbol{\tau}})-\frac{1}{\tau}\boldsymbol{d}^2(u_k^{\boldsymbol{\tau}},u_{k+1}^{\boldsymbol{\tau}})+\frac{1}{4\tau}\boldsymbol{d}^2(u_{k-1}^{\boldsymbol{\tau}},u_{k+1}^{\boldsymbol{\tau}})\end{aligned}\tag{5.1.17}$$

for all $k\in\mathbb{N}$*, and for all* $w\in\mathcal{D}(\mathcal{E})$*.*

Proof. This follows from Assumption (E3). Choose $\gamma_0=u_{k+1}^{\boldsymbol{\tau}}$ and $\gamma_1=w$, and let γ_s be the corresponding connecting curve such that (5.1.2) holds. Combine (5.1.2) with the fact that $u_{k+1}^{\boldsymbol{\tau}}$ minimizes $\Psi(\tau,u_{k-1}^{\boldsymbol{\tau}},u_k^{\boldsymbol{\tau}},\cdot)$ to obtain, for each $s\in(0,1)$,

$$\begin{aligned}0&\leq\Psi(\tau,u_{k-1}^{\boldsymbol{\tau}},u_k^{\boldsymbol{\tau}};\gamma_s)-\Psi(\tau,u_{k-1}^{\boldsymbol{\tau}},u_k^{\boldsymbol{\tau}};u_{k+1}^{\boldsymbol{\tau}})\\&\leq s\Psi(\tau,u_{k-1}^{\boldsymbol{\tau}},u_k^{\boldsymbol{\tau}};w)-s\Psi(\tau,u_{k-1}^{\boldsymbol{\tau}},u_k^{\boldsymbol{\tau}};u_{k+1}^{\boldsymbol{\tau}})-\frac{1}{2}\Big(\frac{3}{2\tau}+\lambda\Big)s(1-s)\boldsymbol{d}^2(u_{k+1}^{\boldsymbol{\tau}},w).\end{aligned}$$

Divide by $s\in(0,1)$ and pass to the limit $s\searrow 0$. This yields

$$0\leq\Psi(\tau,u_{k-1}^{\boldsymbol{\tau}},u_k^{\boldsymbol{\tau}};w)-\Psi(\tau,u_{k-1}^{\boldsymbol{\tau}},u_k^{\boldsymbol{\tau}};u_{k+1}^{\boldsymbol{\tau}})-\frac{1}{2}\Big(\frac{3}{2\tau}+\lambda\Big)\boldsymbol{d}^2(u_{k+1}^{\boldsymbol{\tau}},w),$$

which, by definition of Ψ, is the desired inequality (5.1.17). □

5.1.4 Convergence

Once again, we recall that $(\mathbf{X},\boldsymbol{d})$ is a separable and complete metric space, on which a functional $\mathcal{E}:\mathbf{X}\to\mathbb{R}\cup\{\infty\}$ is given, that satisfies Assumptions (E1)–(E3), with (5.1.3). Our main result is the following.

Theorem 5.1.11 (Convergence result). *Consider a sequence of equidistant partition* $\boldsymbol{\tau}_n=(\tau_n,2\tau_n,3\tau_n,\ldots)$ *with vanishing step sizes* $\tau_n\in(0,\tau_*)$ *which are strictly decreasing, and which are such that the quotients* τ_n/τ_{n+1} *are natural numbers. Let further initial data* $(u_0^{\boldsymbol{\tau}_n},u_1^{\boldsymbol{\tau}_n})$ *be given that satisfy (I1), and such that* $u_0^{\boldsymbol{\tau}_n}\to u_0$*.*

Then the associated discrete solution $(u_k^{\boldsymbol{\tau}_n})_{k\in\mathbb{N}}$ *is well-defined for each* n*, and the sequence of piecewise constant interpolations* $(\overline{u}_t^{\boldsymbol{\tau}_n})_{n\in\mathbb{N}}$ *converges locally uniformly with respect to time to a solution* $u_t^*\in\mathrm{AC}^2(0,\infty;(\mathbf{X},\boldsymbol{d}))$ *of the gradient flow for* $\mathcal{E}$*, i.e., the limit* u_* *satisfies* (2.3.2)*. More precisely, for every time horizon* $T>0$*, there is a constant* C *that can be expressed in terms of* d_1*,* d_2 *and* T *alone, such that*

$$\boldsymbol{d}\big(\overline{u}_t^{\boldsymbol{\tau}_n},u_t^*\big)\leq C\left(\boldsymbol{d}(u_0^{\boldsymbol{\tau}_n},u_0)+\sqrt{\tau_n}\right)\tag{5.1.18}$$

for all $t\in[0,T]$*.*

Remark 5.1.12. The hypothesis that consecutive τ_n's have an integer quotient has been made in order not to make the already quite technical proof even more technical. Under that hypothesis, the time grid associated with some τ_n is always a refinement of the grid for τ_m if $n > m$. That simplifies our calculations considerably.

Comparison Principle

The main ingredient of the proof of Theorem 5.1.11 is the following comparison principle, which estimates the rate at which two discrete solutions with almost identical initial data may diverge from each other.

Theorem 5.1.13 (Comparison principle)**.** *In the setting of Theorem 5.1.11, fix two equidistant partitions $\boldsymbol{\tau}$ and $\boldsymbol{\eta}$ with time step sizes $\tau := \tau_n$ and $\eta := \tau_m$ with $m > n$. Then, there is a constant C, expressible in terms of d_1, d_2 and T alone, such that*

$$\boldsymbol{d}^2\left(\overline{u}_t^{\boldsymbol{\tau}}, \overline{v}_t^{\boldsymbol{\eta}}\right) \leq C\left(\boldsymbol{d}^2(u_0^{\boldsymbol{\tau}}, v_0^{\boldsymbol{\eta}}) + \tau\right) \tag{5.1.19}$$

for all $t \in [0, T]$.

Proof. By hypothesis, define $R := \tau/\eta \in \mathbb{N}$. The basic idea is to derive bounds on the distance between the discrete solutions $(u_k^{\boldsymbol{\tau}})_{k\in\mathbb{N}}$ and $(v_\ell^{\boldsymbol{\eta}})_{\ell\in\mathbb{N}}$ at comparable times, i.e., for $(k-1)R \leq \ell \leq Rk$, by using the time-discrete EVI (5.1.17) for each of the two solutions and substituting the respective other solution for the “observer point” w. More specifically, multiplication of (5.1.17) for $u_{k+1}^{\boldsymbol{\tau}}$ by $(4\tau)/(3+2\lambda\tau)$ yields

$$\begin{aligned}
&\boldsymbol{d}^2(u_{k+1}^{\boldsymbol{\tau}}, w) - \frac{4}{3+2\lambda\tau}\boldsymbol{d}^2(u_k^{\boldsymbol{\tau}}, w) + \frac{1}{3+2\lambda\tau}\boldsymbol{d}^2(u_{k-1}^{\boldsymbol{\tau}}, w) \\
&\leq \frac{4\tau}{3+2\lambda\tau}\Big(\mathcal{E}(w) - \mathcal{E}(u_{k+1}^{\boldsymbol{\tau}}) - \frac{1}{\tau}\boldsymbol{d}^2(u_k^{\boldsymbol{\tau}}, u_{k+1}^{\boldsymbol{\tau}}) + \frac{1}{4\tau}\boldsymbol{d}^2(u_{k-1}^{\boldsymbol{\tau}}, u_{k+1}^{\boldsymbol{\tau}})\Big).
\end{aligned} \tag{5.1.20}$$

For brevity, we introduce

$$g_\tau := \frac{1}{2+\sqrt{1-2\lambda\tau}} = \frac{1}{3} + \mathcal{O}(\tau), \quad h_\tau := 2 - \sqrt{1-2\lambda\tau} = 1 + \mathcal{O}(\tau),$$

where the Landau symbol $\mathcal{O}(\tau)$ is understood for the limit $\tau \to 0$. Note, with this definition one has $\lambda_\tau := \frac{\log(h_\tau)}{\tau} = \lambda + \mathcal{O}(\tau)$. Furthermore, define

$$\begin{aligned}
a_k^{\boldsymbol{\tau}}(\overline{u}_t^{\boldsymbol{\tau}}; w) &:= h_\tau \boldsymbol{d}^2(u_{k+1}^{\boldsymbol{\tau}}, w) - \boldsymbol{d}^2(u_k^{\boldsymbol{\tau}}, w), \\
b_k^{\boldsymbol{\tau}}(\overline{u}_t^{\boldsymbol{\tau}}; w) &:= 4g_\tau\Big(\mathcal{E}(w) - \mathcal{E}(u_{k+1}^{\boldsymbol{\tau}}) - \frac{1}{\tau}\boldsymbol{d}^2(u_k^{\boldsymbol{\tau}}, u_{k+1}^{\boldsymbol{\tau}}) + \frac{1}{4\tau}\boldsymbol{d}^2(u_{k-1}^{\boldsymbol{\tau}}, u_{k+1}^{\boldsymbol{\tau}})\Big).
\end{aligned}$$

With these notations, the variational inequality (5.1.20) attains the following form:

$$a_k^{\boldsymbol{\tau}}(\overline{u}_t^{\boldsymbol{\tau}}; w) \leq g_\tau\, a_{k-1}^{\boldsymbol{\tau}}(\overline{u}_t^{\boldsymbol{\tau}}; w) + \tau b_k^{\boldsymbol{\tau}}(\overline{u}_t^{\boldsymbol{\tau}}; w).$$

An iteration of this inequality yields

$$a_k^{\boldsymbol{\tau}}(\overline{u}_t^{\boldsymbol{\tau}}; w) \leq g_\tau^k a_0^{\boldsymbol{\tau}}(\overline{u}_t^{\boldsymbol{\tau}}; w) + \tau \sum_{n=1}^{k} g_\tau^{k-n}\, b_n^{\boldsymbol{\tau}}(\overline{u}_t^{\boldsymbol{\tau}}; w). \tag{5.1.21}$$

Analogously, define $g_\eta, h_\eta, \lambda_\eta$, as well as $a_\ell^{\boldsymbol{\eta}}(\overline{v}_t^{\boldsymbol{\eta}}; w), b_\ell^{\boldsymbol{\eta}}(\overline{v}_t^{\boldsymbol{\eta}}; w)$, replacing $u_{k-1}^{\boldsymbol{\tau}}, u_k^{\boldsymbol{\tau}}, u_{k+1}^{\boldsymbol{\tau}}$ and τ with $v_{\ell-1}^{\boldsymbol{\eta}}, v_\ell^{\boldsymbol{\eta}}, v_{\ell+1}^{\boldsymbol{\eta}}$ and η, respectively. By the same argument as above, one obtains a corresponding estimate for $a_\ell^{\boldsymbol{\eta}}(\overline{v}_t^{\boldsymbol{\eta}}; w)$.

Now fix a time $t \in [0, T]$, and define the three quantities $N := \max\{n \mid n\tau \leq t\}$, $M := \max\{m \mid m\eta \leq t\}$, and $L := M - RN$. Further more, introduce

$$q_{k,\ell} := h_\tau^k h_\eta^\ell \boldsymbol{d}^2(u_k^{\boldsymbol{\tau}}, v_\ell^{\boldsymbol{\eta}}) = e^{\lambda_\tau k\tau + \lambda_\eta \ell\eta} \boldsymbol{d}^2\big(\overline{u}_{k\tau}^{\boldsymbol{\tau}}, \overline{v}_{\ell\eta}^{\boldsymbol{\eta}}\big).$$

The goal is to derive an estimate on the difference

$$q_{N,M} - q_{0,0} = h_\tau^N h_\eta^M \boldsymbol{d}^2(u_N^{\boldsymbol{\tau}}, v_M^{\boldsymbol{\eta}}) - \boldsymbol{d}^2(u_0^{\boldsymbol{\tau}}, v_0^{\boldsymbol{\eta}}).$$

We expand this difference into telescopic sums firstly with respect to k:

$$q_{N,M} - q_{0,0} = \big(q_{N,M} - q_{N,RN}\big) + \sum_{k=0}^{N-1} \big(q_{k+1,R(k+1)} - q_{k,Rk}\big)$$

and secondly in ℓ such that we get in the end

$$q_{N,M} - q_{0,0} = \sum_{\ell=RN}^{M-1} (q_{N,\ell+1} - q_{N,\ell}) + \sum_{k=0}^{N-1} \Big((q_{k+1,Rk} - q_{k,Rk}) + \sum_{\ell=Rk}^{R(k+1)-1} (q_{k+1,\ell+1} - q_{k+1,\ell}) \Big).$$

By definition of a and b, the differences inside the sums satisfy

$$q_{k+1,\ell} - q_{k,\ell} = h_\tau^k h_\eta^\ell a_k^{\boldsymbol{\tau}}(\overline{u}_t^{\boldsymbol{\tau}}; v_\ell^{\boldsymbol{\eta}}), \qquad q_{k,\ell+1} - q_{k,\ell} = h_\tau^k h_\eta^\ell a_\ell^{\boldsymbol{\eta}}(\overline{v}_t^{\boldsymbol{\eta}}; u_k^{\boldsymbol{\tau}}).$$

Insert this above and use the estimates (5.1.21) to obtain

$$h_\tau^N h_\eta^M \boldsymbol{d}^2(u_N^{\boldsymbol{\tau}}, v_M^{\boldsymbol{\eta}}) - \boldsymbol{d}^2(u_0, v_0) \leq\; I_{N,M}^{\boldsymbol{\tau},\boldsymbol{\eta}}(\overline{u}_t^{\boldsymbol{\tau}}, \overline{v}_t^{\boldsymbol{\eta}})$$

$$:= \sum_{\ell=RN}^{M-1} h_\tau^N h_\eta^\ell \Big[g_\eta^\ell a_0^{\boldsymbol{\eta}}(\overline{v}_t^{\boldsymbol{\eta}}; u_N^{\boldsymbol{\tau}}) + \eta \sum_{n=1}^{\ell} g_\eta^{\ell-n}\; b_n^{\boldsymbol{\eta}}(\overline{v}_t^{\boldsymbol{\eta}}; u_N^{\boldsymbol{\tau}}) \Big] \tag{5.1.22}$$

$$+ \sum_{k=0}^{N-1} h_\tau^k h_\eta^{Rk} \Big[g_\tau^k a_0^{\boldsymbol{\tau}}(\overline{u}_t^{\boldsymbol{\tau}}; v_{Rk}^{\boldsymbol{\eta}}) + \tau \sum_{n=1}^{k} g_\tau^{k-n}\; b_n^{\boldsymbol{\tau}}(\overline{u}_t^{\boldsymbol{\tau}}; v_{Rk}^{\boldsymbol{\eta}}) \Big] \tag{5.1.23}$$

$$+ \sum_{k=0}^{N-1} \sum_{\ell=Rk}^{R(k+1)-1} h_\tau^{k+1} h_\eta^\ell \Big[g_\eta^\ell a_0^{\boldsymbol{\eta}}(\overline{v}_t^{\boldsymbol{\eta}}; u_{k+1}^{\boldsymbol{\tau}}) + \eta \sum_{n=1}^{\ell} g_\eta^{\ell-n}\; b_n^{\boldsymbol{\eta}}(\overline{v}_t^{\boldsymbol{\eta}}; u_{k+1}^{\boldsymbol{\tau}}) \Big]. \tag{5.1.24}$$

The core part of the proof of Theorem 5.1.13 is to show that under the given hypotheses,

$$I^{\boldsymbol{\tau},\boldsymbol{\eta}}_{N,M}(\overline{u}^{\boldsymbol{\tau}}_t,\overline{v}^{\boldsymbol{\eta}}_t)\le C'\tau. \tag{5.1.25}$$

The proof of (5.1.25) can be found at the end of this section. In conclusion, we have

$$e^{\lambda_\tau t+\lambda_\eta t}\boldsymbol{d}^2(\overline{u}^{\boldsymbol{\tau}}_t,\overline{v}^{\boldsymbol{\eta}}_t)\le h^N_\tau h^M_\eta \boldsymbol{d}^2(u^{\boldsymbol{\tau}}_N,v^{\boldsymbol{\eta}}_M)\le C'\tau+\boldsymbol{d}^2(u_0,v_0),$$

which implies the inequality (5.1.19) with $C=e^{-2\lambda T}(1+C')$. □

Proof of the Main Theorem

With Theorem 5.1.13 at hand, we finish the proof of Theorem 5.1.11.

Proof of Theorem 5.1.11. From Theorem 5.1.13 it follows that $(\overline{u}^{\boldsymbol{\tau}_n}_t)_{n\in\mathbb{N}}$ is a Cauchy family with respect to uniform convergence on each interval $[0,T]$. Indeed, from estimate (5.1.19), it follows that the values $(\overline{u}^{\boldsymbol{\tau}_n}_t)_{n\in\mathbb{N}}$ converge in the complete metric space $(\mathbf{X},\boldsymbol{d})$ to a limit u^*_t, uniformly for $t\in[0,T]$, and that the estimate (5.1.18) holds. Since this argument holds for arbitrary $T>0$, the limit u^*_t is defined for all $t\ge 0$.

To prove absolute continuity of the limit curve u_*, we argue as usual: we assign time-discrete derivatives A^n_t to the interpolated solutions $\overline{u}^{\boldsymbol{\tau}_n}_t$ by

$$A^n_t:=\frac{\boldsymbol{d}(\overline{u}^{\boldsymbol{\tau}_n}_{t-\tau_n},\overline{u}^{\boldsymbol{\tau}_n}_t)}{\tau_n}=\frac{\boldsymbol{d}(u^{\boldsymbol{\tau}_n}_{k\text{-}1},u^{\boldsymbol{\tau}_n}_k)}{\tau_n}\qquad \text{for } t\in((k-1)\tau_n,k\tau_n].$$

Thanks to the classical estimate (5.1.14), A^n_t is uniformly bounded in $\mathrm{L}^2(0,T)$. Hence, A^n_t possesses a $\mathrm{L}^2(0,T)$-weakly convergent subsequence (not relabelled) with limit $A_t\in\mathrm{L}^2(0,T)$. Choose arbitrary s,t with $0\le s\le t\le T$, and define $k^n_r:=\max\{k\mid k\tau_n\le r\}$, then

$$\boldsymbol{d}(\overline{u}^{\boldsymbol{\tau}_n}_s,\overline{u}^{\boldsymbol{\tau}_n}_t)\le\sum_{k=k^n_s+1}^{k^n_t}\boldsymbol{d}(u^{\boldsymbol{\tau}_n}_{k\text{-}1},u^{\boldsymbol{\tau}_n}_k)=\int_{k^n_s\tau_n}^{k^n_t\tau_n}A^n_r\,\mathrm{d}r.$$

In the limit $n\to\infty$, this yields

$$\boldsymbol{d}(u^*_s,u^*_t)=\lim_{n\to\infty}\boldsymbol{d}(\overline{u}^{\boldsymbol{\tau}_n}_s,\overline{u}^{\boldsymbol{\tau}_n}_t)\le\lim_{n\to\infty}\int_{k^n_s\tau_n}^{k^n_t\tau_n}A^n_r\,\mathrm{d}r=\int_s^t A_r\,\mathrm{d}r.$$

Hence $u_*\in\mathrm{AC}^2(0,\infty;(\mathbf{X},\boldsymbol{d}))$.

It remains to prove that the limit curve u_* satisfies the integrated form (2.3.2) of the EVI. Again, let $0\le s\le t\le T$, and define k^n_r be as above. Multiply the time-discrete EVI (5.1.17) for $(u^{\boldsymbol{\tau}_n}_k)_{k\in\mathbb{N}}$ by τ_n, and sum from $k=k^n_s$ to $k=k^n_t-1$. We define the left-hand side:

$$J^n_{(1)}(s,t):=\tau_n\sum_{k=k^n_s}^{k^n_t-1}\Big[\Big(\frac{3}{4\tau_n}+\frac{\lambda}{2}\Big)\boldsymbol{d}^2(u^{\boldsymbol{\tau}_n}_{k+1},w)-\frac{1}{\tau_n}\boldsymbol{d}^2(u^{\boldsymbol{\tau}_n}_k,w)+\frac{1}{4\tau_n}\boldsymbol{d}^2(u^{\boldsymbol{\tau}_n}_{k\text{-}1},w)\Big].$$

Consequently, after elementary manipulations we have

$$J^n_{(1)}(s,t) = \frac{\lambda}{2}\int_{k^n_s\tau_n}^{k^n_t\tau_n} \boldsymbol{d}^2(\overline{u}^{\tau_n}_r, w)\,\mathrm{d}r + \frac{1}{4}\Big[\big(3\boldsymbol{d}^2(u^{\tau_n}_{k^n_t}, w) - \boldsymbol{d}^2(u^{\tau_n}_{k^n_t-1}, w)\big) - \big(3\boldsymbol{d}^2(u^{\tau_n}_{k^n_s}, w) - \boldsymbol{d}^2(u^{\tau_n}_{k^n_s-1}, w)\big)\Big].$$

Thanks to the r-uniform convergence of $\overline{u}^{\tau_n}_r$ to u^*_r, and since u^*_t is continuous, we obtain in the limit

$$\lim_{n\to\infty} J^n_{(1)}(s,t) = \frac{\lambda}{2}\int_s^t \boldsymbol{d}^2\big(u^*_r, w\big)\,\mathrm{d}r + \frac{1}{2}\boldsymbol{d}^2\big(u^*_t, w\big) - \frac{1}{2}\boldsymbol{d}^2\big(u^*_s, w\big).$$

On the other hand, after summation of the right-hand side of (5.1.17), we estimate once again with the help of the elementary inequality (5.1.13) and thus obtain

$$\begin{aligned} J^n_{(2)}(s,t) &:= \tau_n \sum_{k=k^n_s}^{k^n_t-1} \Big[\mathcal{E}(w) - \mathcal{E}(u^{\tau_n}_{k+1}) - \frac{1}{\tau_n}\boldsymbol{d}^2(u^{\tau_n}_k, u^{\tau_n}_{k+1}) + \frac{1}{4\tau_n}\boldsymbol{d}^2(u^{\tau_n}_{k-1}, u^{\tau_n}_{k+1})\Big] \\ &\le \int_{k^n_s\tau_n}^{k^n_t\tau_n} \big[\mathcal{E}(w) - \mathcal{E}\big(\overline{u}^{\tau_n}_r\big)\big]\,\mathrm{d}r - \frac{1}{2}\boldsymbol{d}^2(u^{\tau_n}_{k^n_t-1}, u^{\tau_n}_{k^n_t}) + \frac{1}{2}\boldsymbol{d}^2(u^{\tau_n}_{k^n_s-1}, u^{\tau_n}_{k^n_s}). \end{aligned}$$

Again, thanks to local uniform convergence of $\overline{u}^{\tau_n}_t$ to the continuous limit u^*_t, and since $\mathcal{E}$ is lower semi-continuous thanks to Assumption (E1), Fatou's lemma yields that

$$\lim_{n\to\infty} J^n_{(2)}(s,t) \le \int_s^t \big[\mathcal{E}(w) - \mathcal{E}\big(u^*_r\big)\big]\,\mathrm{d}r.$$

Since $J^n_{(1)}(s,t) \le J^n_{(2)}(s,t)$ for all n by (5.1.17), the respective inequality follows for the limits, that is

$$\frac{\lambda}{2}\int_s^t \boldsymbol{d}^2\big(u^*_r, w\big)\,\mathrm{d}r + \frac{1}{2}\boldsymbol{d}^2\big(u^*_t, w\big) - \frac{1}{2}\boldsymbol{d}^2\big(u^*_s, w\big) \le \int_s^t \big[\mathcal{E}(w) - \mathcal{E}\big(u^*_r\big)\big]\,\mathrm{d}r.$$

This implies the integrated EVI (2.3.2). □

Proof of the Estimate (5.1.25)

This is the most technical part of the convergence proof, that uses only elementary inequalities and the classical estimates (5.1.14) (which, in turn, are valid thanks to (I1)). Throughout this section, we adopt the convenient notation that C is a generic constant, which is in principle expressible in terms of d_1 and d_2 from (I1) and the terminal time T alone, and whose value may change from one line to the next.

To begin with, observe that since we assumed $\lambda \le 0$, we have that $g_\tau \le \frac{1}{3}$ and $g_\eta \le \frac{1}{3}$, and therefore

$$\sum_{k=0}^{\infty} g_\tau^k \leq \frac{3}{2}, \quad \sum_{\ell=0}^{\infty} g_\eta^\ell \leq \frac{3}{2}. \tag{5.1.26}$$

Further, we have that $h_\tau \leq 1$ and $h_\eta \leq 1$, which means that

$$h_\tau^k h_\eta^\ell \leq 1 \tag{5.1.27}$$

for arbitrary $k, \ell \geq 0$. On the other hand, since $h_\tau \leq 1$ and due to (5.1.16), it follows that

$$a_0^{\boldsymbol{\tau}}(\overline{u}_t^{\boldsymbol{\tau}}; w) = h_\tau \boldsymbol{d}^2(u_1^{\boldsymbol{\tau}}, w) - \boldsymbol{d}^2(u_0^{\boldsymbol{\tau}}, w) \leq \boldsymbol{d}^2(u_1^{\boldsymbol{\tau}}, w) - \boldsymbol{d}^2(u_0^{\boldsymbol{\tau}}, w) \leq 2\boldsymbol{d}(u_0^{\boldsymbol{\tau}}, u_1^{\boldsymbol{\tau}})\boldsymbol{d}(u_1^{\boldsymbol{\tau}}, w).$$

Substituting $w := v_\ell^{\boldsymbol{\eta}}$, we obtain by the triangle inequality, and thanks to estimate (5.1.14), that

$$a_0^{\boldsymbol{\tau}}(\overline{u}_t^{\boldsymbol{\tau}}; v_\ell^{\boldsymbol{\eta}}) \leq 2\boldsymbol{d}(u_0^{\boldsymbol{\tau}}, u_1^{\boldsymbol{\tau}})\big[\boldsymbol{d}(u_*, u_1^{\boldsymbol{\tau}}) + \boldsymbol{d}(u_*, v_\ell^{\boldsymbol{\eta}})\big] \leq C\boldsymbol{d}(u_0^{\boldsymbol{\tau}}, u_1^{\boldsymbol{\tau}}) \leq C d_2 \tau \tag{5.1.28}$$

where we have used that (I2) holds with constant d_2. Analogously, one derives

$$a_0^{\boldsymbol{\tau}}(\overline{v}_t^{\boldsymbol{\eta}}; u_k^{\boldsymbol{\tau}}) \leq C\eta. \tag{5.1.29}$$

With (5.1.26), (5.1.27), (5.1.28) and (5.1.29) at hand, it is now straight-forward to estimate the terms inside $I_{N,M}^{\boldsymbol{\tau};\boldsymbol{\eta}}(\overline{u}_t^{\boldsymbol{\tau}}, \overline{v}_t^{\boldsymbol{\eta}})$ involving $a_0^{\boldsymbol{\tau}}$ or $a_0^{\boldsymbol{\eta}}$. For the expression in (5.1.22),

$$\sum_{\ell=RN}^{M-1} h_\tau^N h_\eta^\ell g_\eta^\ell a_0^{\boldsymbol{\eta}}(\overline{v}_t^{\boldsymbol{\eta}}; u_N^{\boldsymbol{\tau}}) \leq \sum_{\ell=RN}^{M-1} g_\eta^\ell C\eta \leq \frac{3}{2}C\eta.$$

For (5.1.23),

$$\sum_{k=0}^{N-1} h_\tau^k h_\eta^{Rk} g_\tau^k a_0^{\boldsymbol{\tau}}(\overline{u}_t^{\boldsymbol{\tau}}; v_{Rk}^{\boldsymbol{\eta}}) \leq \sum_{k=0}^{N-1} g_\tau^k C\tau \leq \frac{3}{2}C\tau.$$

And finally, for (5.1.24),

$$\sum_{k=0}^{N-1} \sum_{l=Rk}^{R(k+1)-1} h_\tau^{k+1} h_\eta^\ell g_\eta^\ell a_0^{\boldsymbol{\eta}}(\overline{v}_t^{\boldsymbol{\eta}}; u_{k+1}^{\boldsymbol{\tau}}) \leq \sum_{\ell=0}^{RN} g_\eta^\ell C\eta \leq \frac{3}{2}C\eta.$$

We turn to estimate the terms involving $b_k^{\boldsymbol{\tau}}$ and $b_\ell^{\boldsymbol{\eta}}$. First, by (5.1.13),

$$\begin{aligned} b_k^{\boldsymbol{\tau}}(\overline{u}_t^{\boldsymbol{\tau}}; w) \leq & 4g_\tau\Big(\mathcal{E}(w) - \mathcal{E}(u_{k+1}^{\boldsymbol{\tau}}) + \frac{1}{2\tau}\boldsymbol{d}^2(u_{k-1}^{\boldsymbol{\tau}}, u_k^{\boldsymbol{\tau}})\Big) \\ \leq & 4g_\tau\left(\mathcal{E}(w) - \mathcal{E}(u_{k+1}^{\boldsymbol{\tau}})\right) + \frac{1}{\tau}\boldsymbol{d}^2(u_{k-1}^{\boldsymbol{\tau}}, u_k^{\boldsymbol{\tau}}), \end{aligned}$$

where we have used that $g_\tau \le 1/2$. An analogous estimate holds for $b_\ell^{\boldsymbol{\eta}}$. Substitute this into the expression for $I_{N,M}^{\boldsymbol{\tau},\boldsymbol{\eta}}(\overline{u}_t^{\boldsymbol{\tau}}, \overline{v}_t^{\boldsymbol{\eta}})$ in (5.1.22)–(5.1.24). This gives rise to two groups of terms: one related to the metric, the other related to $\mathcal{E}$.

We begin by estimating the terms related to the metric. This is done using the classical estimate (5.1.14): for the expression in (5.1.22),

$$\eta \sum_{\ell=RN}^{M-1} h_\tau^N h_\eta^\ell \sum_{n=1}^{\ell} g_\eta^{\ell-n} \frac{\boldsymbol{d}^2(v_{n\text{-}1}^{\boldsymbol{\eta}}, v_n^{\boldsymbol{\eta}})}{\eta} \le \eta \sum_{\ell=RN}^{M-1} \sum_{n=1}^{\ell} \frac{\boldsymbol{d}^2(v_{n\text{-}1}^{\boldsymbol{\eta}}, v_n^{\boldsymbol{\eta}})}{\eta} \le R\eta \sum_{\ell=1}^{M-1} \frac{\boldsymbol{d}^2(v_{\ell\text{-}1}^{\boldsymbol{\eta}}, v_\ell^{\boldsymbol{\eta}})}{\eta} \le C\tau.$$

Here we have used that $M - NR \le R$ and that $R\eta = \tau$. For (5.1.23),

$$\tau \sum_{k=0}^{N-1} h_\tau^k h_\eta^{Rk} \sum_{n=1}^{k} g_\tau^{k-n} \frac{\boldsymbol{d}^2(u_{n\text{-}1}^{\boldsymbol{\tau}}, u_N^{\boldsymbol{\tau}})}{\tau} \le \tau \sum_{n=1}^{N-1} \Big[\Big(\sum_{k=n}^{N-1} g_\tau^{k-n} \Big) \frac{\boldsymbol{d}^2(u_{n\text{-}1}^{\boldsymbol{\tau}}, u_N^{\boldsymbol{\tau}})}{\tau} \Big] \le \frac{3}{2} C\tau.$$

Finally, for (5.1.24),

$$\begin{aligned} \eta \sum_{k=0}^{N-1} \sum_{\ell=Rk}^{R(k+1)-1} h_\tau^{k+1} h_\eta^\ell \sum_{n=1}^{\ell} g_\eta^{\ell-n} \frac{\boldsymbol{d}^2(v_{n\text{-}1}^{\boldsymbol{\eta}}, v_n^{\boldsymbol{\eta}})}{\eta} &\le \eta \sum_{\ell=1}^{RN} \sum_{n=1}^{\ell} g_\eta^{\ell-n} \frac{\boldsymbol{d}^2(v_{n\text{-}1}^{\boldsymbol{\eta}}, v_n^{\boldsymbol{\eta}})}{\eta} \\ &\le \eta \sum_{n=1}^{RN} \Big[\Big(\sum_{\ell=n}^{RN} g_\eta^{\ell-n} \Big) \frac{\boldsymbol{d}^2(v_{n\text{-}1}^{\boldsymbol{\eta}}, v_n^{\boldsymbol{\eta}})}{\eta} \Big] \le \frac{3}{2} C\eta. \end{aligned}$$

The estimates on the expressions involving the differences of the energy values are a bit more involved. To simplify calculations, we use that the b's only contain the difference between two values of $\mathcal{E}$; hence adding a constant to $\mathcal{E}$ does not change the b values. Consequently, since $\mathcal{E}(u_k^{\boldsymbol{\tau}})$ and $\mathcal{E}(v_\ell^{\boldsymbol{\eta}})$ are bounded from below thanks to (5.1.14), we may assume without loss of generality that all $\mathcal{E}(u_k^{\boldsymbol{\tau}})$ and $\mathcal{E}(v_\ell^{\boldsymbol{\eta}})$ are non-negative.

The contribution of the $\mathcal{E}$ terms to (5.1.22) is immediately controlled, recalling (5.1.26), (5.1.27), and that $M < R(N+1)$:

$$4\eta \sum_{\ell=RN}^{M-1} h_\tau^N h_\eta^\ell \sum_{n=1}^{\ell} g_\eta^{\ell+1-n} \big[\mathcal{E}(u_N^{\boldsymbol{\tau}}) - \mathcal{E}(v_n^{\boldsymbol{\eta}}) \big] \le 4R\eta \mathcal{E}(u_N^{\boldsymbol{\tau}}) 4 \sum_{\ell=1}^{M} g_\eta^l \le \frac{3}{2} C\tau.$$

Collecting all terms containing evaluations of $\mathcal{E}$ in (5.1.23) and (5.1.24) yields:

$$\begin{aligned} &4\tau \sum_{k=0}^{N-1} h_\tau^k h_\eta^{Rk} \sum_{n=1}^{k} g_\tau^{k-n+1} \big[\mathcal{E}(v_{Rk}^{\boldsymbol{\eta}}) - \mathcal{E}(u_{n+1}^{\boldsymbol{\tau}}) \big] \\ &+ 4\eta \sum_{k=0}^{N-1} \sum_{\ell=Rk}^{R(k+1)-1} h_\tau^{k+1} h_\eta^\ell \sum_{n=1}^{\ell} g_\eta^{\ell-n+1} \big[\mathcal{E}(u_{k+1}^{\boldsymbol{\tau}}) - \mathcal{E}(v_{n+1}^{\boldsymbol{\eta}}) \big] \\ &= 4I_1 + 4I_2, \end{aligned}$$

where I_1 and I_2 collect terms with $\mathcal{E}(u_k^{\boldsymbol{\tau}})$ and $\mathcal{E}(v_\ell^{\boldsymbol{\eta}})$, respectively,

$$
\begin{aligned}
I_1 &= \tau \sum_{k=0}^{N-1} \mathcal{E}(u_{k+1}^{\boldsymbol{\tau}}) \Big[\frac{1}{R} \sum_{\ell=Rk}^{R(k+1)-1} h_\tau^{k+1} h_\eta^\ell \sum_{n=1}^{\ell} g_\eta^{\ell-n+1} - \sum_{n=k}^{N-1} h_\tau^n h_\eta^{Rn} g_\tau^{n-k+1} \Big] \\
&\quad + \tau \mathcal{E}(u_1^{\boldsymbol{\tau}}) \sum_{n=0}^{N-1} h_\tau^n h_\eta^{Rn} g_\tau^{n+1}, \\
I_2 &= \tau \sum_{k=0}^{N-1} \Big[\mathcal{E}(v_{Rk}^{\boldsymbol{\eta}}) h_\tau^k h_\eta^{Rk} \sum_{n=1}^{k} g_\tau^{k-n+1} \Big] - \eta \sum_{k=0}^{N-1} \sum_{\ell=Rk}^{R(k+1)-1} \sum_{n=1}^{\ell} h_\tau^{k+1} h_\eta^\ell g_\eta^{\ell-n+1} \mathcal{E}(v_{n+1}^{\boldsymbol{\eta}}).
\end{aligned}
$$

For I_1, we obtain

$$
\begin{aligned}
I_1 &\le \tau \sum_{k=0}^{N-1} \mathcal{E}(u_{k+1}^{\boldsymbol{\tau}})\, h_\tau^k h_\eta^{Rk} \Big[\frac{1}{R} h_\tau \sum_{\ell=0}^{R-1} \Big(h_\eta^\ell \sum_{n=1}^{Rk+l} g_\eta^n \Big) - g_\tau \sum_{n=0}^{N-k-1} (h_\tau h_\eta^R g_\tau)^n \Big] + \tau \mathcal{E}(u_1^{\boldsymbol{\tau}}) \sum_{n=0}^{N-1} g_\tau^n \\
&\le \tau \sum_{k=0}^{N-1} \mathcal{E}(u_{k+1}^{\boldsymbol{\tau}})\, h_\tau^k h_\eta^{Rk} \Big[\frac{1}{R} \frac{1}{3} \Big(\sum_{l=0}^{R-1} \frac{3}{2} \Big) - g_\tau \frac{1 - (h_\tau h_\eta^R g_\tau)^{N-k}}{1 - h_\tau h_\eta^R g_\tau} \Big] + C\tau \\
&\le \tau \sum_{k=0}^{N} \mathcal{E}(u_{k+1}^{\boldsymbol{\tau}})\, h_\tau^k h_\eta^{Rk} \left[\frac{1}{2} - \frac{g_\tau}{1 - h_\tau h_\eta^R g_\tau} + \frac{g_\tau^{N-k+1}}{1 - h_\tau h_\eta^R g_\tau} \right] + C\tau.
\end{aligned}
$$

Recalling that $-1 \le \lambda\tau \le 0$, and observing that both g_τ and h_τ are convex functions of τ, a Taylor expansion yields that

$$
g_\tau \ge \frac{1}{3}\Big(1 + \frac{1}{3}\lambda\tau\Big) \ge 0, \quad h_\tau \ge 1 + \lambda\tau \ge 0, \tag{5.1.30}
$$

and similarly for g_η and h_η. Therefore, in combination with Bernoulli's inequality,

$$
h_\tau h_\eta^R g_\tau \ge (1 + \lambda\tau)(1 + R\lambda\eta) \frac{1}{3}\Big(1 + \frac{1}{2}\lambda\tau\Big) \ge \frac{1}{3}(1 + \lambda\tau)^3.
$$

With another application of (5.1.30),

$$
\begin{aligned}
\frac{1}{2} - \frac{g_\tau}{1 - h_\tau h_\eta^R g_\tau} &\le \frac{1}{2} - \frac{\frac{1}{3}(1 + \lambda\tau)}{1 - \frac{1}{3}(1 + \lambda\tau)^3} \le \frac{1}{2} - \frac{(1 + \lambda\tau)^3}{3 - (1 + \lambda\tau)^3} \\
&= \frac{3}{2} \cdot \frac{1 - (1 + \lambda\tau)^3}{3 - (1 + \lambda\tau)^3} \le \frac{3}{2} \cdot \frac{(-3)\lambda\tau}{2} = -\frac{9}{4}\lambda\tau.
\end{aligned} \tag{5.1.31}
$$

This yields, in combination with the bound (5.1.14) on $\mathcal{E}$,

$$
\begin{aligned}
I_1 \le& \tau \sum_{k=0}^{N-1} \mathcal{E}(u_{k+1}^{\boldsymbol{\tau}})\, h_\tau^k h_\eta^{Rk} \Big[\frac{g_\tau^{N-k+1}}{1 - h_\tau h_\eta^R g_\tau} - \frac{9}{4}\lambda\tau \Big] + C\tau \\
\le& \tau C \sum_{k=1}^{N} \left[\frac{3}{2} g_\tau^{N-k+1} - \frac{9}{4}\lambda\tau \right] + C\tau \le C[1 + (-\lambda)T]\ \tau.
\end{aligned}
$$

We turn to the estimate of I_2. In order to merge the difference of the two sums into a single sum — similar to what we did for I_1 above — we are going to apply a shift of no more than R to the indices inside $\mathcal{E}(v^{\boldsymbol{\eta}}_{Rk})$. To control the error introduced by that shift, observe that an iteration of the energy estimate (5.1.11) yields

$$\mathcal{E}(v^{\boldsymbol{\eta}}_{Rk}) \le \mathcal{E}(v^{\boldsymbol{\eta}}_{\ell}) + \frac{1}{4\eta}\boldsymbol{d}^2(v^{\boldsymbol{\eta}}_{\ell\text{-}1}, v^{\boldsymbol{\eta}}_{\ell})$$

as soon as $0 \le \ell \le Rk$. Further, for such k and ℓ, we have $h_\tau^k \le h_\tau^{\ell/R}$ since $h_\tau \le 1$. This allows us to estimate the first sum in I_2 as follows:

$$\begin{aligned}\tau \sum_{k=1}^{N-1} \Big[\mathcal{E}(v^{\boldsymbol{\eta}}_{Rk}) h_\tau^k h_\eta^{Rk} g_\tau \sum_{n=1}^{k} g_\tau^{k-n}\Big] &\le \frac{\tau}{2}\sum_{k=1}^{N-1} h_\tau^k h_\eta^{Rk}\mathcal{E}(v^{\boldsymbol{\eta}}_{Rk}) \\ &\le \frac{\tau}{2R}\sum_{k=1}^{N}\sum_{\ell=R(k-1)+1}^{Rk} h_\tau^{\ell/R} h_\eta^{\ell}\Big[\mathcal{E}(v^{\boldsymbol{\eta}}_{\ell}) + \frac{1}{4\eta}\boldsymbol{d}^2(v^{\boldsymbol{\eta}}_{\ell\text{-}1}, v^{\boldsymbol{\eta}}_{\ell})\Big] \\ &\le \frac{\eta}{2}\sum_{\ell=1}^{R(N-1)} h_\tau^{\ell/R} h_\eta^{\ell}\mathcal{E}(v^{\boldsymbol{\eta}}_{\ell}) + \frac{\eta}{4}\sum_{\ell=1}^{R(N-1)} \frac{\boldsymbol{d}^2(v^{\boldsymbol{\eta}}_{\ell\text{-}1}, v^{\boldsymbol{\eta}}_{\ell})}{2\eta}.\end{aligned}$$

Next, use the classical estimate (5.1.14) and a lower bound for h_τ to obtain

$$\tau \sum_{k=1}^{N-1} \Big[\mathcal{E}(v^{\boldsymbol{\eta}}_{Rk}) h_\tau^k h_\eta^{Rk} g_\tau \sum_{n=1}^{k} g_\tau^{k-n}\Big] \le \frac{\eta}{2}\sum_{\ell=0}^{RN-1} h_\tau^{\ell/R} h_\eta^{\ell}\mathcal{E}(v^{\boldsymbol{\eta}}_{\ell+1}) + C\eta.$$

The second sum in I_2 is estimated as follows, using that $h_\tau^k \ge h_\tau^{\ell/R}$ for $Rk \le \ell$:

$$\begin{aligned}&\eta \sum_{k=0}^{N-1}\sum_{\ell=Rk}^{R(k+1)-1}\sum_{n=1}^{l} h_\tau^{k+1} h_\eta^{\ell} g_\eta^{\ell-n+1}\mathcal{E}(v^{\boldsymbol{\eta}}_{n+1}) \ge \eta h_\tau g_\eta \sum_{\ell=0}^{RN-1}\sum_{n=1}^{\ell} h_\tau^{\ell/R} h_\eta^{\ell} g_\eta^{\ell-n}\mathcal{E}(v^{\boldsymbol{\eta}}_{n+1}) \\ &= \eta g_\eta h_\tau \sum_{n=0}^{RN-1}\mathcal{E}(v^{\boldsymbol{\eta}}_{n+1}) h_\tau^{n/R} h_\eta^{n} \sum_{\ell=0}^{RN-n-1}\big(h_\tau^{1/R} h_\eta g_\eta\big)^n - \eta h_\tau g_\eta \mathcal{E}(v^{\boldsymbol{\eta}}_1)\sum_{\ell=0}^{RN-1} h_\tau^{\ell/R} h_\eta^{\ell} g_\eta^{\ell} \\ &\ge \eta g_\eta h_\tau \sum_{n=0}^{RN-1}\mathcal{E}(v^{\boldsymbol{\eta}}_{n+1}) h_\tau^{n/R} h_\eta^{n}\frac{1-(h_\tau^{1/R} h_\eta g_\eta)^{RN-n}}{1-h_\tau^{1/R} h_\eta g_\eta} - C\eta.\end{aligned}$$

Substituting these estimates into the expression for I_2 yields a single sum,

$$I_2 \le C\eta + \eta\sum_{\ell=0}^{RN-1} h_\tau^{\ell/R} h_\eta^{\ell}\mathcal{E}(v^{\boldsymbol{\eta}}_{\ell+1})\Big[\frac{1}{2} - \frac{g_\eta h_\tau}{1-h_\tau^{1/R} h_\eta g_\eta} + \frac{g_\eta^{RN-\ell+1}}{1-h_\tau^{1/R} h_\eta g_\eta}\Big].$$

Arguing similarly as in the derivation of (5.1.31), we estimate

$$h_\tau^{1/R} h_\eta g_\eta \ge (1+\lambda\tau)^{1/R}(1+\lambda\eta)\frac{1}{3}\Big(1+\frac{1}{3}\lambda\eta\Big) \ge \frac{1}{3}(1+\lambda\tau)^3.$$

and consequently, just as in (5.1.31),

$$\frac{1}{2} - \frac{g_\eta h_\tau}{1 - h_\tau^{1/R} h_\eta g_\eta} \leq \frac{1}{2} - \frac{\frac{1}{3}(1+\lambda\tau)^2}{1 - \frac{1}{3}(1+\lambda\tau)^3} \leq -\frac{9}{4}\lambda\tau.$$

In conclusion, we obtain with the help of (5.1.14) that

$$I_2 \leq C\eta + \eta \sum_{\ell=0}^{RN-1} \mathcal{E}(v_{\ell+1}^{\boldsymbol{\eta}}) \Big[\frac{3}{2} g_\eta^{RN+1-\ell} + \frac{9}{4}\lambda\tau\Big] C[1 + (-\lambda)T]\ \tau.$$

Collecting all terms, we finally obtain the desired estimate (5.1.25). □

5.2 Application to Non-linear Fokker-Planck Equation

This section, based on my own work [81], is concerned with the proof of well-posedness and convergence of a formally higher-order semi-discretization in time, inspired by the Backward Differentiation Formula 2 (BDF2), applied to the *non-linear Fokker-Planck equation with no-flux boundary condition*:

$$\begin{aligned} \partial_t \rho_t = \Delta(\rho_t^m) + \operatorname{div}(\rho_t \nabla V) + \operatorname{div}(\rho_t \nabla (W * \rho_t)) \qquad &\text{in } (0,\infty)\times\Omega, \\ \boldsymbol{n}\cdot \mathrm{D}\,\rho = 0, \quad \text{on } (0,\infty)\times\partial\Omega, \qquad\qquad & \rho(0,x) = \rho_0(x) \quad \text{in } \Omega. \end{aligned} \tag{5.2.1}$$

We consider (5.2.1) as an evolutionary equation in the space of probability measures $\mathcal{P}_2(\Omega)$ with finite second moment (i.e $\boldsymbol{M}_2(\mu) := \int_\Omega \|x\|^2 \,\mathrm{d}\mu(x) < \infty$), where $\Omega = \mathbb{R}^{\mathrm{d}}$ or $\Omega \subset \mathbb{R}^{\mathrm{d}}$ is an open, bounded, and connected domain with Lipschitz-continuous boundary $\partial\Omega$ and normal derivative $\boldsymbol{n}$. Indeed, if (5.2.1) is initialized with $\rho_0 \in \mathcal{P}_2(\Omega)$ then there exists a weak solution $\rho_t^* : [0,\infty)\times\Omega \to \mathbb{R}_{\geq 0}$ such that $\rho_0^* = \rho_0$ and $\rho_t^* \in \mathcal{P}_2(\Omega)$ for each $t > 0$.

As recalled in section 2.4, the modern approach towards the theoretical analysis of equation (5.2.1) is the gradient flow structure in the L^2-Wasserstein space $(\mathcal{P}_2(\Omega), \mathbf{W}_2)$, see [4, 54, 79, 87, 92, 93]. The corresponding free energy functional $\mathcal{E} : \mathcal{P}_2(\Omega) \to \mathbb{R}\cup\{\infty\}$ for (5.2.1) is given by:

$$\mathcal{E}(\mu) := \begin{cases} \int_\Omega \rho\log(\rho) + V\rho + \frac{1}{2}(W*\rho)\rho \,\mathrm{d}x & \text{if } m = 1, \\ \int_\Omega \frac{1}{m-1}\rho^m + V\rho + \frac{1}{2}(W*\rho)\rho \,\mathrm{d}x & \text{if } m > 1, \end{cases} \tag{5.2.2}$$

provided that $\mu = \rho\mathcal{L}^d$ and the integrals on the right-hand side are well-defined otherwise we set $\mathcal{E}(\mu) = \infty$. We want to emphasize, that we don't assume any convexity property on the confinement potential V nor on the interaction kernel W. Hence, the corresponding free energy functional $\mathcal{E}$ does not satisfy the convexity assumption (E3) from the previous section 5.1 and we cannot apply the theory developed therein.

Method. Using the notation of the L^2-Wasserstein framework, the approximation scheme via the variational formulation of the BDF2 method reads than as:

Scheme. For each equidistant partition $\boldsymbol{\tau} = (\tau, 2\tau, 3\tau, \ldots)$ with sufficiently small time step $\tau > 0$, let a pair of initial conditions $(\rho_0^{\boldsymbol{\tau}}, \rho_1^{\boldsymbol{\tau}})$ be given that approximate ρ_0. Then define inductively a discrete solution $(\rho_k^{\boldsymbol{\tau}})_{k\in\mathbb{N}}$ such that each $\rho_{k+1}^{\boldsymbol{\tau}}$ with $k \in \mathbb{N}$ is a minimizer of the following functional,

$$\rho \mapsto \Psi(\tau, \rho_{k\text{-}1}^{\boldsymbol{\tau}}, \rho_k^{\boldsymbol{\tau}}; \rho) := \frac{1}{\tau}\mathbf{W}_2^2(\rho_k^{\boldsymbol{\tau}}, \rho) - \frac{1}{4\tau}\mathbf{W}_2^2(\rho_{k\text{-}1}^{\boldsymbol{\tau}}, \rho) + \mathcal{E}(\rho).$$

Define the corresponding piecewise constant interpolation in time $\overline{\rho}_t^{\boldsymbol{\tau}} : [0,\infty) \to \mathcal{P}_2(\Omega)$ of the discrete solution $\rho_k^{\boldsymbol{\tau}}$ in time via

$$\overline{\rho}_0^{\boldsymbol{\tau}} = \rho_0^{\boldsymbol{\tau}}, \quad \overline{\rho}_t^{\boldsymbol{\tau}} = \rho_k^{\boldsymbol{\tau}} \quad \text{for } t \in ((k-1)\tau, k\tau] \quad \text{and } k \in \mathbb{N}.$$

Strategy of the Proof. The structure of the proof is similar to the procedure from section 3.2. However, existence of a minimizer ρ_k^τ is a priori not clear, since the negative L^2-Wasserstein distance is not lower semi-continuous with respect to narrow convergence. We circumvent this problem by considering the auxiliary functional $\mathcal{A}(\mu) := 4\mathbf{W}_2^2(\nu,\mu) - \mathbf{W}_2^2(\eta,\mu)$ which turns out to be lower semi-continuous, see section 5.2.2. In section 5.2.3, we derive the intrinsic properties of the discrete solution $(\rho_k^\tau)_{k\in\mathbb{N}}$ by the use of the variational formulation of the BDF2 method. We want to mention here the time-discrete Euler-Lagrange equations which are obtained by perturbing each minimizer ρ_k^τ along solutions to the transport equation. These Euler-Lagrange equations comprise enough structural information to deduce the refined a priori estimates on the regularity of $(\rho_k^\tau)^m$ in the $\mathrm{BV}(\Omega)$-norm. In section 5.2.4 we complete the proof and show the strong convergence of the approximation to the weak solution of the non-linear Fokker-Planck equation (5.2.1) with respect to the narrow-topology and with respect to the strong $\mathrm{L}^p(0,T;\mathrm{L}^m(\Omega))$-topology.

5.2.1 Setup and Assumptions

Similarly to [70], the *BDF2 penalization* $\Psi : (0,\tau_*)\times(\mathcal{P}_2(\Omega))^3 \to \mathbb{R}\cup\{\infty\}$ of the original energy functional $\mathcal{E}$ is defined by

$$\Psi(\tau,\eta,\nu;\cdot) : \mathcal{P}_2(\Omega) \to \mathbb{R}\cup\{\infty\};\ \Psi(\tau,\eta,\nu;\rho) := \frac{1}{\tau}\mathbf{W}_2^2(\nu,\rho) - \frac{1}{4\tau}\mathbf{W}_2^2(\eta,\rho) + \mathcal{E}(\rho),$$

where we assume an upper bound of the step sizes τ_*, i.e.,

$$\tau_* < 1/(12d_1 + 8d_2) \tag{5.2.3}$$

In the sequel, the Assumptions on the external potential V and on the interaction kernel W reads as follows:

Assumption 5.2.1. The confinement potential V and the interaction kernel W satisfy:

(F1) $V \in \mathcal{C}^1(\Omega)$, $W \in \mathcal{C}^1(\mathbb{R}^\mathrm{d})$, and W is symmetric.

(F2) There exists some non-negative constant d_1 such that

$$|V(x)|\,,\ |W(x)|\,,\ \|\nabla V(x)\|\,,\ \|\nabla W(x)\| \le d_1\big(1 + \|x\|^2\big).$$

Note, with these definition at hand, the free energy functional $\mathcal{E}$ satisfies the usual *LSCC*-conditions from [4].

Later, in section 5.2.3 we will need further Assumptions on the approximation $(\rho_0^\tau, \rho_1^\tau)$ of the initial datum ρ_0.

Assumption 5.2.2. There are non-negative constants d_3, d_4 such that for all $\tau \in (0,\tau_*)$:

(I1) $\mathbf{W}_2^2(\rho_0^\tau,\rho_1^\tau) \le d_3\tau$ and $\mathbf{W}_2^2(\rho_0^\tau,\rho_0) \le d_3\tau$.

(I2) $\mathcal{U}_m(\rho_0^\tau) \le d_4$, $\mathcal{U}_m(\rho_1^\tau) \le d_4$, and $\|\rho_1^\tau\|^2_{\mathrm{BV}(\Omega)} \le d_4/\tau$.

5.2.2 Basic Properties of the BDF2 Penalization Ψ

Before we prove the solvability of problem (5.3.2) we establish two basic properties of the BDF2 penalization $\Psi(\tau,\eta,\nu;\,\cdot\,)$: Boundedness from below and lower semi-continuous with respect to narrow convergence.

Recall that Assumptions (F1)&(F2) are supposed to hold.

Lemma 5.2.3 (Lower Bound)**.** *There exist a non-negative constant d_2 such that the BDF2 penalization Ψ satisfies for each $\tau > 0$ and for all $\rho,\eta,\nu \in \mathcal{P}_2(\Omega)$:*

$$\Psi(\tau,\eta,\nu;\rho) \geq \Big(\frac{1}{8\tau} - \frac{3}{2}d_1 - d_2\Big)\boldsymbol{M}_2(\rho) - \frac{1}{\tau}\boldsymbol{M}_2(\nu) - \frac{3}{4\tau}\boldsymbol{M}_2(\eta) - d_2 - \frac{3}{2}d_1 \tag{5.2.4}$$

where d_2 is the constant from the Carleman estimate, cf. [54].

Remark 5.2.4. The upper bound for τ_* is chosen in such a way, that $\rho \mapsto \Psi(\tau,\eta,\nu;\rho)$ is bounded from below by a constant.

Proof. Without loss of generality we can assume ρ is an absolutely continuous measure with density ρ. Observe that $\mathcal{H}$ is not bounded from below by a constant on $\mathcal{P}_2(\Omega)$. However, we derive from the Carleman estimate a lower bound of $\mathcal{H}$ in terms of the second moment $\boldsymbol{M}_2$, cf. [54], i.e., there exist non-negative constants $d_2 \geq 0$ and $\gamma \in (\frac{d}{d+2}, 1)$ such that

$$\mathcal{U}_m(\rho) \geq \mathcal{H}(\rho) \geq -d_2(1+\boldsymbol{M}_2(\rho))^\gamma \geq -d_2(1+\boldsymbol{M}_2(\rho)).$$

By (F2) the external potential V and the interaction kernel W grow at most quadratically at infinity, the corresponding energies can be estimated from below in terms of the second moment $\boldsymbol{M}_2$ by

$$\begin{aligned}\mathcal{V}(\rho) + \mathcal{W}(\rho) \geq& - d_1 \int_\Omega (1 + \|x\|^2)\rho(x)\,\mathrm{d}x - \frac{1}{2}d_1 \int_{\Omega^2} (1 + \|x-y\|^2)\rho(x)\rho(y)\,\mathrm{d}x\,\mathrm{d}y \\ =& -\frac{3}{2}d_1(1+\boldsymbol{M}_2(\rho)).\end{aligned}$$

From the elementary inequality $\|x\|^2 - 2\|y\|^2 \leq 2\|x-y\|^2 \leq 3\|x\|^2 + 6\|y\|^2$ and from the definition of $\mathbf{W}_2$ it follows immediately

$$\boldsymbol{M}_2(\rho) - 2\boldsymbol{M}_2(\nu) \leq 2\mathbf{W}_2^2(\rho,\nu) \leq 3\boldsymbol{M}_2(\rho) + 6\boldsymbol{M}_2(\nu) \qquad \text{for all } \rho,\nu \in \mathcal{P}_2(\Omega). \tag{5.2.5}$$

Combining all three inequalities, we can deduce the following lower bound:

$$\begin{aligned}\Psi(\tau,\eta,\nu;\rho) \geq& \frac{1}{2\tau}\boldsymbol{M}_2(\rho) - \frac{1}{\tau}\boldsymbol{M}_2(\nu) - \frac{3}{8\tau}\boldsymbol{M}_2(\rho) - \frac{3}{4\tau}\boldsymbol{M}_2(\eta) \\ & - d_2(1+\boldsymbol{M}_2(\rho)) - \frac{3}{2}d_1(1+\boldsymbol{M}_2(\rho)),\end{aligned}$$

which is equivalent to the desired inequality (5.2.4). □

Lemma 5.2.5 (Lower Semi-continuity)**.** *For each $\tau > 0$ and for all $\eta, \nu \in \mathcal{P}_2(\Omega)$ the BDF2 penalization $\Psi(\tau, \eta, \nu; \cdot)$ is lower semi-continuous with respect to narrow convergence.*

Proof. Due to the lower semi-continuity with respect to narrow convergence of the internal energy $\mathcal{U}_m$, the external potential $\mathcal{V}$, and the interaction energy $\mathcal{W}$, the free energy functional $\mathcal{E}$ is also lower semi-continuous with respect to narrow convergence as a sum of lower semi-continuous functions.

Thus it remains to prove the lower semi-continuity of the auxiliary functional $\mathcal{A} : \mathcal{P}_2(\Omega) \to \mathbb{R}$, defined via

$$\mathcal{A}(\rho) := 4\mathbf{W}_2^2(\nu, \rho) - \mathbf{W}_2^2(\eta, \rho).$$

First, we simplify the auxiliary functional $\mathcal{A}$. Let $\boldsymbol{p}^1 \in \Gamma(\rho, \nu)$ and $\boldsymbol{p}^2 \in \Gamma(\rho, \eta)$ be two optimal transport plans. Further, introduce the special three-plan $\boldsymbol{p} \in \Gamma(\rho, \nu, \eta) := \{\boldsymbol{p} \in \mathscr{P}(\Omega \times \Omega \times \Omega) : (\pi_1)_\# \boldsymbol{p} = \rho,\ (\pi_2)_\# \boldsymbol{p} = \nu,\ (\pi_3)_\# \boldsymbol{p} = \eta\}$ such that $\boldsymbol{p}$ has marginal with respect to the x- and y-components equals to $\boldsymbol{p}^1$ and the marginal with respect to the x- and z-components is equal to $\boldsymbol{p}^2$, i.e., $(\pi_1, \pi_2)_\# \boldsymbol{p} = \boldsymbol{p}^1$ and $(\pi_1, \pi_3)_\# \boldsymbol{p} = \boldsymbol{p}^2$. The existence of such a three-plan is guaranteed by the gluing lemma, see [4, Lemma 5.3.2]. Then, we can rewrite the auxiliary functional $\mathcal{A}$ as

$$\begin{aligned} \mathcal{A}(\rho) &= \int_{\Omega^2} 4\left\|x - y\right\|^2 \mathrm{d}\boldsymbol{p}^1(x, y) - \int_{\Omega^2} \left\|x - z\right\|^2 \mathrm{d}\boldsymbol{p}^2(x, z) \\ &= \int_{\Omega^3} 4\left\|x - y\right\|^2 - \left\|x - z\right\|^2 \mathrm{d}\boldsymbol{p}(x, y, z). \end{aligned} \tag{5.2.6}$$

Now, let $(\rho_n)_{n\in\mathbb{N}}$ be a narrowly converging sequence with limit $\rho_* \in \mathcal{P}_2(\Omega)$. Since $(\rho_n)_{n\in\mathbb{N}}$ is narrowly converging to ρ_*, the sequences $(\boldsymbol{p}_n^1)_{n\in\mathbb{N}}$ and $(\boldsymbol{p}_n^2)_{n\in\mathbb{N}}$ are relatively compact in $\mathscr{P}_2(\Omega^2)$ with respect to narrow convergence and any limit point is an optimal transport plan, see [4, Proposition 7.1.3]. Thus we can extract a non-relabelled subsequence such that $(\boldsymbol{p}_n^1)_{n\in\mathbb{N}}$ and $(\boldsymbol{p}_n^2)_{n\in\mathbb{N}}$ converge narrowly to an optimal transport plan $\boldsymbol{p}_*^1 \in \Gamma(\rho_*, \nu)$ and to an optimal transport plan $\boldsymbol{p}_*^2 \in \Gamma(\rho_*, \eta)$, respectively. By the same argument, the sequence $(\boldsymbol{p}_n)_{n\in\mathbb{N}}$ of three-plans is relatively compact in $\mathscr{P}_2(\Omega^3)$ with respect to narrow convergence. Therefore we can extract a further non-relabelled subsequence such that $(\boldsymbol{p}_n)_{n\in\mathbb{N}}$ narrowly converges to some three-plan $\boldsymbol{p}_* \in \Gamma(\rho_*, \nu, \eta)$. Taking marginals is continuous with respect to narrow convergence, so we have $(\pi_1, \pi_2)_\# \boldsymbol{p}_* = \boldsymbol{p}_*^1$ and $(\pi_1, \pi_3)_\# \boldsymbol{p}_* = \boldsymbol{p}_*^2$, i.e., this limit three-plan $\boldsymbol{p}_*$ is admissible in (5.2.6).

Next, we want to apply the lower semi-continuity result [4, Lemma 5.1.7] to the alternative representation of $\mathcal{A}$. The uniform integrability of the negative part of the integrand in (5.2.6) with respect to $(\boldsymbol{p}_n)_{n\in\mathbb{N}}$ in the sense of [4] follows by the elementary inequality

$$4\left\|x - y\right\|^2 - \left\|x - z\right\|^2 \geq \frac{1}{2}\left\|x\right\|^2 - 4\left\|y\right\|^2 - 3\left\|z\right\|^2 \geq -4\big(\left\|y\right\|^2 + \left\|z\right\|^2\big).$$

Thus the lower bound on $4\left\|x - y\right\|^2 - \left\|x - z\right\|^2$ is independent of x. Since the second moments of ν and η are finite that difference is uniform integrable with respect to the family $(\boldsymbol{p}_n)_{n\in\mathbb{N}}$. Hence, we can invoke [4, Lemma 5.1.7] to conclude

$$\int_{\Omega^3} 4\,\|x-y\|^2 - \|x-z\|^2 \,\mathrm{d}\boldsymbol{p}_*(x,y,z) \le \liminf_{n\to\infty} \int_{\Omega^3} 4\,\|x-y\|^2 - \|x-z\|^2 \,\mathrm{d}\boldsymbol{p}_n(x,y,z).$$

Therefore the auxiliary function $\rho \mapsto \mathcal{A}(\rho) = 4\mathbf{W}_2^2(\nu,\rho) - \mathbf{W}_2^2(\eta,\rho)$ is lower semi-continuous with respect to narrow convergence. □

Recall that the well-posedness of a single step of the BDF2 scheme is equivalent to the existence of a minimizer in (5.3.2). The augmented energy functional Ψ shares no uniform semi-convexity as in the case of [70], so we cannot exploit the convexity to ensure the existence of a minimizer. Nevertheless, a standard technique from the calculus of variations yields the existence of a minimizer.

Theorem 5.2.6 (Existence of a Minimizer)**.** *For each $\tau \in (0,\tau_*)$ and for all $\eta,\nu \in \mathcal{P}_2(\Omega)$, there exists an absolutely continuous minimizer $\rho_* \in \mathcal{D}(\mathcal{E})$ of the map $\rho \mapsto \Psi(\tau,\eta,\nu;\rho)$.*

Proof. Take a minimizing sequence $(\rho_n)_{n\in\mathbb{N}}$ for the BDF2 penalization $\rho \mapsto \Psi(\tau,\eta,\nu;\rho)$. To extract a convergent subsequence, we use the auxiliary inequality (5.2.4). Since $\tau < \tau_*$, the pre-factor of the second moment $\boldsymbol{M}_2(\rho)$ in (5.2.4) is positive. Hence, the second moment $(\boldsymbol{M}_2(\rho_n))_{n\in\mathbb{N}}$ of the minimizing sequence $(\rho_n)_{n\in\mathbb{N}}$ is bounded. Also the internal energy $\mathcal{U}_m(\rho_n)$ of the minimizing sequence is bounded, since

$$\begin{aligned}\mathcal{U}_m(\rho_n) \le& \Psi(\tau,\eta,\nu;\rho_n) + \frac{1}{4\tau}\mathbf{W}_2^2(\eta,\rho_n) - \mathcal{V}(\rho_n) - \mathcal{W}(\rho_n)\\ \le& \sup_{n\in\mathbb{N}} \left[\Psi(\tau,\eta,\nu;\rho_n) + C(1+\boldsymbol{M}_2(\rho_n))\right] < \infty.\end{aligned}$$

Due to the super-linear growth of $\rho \mapsto \rho\log(\rho)$ and of $\rho \mapsto \rho^m$, we can apply the Dunford-Pettis Theorem to the densities $(\rho_n)_{n\in\mathbb{N}}$ and we can extract a non-relabelled subsequence $(\rho_n)_{n\in\mathbb{N}}$ converging weakly in $\mathrm{L}^1(\Omega)$. Since $C^b(\Omega) \subset \mathrm{L}^\infty(\Omega) \cong (\mathrm{L}^1(\Omega))^*$, in this case we can deduce from the weak convergence in $\mathrm{L}^1(\Omega)$ of the sequence of densities the narrow convergence of the corresponding measures. Summarized, the sequence $(\rho_n)_{n\in\mathbb{N}}$ also converges narrowly to an absolutely continuous measure $\rho_* \in \mathcal{P}_2(\Omega)$ with density ρ_*. By the lower semi-continuity of the $\mathrm{L}^m(\Omega)$-norm with respect to narrow convergence it follows $\rho_* \in \mathcal{D}(\mathcal{E})$.

To prove that ρ_* is indeed a minimizer we use the lower semi-continuity of the BDF2 penalization Ψ, proven in Lemma 5.2.5, to conclude

$$\Psi(\tau,\eta,\nu;\rho_*) \le \liminf_{n\to\infty} \Psi(\tau,\eta,\nu;\rho_n) = \inf_{\rho\in\mathcal{P}_2(\Omega)} \Psi(\tau,\eta,\nu;\rho).$$

Indeed, the limit measure with density ρ_* is a minimizer of the BDF2 penalization $\Psi(\tau,\eta,\nu;\,\cdot\,)$. □

5.2.3 Intrinsic Properties of the BDF2 Scheme

Given an equidistant partition $\boldsymbol{\tau} = (\tau, 2\tau, 3\tau, \ldots)$ of fixed time step size $\tau \in (0, \tau_*)$ and a pair of initial data $(\rho_0^{\boldsymbol{\tau}}, \rho_1^{\boldsymbol{\tau}})$ which approximates the initial datum ρ_0. Then, the *discrete solution* $(\rho_k^{\boldsymbol{\tau}})_{k\in\mathbb{N}}$ for $\mathcal{E}$ on $(\mathcal{P}_2(\Omega), \mathbf{W}_2)$ defined in (5.0.3) and equivalently defined by the recursive formula

$$\rho_{k+1}^{\boldsymbol{\tau}} \in \underset{\rho\in\mathcal{P}_2(\Omega)}{\operatorname{argmin}}\ \Psi(\tau, \rho_{k-1}^{\boldsymbol{\tau}}, \rho_k^{\boldsymbol{\tau}}; \rho) \quad \text{for } k \in \mathbb{N} \tag{5.2.7}$$

is well-posed by theorem 5.2.6. The rest of this section is devoted to derive structural properties of the BDF2 scheme, namely: *Step size independent estimates*, *discrete Euler-Lagrange equations*, *better a priori estimates.*

Step Size Independent Estimates. Next, we deduce the almost energy diminishing property and the step size independent bounds. We want to emphasize that these estimates are intrinsic properties of the scheme, which do not rely on any uniform semi-convexity of the augmented energy functional Ψ. The original proof of those estimates can be found in [70] and for the sake of the completeness, we recall a proof adapted to the L^2-Wasserstein formalism.

The first result is an auxiliary inequality which will be used to derive the step size independent bounds. Despite the auxiliary character of this inequality, we want to emphasize that this property is of interest by itself, since we can give a precise estimate of the energy decay of the BDF2 scheme in every step.

Lemma 5.2.7 (Almost Energy Diminishing)**.** *For each time step size $\tau \in (0, \tau_*)$ the discrete solution $(\rho_k^{\boldsymbol{\tau}})_{k\in\mathbb{N}}$ satisfies*

$$\mathcal{E}(\rho_k^{\boldsymbol{\tau}}) + \frac{1}{2\tau}\mathbf{W}_2^2(\rho_{k-1}^{\boldsymbol{\tau}}, \rho_k^{\boldsymbol{\tau}}) \leq \mathcal{E}(\rho_{k-1}^{\boldsymbol{\tau}}) + \frac{1}{4\tau}\mathbf{W}_2^2(\rho_{k-2}^{\boldsymbol{\tau}}, \rho_{k-1}^{\boldsymbol{\tau}}) \tag{5.2.8}$$

at each step $k = 2, 3, \ldots$.

Proof. Since $\rho_k^{\boldsymbol{\tau}}$ is a minimizer of $\rho \mapsto \Psi(\tau, \rho_{k-2}^{\boldsymbol{\tau}}, \rho_{k-1}^{\boldsymbol{\tau}}; \rho)$, it satisfies

$$\Psi(\tau, \rho_{k-2}^{\boldsymbol{\tau}}, \rho_{k-1}^{\boldsymbol{\tau}}; \rho_k^{\boldsymbol{\tau}}) \leq \Psi(\tau, \rho_{k-2}^{\boldsymbol{\tau}}, \rho_{k-1}^{\boldsymbol{\tau}}; \rho)$$

for all $\rho \in \mathcal{P}_2(\Omega)$. For the specific choice $\rho = \rho_{k-1}^{\boldsymbol{\tau}}$, we obtain

$$\frac{1}{\tau}\mathbf{W}_2^2(\rho_{k-1}^{\boldsymbol{\tau}}, \rho_k^{\boldsymbol{\tau}}) - \frac{1}{4\tau}\mathbf{W}_2^2(\rho_{k-2}^{\boldsymbol{\tau}}, \rho_k^{\boldsymbol{\tau}}) + \mathcal{E}(\rho_k^{\boldsymbol{\tau}}) \leq -\frac{1}{4\tau}\mathbf{W}_2^2(\rho_{k-2}^{\boldsymbol{\tau}}, \rho_{k-1}^{\boldsymbol{\tau}}) + \mathcal{E}(\rho_{k-1}^{\boldsymbol{\tau}}). \tag{5.2.9}$$

By the triangle inequality and the binomial formula,

$$\mathbf{W}_2^2(\rho_{k-2}^{\boldsymbol{\tau}}, \rho_k^{\boldsymbol{\tau}}) \leq 2\mathbf{W}_2^2(\rho_{k-2}^{\boldsymbol{\tau}}, \rho_{k-1}^{\boldsymbol{\tau}}) + 2\mathbf{W}_2^2(\rho_{k-1}^{\boldsymbol{\tau}}, \rho_k^{\boldsymbol{\tau}}).$$

Substitute this in the left-hand side of (5.2.9). This yields (5.2.8) □

The second result is the BDF2-equivalent of the classical estimates in the (time-dependent) Minimizing Movement scheme. For this reason, we have to assume that the approximation $(\rho_0^\tau, \rho_1^\tau)$ of the initial datum ρ_0 satisfy (I1)&(I2) from assumption 5.2.2.

Theorem 5.2.8 (Classical Estimates)**.** *Fix a time horizon $T > 0$. There exists a constant C, depending only on d_1 to d_2 and T, such that the corresponding discrete solutions $(\rho_k^\tau)_{k\in\mathbb{N}}$ satisfy*

$$\sum_{k=1}^{N} \frac{1}{2\tau}\mathbf{W}_2^2(\rho_{k\text{-}1}^\tau, \rho_k^\tau) \leq C, \qquad |\mathcal{U}_m(\rho_N^\tau)| \leq C, \qquad \boldsymbol{M}_2(\rho_N^\tau) \leq C, \tag{5.2.10}$$

for all $\tau \in (0, \tau_*)$ and for all $N \in \mathbb{N}$ with $N\tau \leq T$.

Proof. Sum up inequalities (5.2.8) for $k = 2$ to $K = N$ to obtain after cancellation:

$$\mathcal{E}(\rho_N^\tau) + \frac{1}{4\tau}\sum_{k=2}^{N}\mathbf{W}_2^2(\rho_{k\text{-}1}^\tau, \rho_k^\tau) \leq \mathcal{E}(\rho_1^\tau) + \frac{1}{4\tau}\mathbf{W}_2^2(\rho_1^\tau, \rho_2^\tau). \tag{5.2.11}$$

Next, we want to prove the auxiliary inequality

$$\boldsymbol{M}_2^2(\rho_k^\tau) - \boldsymbol{M}_2^2(\rho_{k\text{-}1}^\tau) \leq 2\mathbf{W}_2(\rho_{k\text{-}1}^\tau, \rho_k^\tau)\boldsymbol{M}_2(\rho_k^\tau). \tag{5.2.12}$$

Without loss of generality we assume $\boldsymbol{M}_2(\rho_{k\text{-}1}^\tau) \geq \boldsymbol{M}_2(\rho_k^\tau)$, otherwise the equality is always true. We use the binomial formula to obtain

$$\begin{aligned}\boldsymbol{M}_2^2(\rho_k^\tau) - \boldsymbol{M}_2^2(\rho_{k\text{-}1}^\tau) &= (\boldsymbol{M}_2(\rho_k^\tau) + \boldsymbol{M}_2(\rho_{k\text{-}1}^\tau))(\boldsymbol{M}_2(\rho_k^\tau) - \boldsymbol{M}_2(\rho_{k\text{-}1}^\tau)) \\ &\leq 2\boldsymbol{M}_2(\rho_k^\tau)(\boldsymbol{M}_2(\rho_k^\tau) - \boldsymbol{M}_2(\rho_{k\text{-}1}^\tau)).\end{aligned}$$

Let δ_0 be the Dirac measure localized at $x = 0$, then by the triangle inequality

$$\boldsymbol{M}_2(\rho) = \mathbf{W}_2(\rho, \delta_0) \leq \mathbf{W}_2(\rho, \nu) + \mathbf{W}_2(\nu, \delta_0) = \mathbf{W}_2(\rho, \nu) + \boldsymbol{M}_2(\nu).$$

This yields (5.2.12). Rewrite the difference of the second moments of ρ_N^τ and ρ_1^τ by means of a telescopic sum and use (5.2.12) to obtain

$$\boldsymbol{M}_2^2(\rho_N^\tau) - \boldsymbol{M}_2^2(\rho_1^\tau) = \sum_{k=2}^{N}\left[\boldsymbol{M}_2^2(\rho_k^\tau) - \boldsymbol{M}_2^2(\rho_{k\text{-}1}^\tau)\right] \leq 2\sum_{k=2}^{N}\mathbf{W}_2(\rho_{k\text{-}1}^\tau, \rho_k^\tau)\boldsymbol{M}_2(\rho_k^\tau).$$

A Cauchy type inequality with (5.2.12) yields

$$\boldsymbol{M}_2^2(\rho_N^\tau) - \boldsymbol{M}_2^2(\rho_1^\tau) \leq \frac{\tau_*}{4}\sum_{k=2}^{N}\frac{\mathbf{W}_2^2(\rho_{k\text{-}1}^\tau, \rho_k^\tau)}{\tau} + \frac{4\tau}{\tau_*}\sum_{k=2}^{N}\boldsymbol{M}_2^2(\rho_k^\tau).$$

Substitute (5.2.11) into this inequality:

$$\boldsymbol{M}_2^2(\rho_N^{\boldsymbol{\tau}}) \leq \boldsymbol{M}_2^2(\rho_1^{\boldsymbol{\tau}}) + \tau_* \Big(\mathcal{E}(\rho_1^{\boldsymbol{\tau}}) + \frac{1}{4\tau} \mathbf{W}_2^2(\rho_1^{\boldsymbol{\tau}}, \rho_2^{\boldsymbol{\tau}}) - \mathcal{E}(\rho_N^{\boldsymbol{\tau}}) \Big) + \frac{4\tau}{\tau_*} \sum_{k=2}^{N} \boldsymbol{M}_2^2(\rho_k^{\boldsymbol{\tau}}).$$

The first term of the right-hand side is estimated by

$$\boldsymbol{M}_2(\rho_1^{\boldsymbol{\tau}}) \leq 2\mathbf{W}_2(\rho_1^{\boldsymbol{\tau}}, \rho_0) + \boldsymbol{M}_2(\rho_0) \leq 2d_3\sqrt{\tau} + 2\boldsymbol{M}_2(\rho_0). \tag{5.2.13}$$

Next, $\mathcal{E}(\rho_1^{\boldsymbol{\tau}})$ is estimated using (I2) and estimate (5.2.13):

$$\mathcal{E}(\rho_1^{\boldsymbol{\tau}}) \leq d_4 + \frac{3}{2} d_1 (1 + \boldsymbol{M}_2(\rho_1^{\boldsymbol{\tau}})) \leq d_4 + \frac{3}{2} d_1 (1 + 2d_3\sqrt{\tau} + 2\boldsymbol{M}_2(\rho_0)). \tag{5.2.14}$$

A lower bound of the energy $\mathcal{E}$ evaluated at $\rho_N^{\boldsymbol{\tau}}$ is derived by the same way as in the prove of Lemma 5.2.3, i.e., there exist constants d_2 and $\gamma \in (\frac{d}{d+1}, 1)$ such that

$$\mathcal{E}(\rho_N^{\boldsymbol{\tau}}) \geq -d_2(1 + \boldsymbol{M}_2(\rho_N^{\boldsymbol{\tau}}))^{\gamma} - \frac{3}{2} d_1 (1 + \boldsymbol{M}_2(\rho_N^{\boldsymbol{\tau}})) \geq -(d_2 + \frac{3}{2} d_1)(2 + \boldsymbol{M}_2^2(\rho_N^{\boldsymbol{\tau}})).$$

Hence, there is a universal constant C, not depending on the step size τ, such that

$$\boldsymbol{M}_2^2(\rho_N^{\boldsymbol{\tau}}) \leq C + \tau_*(d_2 + \frac{3}{2} d_1) \boldsymbol{M}_2^2(\rho_N^{\boldsymbol{\tau}}) + \frac{4\tau}{\tau_*} \sum_{k=2}^{N} \boldsymbol{M}_2^2(\rho_k^{\boldsymbol{\tau}}).$$

We rearrange terms and use the upper bound for τ_* to arrive at the time-discrete Gronwall inequality

$$\boldsymbol{M}_2^2(\rho_N^{\boldsymbol{\tau}}) \leq 2C + \frac{8\tau}{\tau_*} \sum_{k=2}^{N} \boldsymbol{M}_2^2(\rho_k^{\boldsymbol{\tau}}).$$

By induction on N we obtain

$$\boldsymbol{M}_2^2(\rho_N^{\boldsymbol{\tau}}) \leq C \Big(1 + \frac{8\tau}{\tau_*} \Big)^N \leq \widehat{C} \exp\Big(\frac{8N\tau}{\tau_*} \Big) \leq \widehat{C} \exp\Big(\frac{8T}{\tau_*} \Big).$$

So the second moments $\boldsymbol{M}_2(\rho_N^{\boldsymbol{\tau}})$ of the discrete solution are uniformly bounded independent of the step size $\tau \in (0, \tau_*)$ and for all $N \in \mathbb{N}$ with $N\tau < T$.

The remaining estimates can be derived from this. An upper bound for the energy $\mathcal{E}(\rho_N^{\boldsymbol{\tau}})$ follows from (5.2.11) and (5.2.14) combined with (I2). The lower bound on $\mathcal{E}(\rho_N^{\boldsymbol{\tau}})$ follows by the lower bounds on $\mathcal{U}_m, \mathcal{V}$, and $\mathcal{W}$ in terms of the second moment. Hence, the boundedness of $\mathcal{E}(\rho_N^{\boldsymbol{\tau}}), \mathcal{V}(\rho_N^{\boldsymbol{\tau}})$, and $\mathcal{W}(\rho_N^{\boldsymbol{\tau}})$ yields the boundedness of $\mathcal{U}_m(\rho_N^{\boldsymbol{\tau}})$. The upper bound for the kinetic energy follows from the lower bound for the energy $\mathcal{E}(\rho_N^{\boldsymbol{\tau}})$, (5.2.11), and (5.2.14) combined with (I2). □

Discrete Euler-Lagrange Equations. In theorem 5.2.9, we derive the discrete Euler-Lagrange equations for the weak formulation of the non-linear Fokker-Planck equation (5.2.1). The key idea is the JKO-method introduced in [54] and recalled in section 2.4.

Theorem 5.2.9 (Discrete Euler-Lagrange Equations)**.** *The discrete solution* $(\rho_k^\tau)_{k\in\mathbb{N}}$ *obtained by the BDF2 method satisfies for each* $k \in \mathbb{N}\backslash\{1\}$ *and for all vector fields* $\xi \in C_c^\infty(\overline{\Omega}, \mathbb{R}^{\mathrm{d}})$ *with* $\xi \cdot \boldsymbol{n} = 0$ *on* $\partial\Omega$

$$\begin{aligned} 0 = &\int_\Omega - \operatorname{div}(\xi)\,(\rho_k^\tau)^m + \langle \xi, \nabla V\rangle\, \rho_k^\tau + \langle \xi, \nabla W * \rho_k^\tau\rangle\, \rho_k^\tau \,\mathrm{d}x \\ &+ \frac{2}{\tau}\int_{\Omega^2} \langle \xi(x), x-y\rangle \,\mathrm{d}\boldsymbol{p}_k^\tau(x,y) - \frac{1}{2\tau}\int_{\Omega^2} \langle \xi(x), x-z\rangle\,\mathrm{d}\boldsymbol{q}_k^\tau(x,z), \end{aligned} \tag{5.2.15}$$

where $\boldsymbol{p}_k^\tau \in \Gamma(\rho_k^\tau, \rho_{k\text{-}1}^\tau)$ *and* $\boldsymbol{q}_\tau^k \in \Gamma(\rho_k^\tau, \rho_{k\text{-}2}^\tau)$ *are optimal transport plans.*

Proof. Fix $\rho_k^\tau, \rho_{k\text{-}1}^\tau, \rho_{k\text{-}2}^\tau$ and $\xi \in C_c^\infty(\overline{\Omega}, \mathbb{R}^{\mathrm{d}})$ with $\xi \cdot \boldsymbol{n} = 0$ on $\partial\Omega$. We consider the perturbation ρ^s of ρ_k^τ as the solution of the Transport equation with velocity field ξ starting at ρ_k^τ, i.e., ρ^s is the solution of (2.4.6) as in section 2.4. The first variation of the energy $\mathcal{E}$ along the solution to the continuity equation amounts to

$$\frac{\mathrm{d}}{\mathrm{d}s}\left[\mathcal{E}(\rho^s)\right]_{s=0} = \int_\Omega - \operatorname{div}(\xi)\,(\rho_k^\tau)^m + \langle \xi, \nabla V\rangle\, \rho_k^\tau + \langle \xi, \nabla W * \rho_k^\tau\rangle\, \rho_k^\tau \,\mathrm{d}x.$$

The differentiability of the quadratic L^2-Wasserstein distance $\mathbf{W}_2$ along the solution ρ^s is given by [92, Theorem 8.13], since $\rho_{k\text{-}2}^\tau, \rho_{k\text{-}1}^\tau, \rho_k^\tau, \rho^s$ are all absolutely continuous measures. Hence, we can conclude:

$$\begin{aligned} &\frac{\mathrm{d}}{\mathrm{d}s}\left[4\mathbf{W}_2^2(\rho_{k\text{-}1}^\tau, \rho^s) - \mathbf{W}_2^2(\rho_{k\text{-}2}^\tau, \rho^s)\right]_{s=0} \\ &= 8\int_{\Omega^2} \langle \xi(x), x-y\rangle\,\mathrm{d}\boldsymbol{p}_k^\tau(x,y) - 2\int_{\Omega^2}\langle \xi(x), x-z\rangle\,\mathrm{d}\boldsymbol{q}_k^\tau(x,z), \end{aligned}$$

where $\boldsymbol{p}_k^\tau \in \Gamma(\rho_k^\tau, \rho_{k\text{-}1}^\tau)$ and $\boldsymbol{q}_\tau^k \in \Gamma(\rho_k^\tau, \rho_{k\text{-}2}^\tau)$ are optimal transport plans. Since ρ_k^τ is a minimizer of the BDF2 penalization $\Psi(\tau, \rho_{k\text{-}2}^\tau, \rho_{k\text{-}1}^\tau; \cdot)$ and since $s \mapsto \Psi(\tau, \rho_{k\text{-}2}^\tau, \rho_{k\text{-}1}^\tau; \rho^s)$ is differentiable at $s = 0$,

$$\begin{aligned} 0 =& \frac{\mathrm{d}}{\mathrm{d}s}\left[\Psi(\tau, \rho_{k\text{-}2}^\tau, \rho_{k\text{-}1}^\tau; \rho^s)\right]_{s=0} \\ =& \frac{1}{4\tau}\frac{\mathrm{d}}{\mathrm{d}s}\left[4\mathbf{W}_2^2(\rho_{k\text{-}1}^\tau, \rho^s) - \mathbf{W}_2^2(\rho_{k\text{-}2}^\tau, \rho^s)\right]_{s=0} + \frac{\mathrm{d}}{\mathrm{d}s}\left[\mathcal{E}(\rho^s)\right]_{s=0} \\ =& \frac{2}{\tau}\int_{\Omega^2}\langle \xi(x), x-y\rangle\,\mathrm{d}\boldsymbol{p}_k^\tau(x,y) - \frac{1}{2\tau}\int_{\Omega^2}\langle \xi(x), x-z\rangle\,\mathrm{d}\boldsymbol{q}_k^\tau(x,z) \\ &+ \int_\Omega - \operatorname{div}(\xi)\,(\rho_k^\tau)^m + \langle \xi, \nabla V\rangle\, \rho_k^\tau + \langle \xi, \nabla W * \rho_k^\tau\rangle\, \rho_k^\tau \,\mathrm{d}x. \end{aligned}$$

Indeed, we have the desired equality (5.2.15). □

Refined Regularity. The already obtained regularity results for the interpolated solution $\overline{\rho}_t^{\tau}$ are not sufficient to pass to the limit in the first term of the discrete Euler-Lagrange equation (5.2.15). Nevertheless, the following bounds in the $\mathrm{BV}(\Omega)$-norm of $(\rho_k^{\tau})^m$ are sufficient to obtain the desired regularity results. These estimates can be derived from the discrete Euler-Lagrange equation quite naturally.

Proposition 5.2.10 (Step Size Independent Local $\mathrm{BV}(\Omega)$-estimate)**.** *Fix a time horizon $T > 0$. There exists a constant C, depending only on d_1 to d_4 and T, such that the corresponding discrete solutions $(\rho_\tau^k)_{k\in\mathbb{N}}$ satisfy for all $\tau \in (0, \tau_*)$ and for all $k \in \mathbb{N}\backslash\{1\}$ with $k\tau \le T$:*

$$\|(\rho_k^{\tau})^m\|_{\mathrm{BV}(\Omega)} \le C\Big(1 + \frac{\mathbf{W}_2(\rho_{k\text{-}1}^{\tau}, \rho_k^{\tau})}{\tau} + \frac{\mathbf{W}_2(\rho_{k\text{-}2}^{\tau}, \rho_k^{\tau})}{\tau}\Big). \tag{5.2.16}$$

Proof. The $\mathrm{L}^1(\Omega)$-norm of $(\rho_k^{\tau})^m$ is equal to $(m-1)\mathcal{U}_m$ evaluated at ρ_k^{τ}. Hence, we can bound the first term in the definition of the $\mathrm{BV}(\Omega)$-norm uniformly by the classical estimates (5.2.10). In order to estimate the variation of $(\rho_k^{\tau})^m$, we estimate the term inside the supremum of the definition of $V((\rho_k^{\tau})^m, \Omega)$. Thus let $\xi \in C_c^{\infty}(\Omega, \mathbb{R}^{\mathrm{d}})$ with $\|\xi\|_{\infty} \le 1$, then we can estimate the integral term in the definition of the variation of $(\rho_k^{\tau})^m$ with the discrete Euler-Lagrange equations (5.2.15) as follows

$$\begin{aligned}\int_{\Omega} (\rho_k^{\tau})^m \ \mathrm{div}(\xi)\,\mathrm{d}x &= \int_{\Omega} \langle \xi(x), \nabla V\rangle \rho_k^{\tau}(x) + \langle \xi(x), \nabla W * \rho_k^{\tau}\rangle \rho_k^{\tau}(x)\,\mathrm{d}x \\ &\quad + \frac{2}{\tau}\int_{\Omega^2} \langle \xi(x), x-y\rangle \ \mathrm{d}\boldsymbol{p}_k^{\tau}(x,y) - \frac{1}{2\tau}\int_{\Omega^2} \langle \xi(x), x-z\rangle \ \mathrm{d}\boldsymbol{q}_k^{\tau}(x,z).\end{aligned} \tag{5.2.17}$$

By (I2) we have quadratic growth bounds for ∇V and ∇W, so using the step size independent bounds on the second moment (5.2.10), we can estimate the first terms in (5.2.17) as follows:

$$\int_{\Omega} \langle \xi(x), \nabla V\rangle \rho_k^{\tau}(x) + \langle \xi(x), \nabla W * \rho_k^{\tau}\rangle \rho_k^{\tau}(x)\,\mathrm{d}x \le 2d_1 \|\xi\|_{\infty} (1 + \boldsymbol{M}_2(\rho_k^{\tau})) \le 2d_1(1+C).$$

The second integral on the right-hand side of (5.2.17) can be estimated using Jensen's inequality

$$\Big|\int_{\Omega^2} \langle \xi(x), x-y\rangle \ \mathrm{d}\boldsymbol{p}_k^{\tau}(x,y)\Big| \le \|\xi\|_{\infty} \Big(\int_{\Omega^2} \|x-y\|^2 \ \mathrm{d}\boldsymbol{p}_k^{\tau}(x,y)\Big)^{1/2} \le \mathbf{W}_2(\rho_{k\text{-}1}^{\tau}, \rho_k^{\tau}),$$

and similar for the third integral of the right-hand side of (5.2.17). Hence, we have the following upper bound for the variation of $(\rho_k^{\tau})^m$:

$$V((\rho_k^{\tau})^m, \Omega) \le C\Big(1 + \frac{\mathbf{W}_2(\rho_{k\text{-}1}^{\tau}, \rho_k^{\tau})}{\tau} + \frac{\mathbf{W}_2(\rho_{k\text{-}2}^{\tau}, \rho_k^{\tau})}{\tau}\Big).$$

In conclusion, the discrete solution $(\rho_k^{\tau})_{k\in\mathbb{N}}$ satisfies the desired bound (5.2.16). □

As a sort of Corollary of the previous Proposition 5.2.10 we get the main ingredient of the convergence proof of the interpolated solution $\overline{\rho}_t^{\tau_n}$.

Theorem 5.2.11 (Step Size Independent Global $\mathrm{L}^2(0,T;\mathrm{BV}(\Omega))$-estimate)**.** *Fix a time horizon $T>0$. There exists a constant C, depending only on d_1 to d_4 and T, such that the corresponding interpolated solution $\overline{\rho}_t^{\tau}$ satisfies for each $\tau \in (0,\tau_*)$:*

$$\|(\overline{\rho}_t^{\tau})^m\|_{\mathrm{L}^2(0,T;\mathrm{BV}(\Omega))} \leq C. \tag{5.2.18}$$

Proof. We use the classical estimates on the kinetic energy (5.2.10) and the result from Proposition 5.2.10 to estimate the $\mathrm{L}^2(0,T;\mathrm{BV}(\Omega))$-norm of $(\overline{\rho}_t^{\tau})^m$. Let $N_T := \max\{N \in \mathbb{N} \mid N\tau \leq T\}$, then we have with (I2) from Assumption 5.2.2

$$\begin{aligned}
\|(\overline{\rho}_t^{\tau})^m\|^2_{\mathrm{L}^2(0,T;\mathrm{BV}(\Omega))} &\leq \tau\, \|(\rho_1^{\tau})^m\|^2_{\mathrm{BV}(\Omega)} + \sum_{k=2}^{N_T+1} \int_{(k-1)\tau}^{k\tau} \|(\rho_k^{\tau})^m\|^2_{\mathrm{BV}(\Omega)}\,\mathrm{d}t \\
&\leq d_4 + C \sum_{k=2}^{N_T+1} \tau \Big(1 + \frac{\mathbf{W}_2(\rho_{k\text{-}1}^{\tau}, \rho_k^{\tau})}{\tau} + \frac{\mathbf{W}_2(\rho_{k\text{-}2}^{\tau}, \rho_k^{\tau})}{\tau}\Big)^2.
\end{aligned}$$

By the triangle inequality $\mathbf{W}_2(\rho_k^{\tau}, \rho_{k\text{-}2}^{\tau}) \leq \mathbf{W}_2(\rho_{k\text{-}2}^{\tau}, \rho_{k\text{-}1}^{\tau}) + \mathbf{W}_2(\rho_{k\text{-}1}^{\tau}, \rho_k^{\tau})$ in combination with a Cauchy type inequality we obtain

$$\begin{aligned}
\|(\overline{\rho}_t^{\tau})^m\|^2_{\mathrm{L}^2(0,T;\mathrm{BV}(\Omega))} &\leq d_4 + C \sum_{k=2}^{N_T+1} \Big[\tau + \frac{\mathbf{W}_2^2(\rho_{k\text{-}1}^{\tau}, \rho_k^{\tau})}{\tau} + \frac{\mathbf{W}_2^2(\rho_{k\text{-}2}^{\tau}, \rho_{k\text{-}1}^{\tau})}{\tau}\Big] \\
&\leq d_4 + C(T+\tau) + C \sum_{k=1}^{N_T+1} \frac{\mathbf{W}_2^2(\rho_{k\text{-}1}^{\tau}, \rho_k^{\tau})}{\tau}.
\end{aligned}$$

Finally, we can conclude, under the step size independent bounds on the kinetic energy (5.2.10),

$$\|(\overline{\rho}_t^{\tau})^m\|^2_{\mathrm{L}^2(0,T;\mathrm{BV}(\Omega))} \leq d_4 + C(T+\tau) + C := \widetilde{C}$$

for some universal constant $\widetilde{C}$, which only depends on d_1 to d_4 and T, but not on the step size $\tau \in (0,\tau_*)$. Hence, we have proven the desired step-size independent $\mathrm{L}^2(0,T;\mathrm{BV}(\Omega))$-estimate (5.2.18). □

5.2.4 Convergence

In this section we prove our main theorem, the narrow and strong convergence of the approximation $\overline{\rho}_t^{\tau}$ to the solution ρ_t^* of the non-linear Fokker-Planck equation. Our first weak convergence result follows from the step size independent bounds (5.2.10) and the Arzelà-Ascoli theorem, which can be found in [4, Proposition 3.3.1].

Theorem 5.2.12 (Narrow Convergence in $\mathcal{P}_2(\Omega)$)**.** *Given a sequence of equidistant partitions $(\boldsymbol{\tau}_n)_{n\in\mathbb{N}}$ of vanishing step sizes $\tau_n \in (0,\tau_*)$. Then, there exists a (non-relabelled) subsequence $(\tau_n)_{n\in\mathbb{N}}$ and a limit curve $\rho_t^* \in \mathrm{AC}^2(0,\infty;(\mathcal{P}_2(\Omega),\mathbf{W}_2))$ such that for any $t \geq 0$:*

$$\overline{\rho}_t^{\boldsymbol{\tau}_n} \rightharpoonup \rho_t^* \qquad \text{narrowly in the space } \mathcal{P}_2(\Omega) \text{ as } n \to \infty.$$

Proof. Fix $T > 0$ and define the auxiliary function $A_t^n \in \mathrm{L}^2(0,T)$, also called discrete derivative, as

$$A_t^n := \frac{\mathbf{W}_2(\rho_{k\text{-}1}^{\boldsymbol{\tau}_n},\rho_k^{\boldsymbol{\tau}_n})}{\tau_n} \quad \text{for} \quad t \in ((k-1)\tau_n, k\tau_n] \quad \text{and} \quad k \in \mathbb{N}.$$

Using the step size independent bounds (5.2.10) we obtain for $N_T = \max\{N \mid N\tau_n \leq T\}$:

$$\int_0^T (A_t^n)^2 \,\mathrm{d}t \leq \sum_{k=1}^{N_T} \int_{(k-1)\tau_n}^{k\tau_n} \Big(\frac{\mathbf{W}_2(\rho_{k\text{-}1}^{\boldsymbol{\tau}_n},\rho_k^{\boldsymbol{\tau}_n})}{\tau_n}\Big)^2 \,\mathrm{d}t = \sum_{k=1}^{N_T} \frac{\mathbf{W}_2^2(\rho_{k\text{-}1}^{\boldsymbol{\tau}_n},\rho_k^{\boldsymbol{\tau}_n})}{\tau_n} \leq C.$$

Indeed, $A_t^n \in \mathrm{L}^2(0,T)$ and the $\mathrm{L}^2(0,T)$-norm of A_t^n is uniformly bounded independently of the step size τ_n. Therefore, the sequence A_t^n possesses a non-relabelled subsequence weakly convergent in $\mathrm{L}^2(0,T)$ with limit $A_t \in \mathrm{L}^2(0,T)$. To derive an uniform Hölder-estimate for $\overline{\rho}_t^{\boldsymbol{\tau}_n}$, choose $0 \leq s \leq t \leq T$ arbitrary and define $k_t = \max\{k \in \mathbb{N} \mid k\tau_n \leq t\}$, then

$$\mathbf{W}_2(\overline{\rho}_s^{\boldsymbol{\tau}_n},\overline{\rho}_t^{\boldsymbol{\tau}_n}) \leq \sum_{k=k_s+1}^{k_t} \mathbf{W}_2(\rho_{k\text{-}1}^{\boldsymbol{\tau}_n},\rho_k^{\boldsymbol{\tau}_n}) = \sum_{k=k_s+1}^{k_t} \int_{(k-1)\tau_n}^{k\tau_n} \frac{\mathbf{W}_2(\rho_{k\text{-}1}^{\boldsymbol{\tau}_n},\rho_k^{\boldsymbol{\tau}_n})}{\tau_n} \,\mathrm{d}t.$$

Rewriting this in terms of A_t^n yields the auxiliary inequality

$$\mathbf{W}_2(\overline{\rho}_s^{\boldsymbol{\tau}_n},\overline{\rho}_t^{\boldsymbol{\tau}_n}) \;\leq \int_{(s-\tau_n)^+}^t A_r^n \,\mathrm{d}r. \tag{5.2.19}$$

Taking the limit $n \to \infty$ yields, together with $A_t^n \rightharpoonup A_t$ in $\mathrm{L}^2(0,T)$,

$$\limsup_{n\to\infty} \mathbf{W}_2(\overline{\rho}_s^{\boldsymbol{\tau}_n},\overline{\rho}_t^{\boldsymbol{\tau}_n}) \leq \int_s^t A_r \,\mathrm{d}r.$$

Moreover, the second moments of the discrete solutions $(\rho_k^{\boldsymbol{\tau}_n})_{k\in\mathbb{N}}$ are uniformly bounded independently of the step size τ_n and therefore the interpolated solutions $\overline{\rho}_t^{\boldsymbol{\tau}_n}$ is uniformly contained in a set K which is compact with respect to narrow convergence. Hence, we can apply the Arzelà-Ascoli Theorem [4, Proposition 3.3.1] yielding the existence of a non-relabelled subsequence and a limit curve $\rho_t^* : [0,T] \to \mathcal{P}_2(\Omega)$ such that $\overline{\rho}_t^{\boldsymbol{\tau}_n}$ converges narrowly to ρ_t^* for each fixed $t \in [0,T]$. Additionally, the limit curve ρ_t^* is L^2-absolutely continuous with modulus of continuity $A_t \in \mathrm{L}^2(0,T)$. A further diagonal argument in $T \to \infty$ yields the narrow convergence of the interpolated solution $\overline{\rho}_t^{\boldsymbol{\tau}_n}$ to the limit curve ρ_t^* on for any $t \geq 0$ and $\rho_t^* \in \mathrm{AC}^2(0,\infty;(\mathcal{P}_2(\Omega),\mathbf{W}_2))$. □

Theorem 5.2.13 (Strong Convergence in $\mathrm{L}^p(0,T;\mathrm{L}^m(\Omega))$). *Under the same assumptions as in Theorem 5.2.12 and given the limit curve ρ_* therein, then there exists a further (non-relabelled) subsequence $(\boldsymbol{\tau}_n)_{n\in\mathbb{N}}$ such that for all $T>0$, for any $p\in[1,\infty)$ and for any bounded subset $\Theta\subseteq\Omega$:*

$$\overline{\rho}_t^{\boldsymbol{\tau}_n} \to \rho_t^* \qquad \textit{strongly in } \mathrm{L}^p(0,T;\mathrm{L}^m(\Theta)) \textit{ as } n\to\infty.$$

Proof of Theorem 5.2.13 for $\Omega\subsetneq\mathbb{R}^{\mathrm{d}}$. Fix $T>0$. In order to prove the strong convergence result we use the Aubin-Lions Theorem 2.5.4 with the underlying Banach space $\mathbf{X}=\mathrm{L}^m(\Omega)$. We consider the functional $\mathcal{A}:\mathrm{L}^m(\Omega)\to\mathbb{R}$, defined via

$$\mathcal{A}(\rho) := \begin{cases} \|\rho^m\|^2_{\mathrm{BV}(\Omega)} & \text{if } \rho\in\mathcal{P}_2(\Omega) \text{ and } \rho^m\in\mathrm{BV}(\Omega),\\ +\infty & \text{else.}\end{cases}$$

Using the lemma 2.5.2 in the introductory section 2.5 it follows that the functional $\mathcal{A}$ is measurable, lower semi-continuous with respect to the $\mathrm{L}^m(\Omega)$-topology, and has compact sublevels. Next, we choose as pseudo-distance $g=\mathbf{W}_2$ on $\mathrm{L}^m(\Omega)$. The L^2-Wasserstein distance is lower semi-continuous with respect to the $\mathrm{L}^m(\Omega)$-topology and clearly compatible with $\mathcal{A}$.

Next, we verify the assumption (2.5.1) on $(\overline{\rho}_t^{\boldsymbol{\tau}_n})_{n\in\mathbb{N}}$ of Theorem 2.5.4. By the refined $\mathrm{L}^2(0,T;\mathrm{BV}(\Omega))$-estimates of Theorem 5.2.11 it is clear, that the sequence $(\overline{\rho}_t^{\boldsymbol{\tau}_n})_{n\in\mathbb{N}}$ is tight with respect to $\mathcal{A}$, since we have:

$$\sup_{n\in\mathbb{N}}\int_0^T \|(\overline{\rho}_t^{\boldsymbol{\tau}_n})^m\|^2_{\mathrm{BV}(\Omega)}\,\mathrm{d}t = \sup_{n\in\mathbb{N}}\|(\overline{\rho}_t^{\boldsymbol{\tau}_n})^m\|^2_{\mathrm{L}^2(0,T;\mathrm{BV}(\Omega))} \le C<\infty.$$

For the proof of the relaxed averaged weak integral equicontinuity condition of $(\overline{\rho}_t^{\boldsymbol{\tau}_n})_{n\in\mathbb{N}}$ with respect to $\mathbf{W}_2$, we use the auxiliary function A_t^n and the estimate (5.2.19) from the proof of weak convergence results to obtain:

$$\int_0^{T-t}\mathbf{W}_2(\overline{\rho}_{s+t}^{\boldsymbol{\tau}_n},\overline{\rho}_s^{\boldsymbol{\tau}_n})\,\mathrm{d}s \le \int_0^{T-t}\int_{(s-\tau_n)^+}^{s+t} A_r^n\,\mathrm{d}r\,\mathrm{d}s \le (t+\tau_n)\int_0^T A_r^n\,\mathrm{d}r.$$

Indeed, using the weak L^2-convergence of A_t^n to some $A_t\in\mathrm{L}^2_{loc}(0,\infty)$ it follows

$$\begin{aligned}&\liminf_{h\searrow 0}\limsup_{n\to\infty}\frac{1}{h}\int_0^h\int_0^{T-t}\mathbf{W}_2(\overline{\rho}_{s+t}^{\boldsymbol{\tau}_n},\overline{\rho}_s^{\boldsymbol{\tau}_n})\,\mathrm{d}s\,\mathrm{d}t\\ \le&\liminf_{h\searrow 0}\limsup_{n\to\infty}\frac{1}{h}\int_0^h(t+\tau_n)\,\mathrm{d}t\int_0^T A_t^n\,\mathrm{d}t = 0.\end{aligned}$$

Therefore, we can conclude that there exists a non-relabeled subsequence $(\boldsymbol{\tau}_n)_{n\in\mathbb{N}}$ such that $\overline{\rho}_t^{\boldsymbol{\tau}_n}$ converges in $\mathcal{M}(0,T;\mathrm{L}^m(\Omega))$ to some curve ρ_t^+. Due to the uniform bounds in $\mathrm{L}^\infty(0,T;\mathrm{L}^m(\Omega))$, we obtain with remark 2.1.1 also convergence in $\mathrm{L}^p(0,T;\mathrm{L}^m(\Omega))$ as desired. Moreover, the limit curves ρ_t^+ and ρ_t^* have to coincide, since $\overline{\rho}_t^{\boldsymbol{\tau}_n}$ converges also in measure to ρ_t^+ and ρ_t^*, so both limits have to be equal. □

In the case of $\Omega = \mathbb{R}^{\mathrm{d}}$ we have to alter the proof given above, since the embedding of $\mathrm{BV}(\mathbb{R}^{\mathrm{d}})$ into $\mathrm{L}^1(\mathbb{R}^{\mathrm{d}})$ is not compact anymore. So we restrict ourself to the open and bounded sets $\Theta = \mathbb{B}_R(0)$. This subset is clearly open and bounded with Lipschitz-continuous boundary $\partial\Theta$, so the embedding of $\mathrm{BV}(\Theta)$ into $\mathrm{L}^1(\Theta)$ is compact again.

Proof of Theorem 5.2.13 for $\Omega = \mathbb{R}^{\mathrm{d}}$. Fix $T > 0$. Without loss of generality we can assume $\Theta = \mathbb{B}_R(0)$, since every bounded subset $K \subset \mathbb{R}^{\mathrm{d}}$ is contained in a ball with radius R and convergence in $\mathrm{L}^m(0,T;\mathrm{L}^m(\mathbb{B}_R(0)))$ implies convergence in $\mathrm{L}^m(0,T;\mathrm{L}^m(K))$.

As before, we want to use the Aubin-Lions Theorem 2.5.4 for the Banach space $\mathrm{L}^m(\Theta)$ equipped with the natural topology induced by the $\mathrm{L}^m(\Theta)$-norm applied to $(\overline{\rho}_t^{\boldsymbol{\tau}_n}|_\Theta)_{n\in\mathbb{N}}$, the restriction of the density $\overline{\rho}_t^{\boldsymbol{\tau}_n}$ to the subspace Θ. In this case we consider the functional $\widetilde{\mathcal{A}} : \mathrm{L}^m(\Theta) \to \mathbb{R}$, defined via

$$\widetilde{\mathcal{A}}(\rho) := \begin{cases} \|\rho^m\|^2_{\mathrm{BV}(\Theta)} & \text{if } \rho \in \mathcal{M}_f(\Theta) \text{ and } \rho^m \in \mathrm{BV}(\Theta), \\ +\infty & \text{else.} \end{cases}$$

Now, the functional $\widetilde{\mathcal{A}}$ is measurable, lower semi-continuous with respect to the $\mathrm{L}^m(\Theta)$ topology, and has compact sublevels. Since $\widetilde{\mathcal{A}}(\rho|_\Theta) \leq \mathcal{A}(\rho)$, we obtain by the same calculations as above the tightness of $(\overline{\rho}_t^{\boldsymbol{\tau}_n}|_\Theta)_{n\in\mathbb{N}}$ with respect to $\widetilde{\mathcal{A}}$.

Since the measure $\rho|_\Theta$ does not have unit mass anymore, we cannot consider the L^2-Wasserstein distance $\mathbf{W}_2$ as pseudo-distance anymore. However, we can use the following pseudo-distance $\widetilde{g}$:

$$\begin{aligned} \widetilde{g}(\rho,\nu) &:= \inf\left\{\mathbf{W}_2(\widetilde{\rho},\widetilde{\nu}) \mid \widetilde{\rho} \in \Sigma(\rho),\ \widetilde{\nu} \in \Sigma(\nu)\right\}, \\ \Sigma(\rho) &:= \left\{\widetilde{\rho} \in \mathscr{P}(\mathbb{R}^{\mathrm{d}}) \mid \widetilde{\rho}|_\Omega = \rho, \boldsymbol{M}_2(\widetilde{\rho}) \leq C\right\}, \end{aligned}$$

where C is the constant from the classical estimates (5.2.10) for the specific T. Since $\Sigma(\rho)$ and $\Sigma(\nu)$ are compact sets with respect to the narrow topology, the infimum is attained at some pair $\widetilde{\rho}_*, \widetilde{\nu}_*$. The pseudo-distance $\widetilde{g}$ is compatible with $\widetilde{\mathcal{A}}$, i.e., if $\rho^m, \nu^m \in BV(\Theta)$ and $\widetilde{g}(\rho,\nu) = 0$ then $\rho = \nu$ a.e. on Θ. The lower semi-continuity of the pseudo-distance $\widetilde{g}$ with respect to the $\mathrm{L}^m(\Theta)$-topology can be proven as follows. Choose to convergent sequences $\rho_n \to \rho$ and $\nu_n \to \nu$ in $\mathrm{L}^m(\Theta)$ with $\sup_n \widetilde{g}(\rho_n,\nu_n) < \infty$. By the remark from above, there exists $\widetilde{\rho}_n, \widetilde{\nu}_n$ such that $\widetilde{g}(\rho_n,\nu_n) = \mathbf{W}_2(\widetilde{\rho}_n,\widetilde{\nu}_n)$. Since the second moments are by definition of $\Sigma(\rho)$ uniformly bounded, we can extract a non-relabeled convergent subsequence which converges narrowly to $\widetilde{\rho} \in \Sigma(\rho), \widetilde{\nu} \in \Sigma(\nu)$. By the lower semi-continuity of W with respect to narrow convergence, we get in the end

$$\widetilde{g}(\rho,\nu) \leq \mathbf{W}_2(\widetilde{\rho},\widetilde{\nu}) \leq \liminf_{n\to\infty} \mathbf{W}_2(\widetilde{\rho}_n,\widetilde{\nu}_n) = \liminf_{n\to\infty} \mathbf{W}_2(\rho_n,\nu_n).$$

Therefore, the pseudo-distance $\widetilde{g}$ is lower semi-continuous with respect to the $\mathrm{L}^m(\Theta)$-topology. Thus, $\widetilde{g}$ satisfies the assumptions of theorem 2.5.4. Further, one has by definition $\widetilde{g}(\rho|_\Theta, \nu|_\Theta) \leq \mathbf{W}_2(\rho,\nu)$. Thus, we derive, using the same proof as above, the equi-continuity of $(\overline{\rho}_t^{\boldsymbol{\tau}_n}|_\Theta)_{n\in\mathbb{N}}$ with respect to the pseudo-distance $\widetilde{g}$.

Hence, we can conclude that there exists a non-relabeled subsequence of $\overline{\rho}_t^{\tau_n}|_\Theta$ which converges in $\mathcal{M}(0,T;\mathrm{L}^m(\Theta))$ to some limit ρ_t^+. As before, we use the uniform bounds in $\mathrm{L}^\infty(0,T;\mathrm{L}^m(\Theta))$, to obtain the strong convergence in $\mathrm{L}^p(0,T;\mathrm{L}^m(\Theta))$ by Remark 2.1.1. Moreover, the limit curves ρ_t^+ and $\rho_t^*|_\Theta$ have to coincide on Θ, since $\overline{\rho}_t^{\tau_n}|_\Theta$ converges also in measure on Θ to ρ_t^+ and $\rho_t^*|_\Theta$, so both limits have to be equal on Θ. Two diagonal arguments in $T \to \infty$ and $R \to \infty$ yield the desired convergence result. □

To complete the proof of the main theorem 5.0.2, we have to validate that ρ_t^* is indeed a solution to (5.2.1) in the sense of distributions.

Theorem 5.2.14 (Solution of the Non-linear Fokker-Planck Equation)**.** *Under the same assumptions as in Theorem 5.2.13 and given the limit curve ρ_t^* from there. The limit curve ρ_t^* is a solution to the non-linear Fokker-Planck equation with no-flux boundary condition* (5.2.1) *in the following weak sense: For each test function $\varphi_t \in C_c^\infty([0,\infty)\times\overline{\Omega})$ with $\nabla\varphi_t \cdot \boldsymbol{n} = 0$ on $\partial\Omega$ the limit function ρ_t^* satisfies:*

$$\begin{aligned}&\int_0^\infty \int_\Omega -\Delta\varphi_t\,(\rho_t^*)^m + \langle\nabla\varphi_t, \nabla V\rangle\,\rho_t^* + \langle\nabla\varphi_t, \nabla W * \rho_t^*\rangle\,\rho_t^*\,\mathrm{d}x\,\mathrm{d}t\\ &= \int_0^\infty \int_\Omega \partial_t\varphi_t\,\rho_t^*\,\mathrm{d}x\,\mathrm{d}t + \int_\Omega \varphi_0\,\rho_0\,\mathrm{d}x.\end{aligned} \tag{5.2.20}$$

Proof. For simplicity we drop the index n and write for the step size only τ and $\tau \to 0$. Fix $\varphi_t \in C_c^\infty([0,\infty)\times\overline{\Omega})$ with $\xi\cdot\boldsymbol{n} = 0$ on $\partial\Omega$ and let be $T > 0$ and $\Theta \subset \Omega$ be open and bounded such that supp $\varphi_t \subset [0,T]\times\Theta$. Further, define the piecewise constant interpolation $\overline{\varphi}_t^\tau$ of φ_t by

$$\overline{\varphi}_0^\tau = \varphi_0, \quad \overline{\varphi}_t^\tau = \varphi_{k\tau} \quad \text{for } t \in ((k-1)\tau, k\tau] \quad \text{and } k \in \mathbb{N}.$$

For each $k \in \mathbb{N}\backslash\{1\}$ insert the smooth function $x \mapsto \nabla\varphi_{(k-1)\tau}$ in the discrete Euler-Lagrange equation (5.2.15) for the vector field ξ. Summing the resulting equations from $k = 2$ to $N_T + 1$ and multiplying with τ yields:

$$\begin{aligned}0 &= \int_\tau^\infty \int_\Omega -\Delta\overline{\varphi}_t^\tau\,(\overline{\rho}_t^\tau)^m + \langle\nabla\overline{\varphi}_t^\tau, \nabla V\rangle\,\overline{\rho}_t^\tau + \langle\nabla\overline{\varphi}_t^\tau, \nabla W * \overline{\rho}_t^\tau\rangle\,\overline{\rho}_t^\tau\,\mathrm{d}x\,\mathrm{d}t\\ &\quad + \sum_{k=2}^{N_T}\Big[2\int_{\Omega^2}\langle\nabla\varphi_{(k-1)\tau}(x), x-y\rangle\,\mathrm{d}\boldsymbol{p}_k^\tau(x,y) - \frac{1}{2}\int_{\Omega^2}\langle\nabla\varphi_{(k-1)\tau}(x), x-z\rangle\,\mathrm{d}\boldsymbol{q}_k^\tau(x,z)\Big]\\ &=: I_1 + I_2.\end{aligned}$$

Due to the strong convergence in $\mathrm{L}^m(0,T;\mathrm{L}^m(\Theta))$ of $\overline{\rho}_t^\tau$ to ρ_t^* and due to the uniform convergence of $\Delta\overline{\varphi}_t^\tau$ to $\Delta\varphi_t$

$$\lim_{\tau\searrow 0} I_1 = \int_0^T \int_{\widetilde{\Omega}} -\Delta\varphi_t\,(\rho_t^*)^m + \langle\nabla\varphi_t, \nabla V\rangle\,\rho_t^* + \langle\nabla\varphi_t, \nabla W * \rho_t^*\rangle\,\rho_t^*\,\mathrm{d}x\,\mathrm{d}t.$$

To rewrite I_2, we use, as in [54], the second order Taylor expansion for a time independent function φ, to obtain

$$\begin{aligned}
&\Big| \int_\Omega \varphi(y)\rho^{\tau}_{k\text{-}1}(y)\,\mathrm{d}y - \int_\Omega \varphi(x)\rho^{\tau}_{k}(x)\,\mathrm{d}x - \int_{\Omega^2} \langle \nabla\varphi(x), y-x\rangle \;\mathrm{d}\boldsymbol{p}^{\tau}_k(x,y)\Big| \\
=&\Big| \int_{\Omega^2} \varphi(y) - \varphi(x) - \langle \nabla\varphi(x), y-x\rangle \;\mathrm{d}\boldsymbol{p}^{\tau}_k(x,y)\Big| \\
\le&\frac{1}{2}\,\|\mathrm{Hess}\,\varphi\|_\infty \int_{\Omega^2} \|x-y\|^2\,\mathrm{d}\boldsymbol{p}^{\tau}_k(x,y) \\
=&\frac{1}{2}\,\|\mathrm{Hess}\,\varphi\|_\infty\, \mathbf{W}_2^2(\rho^{\tau}_{k\text{-}1},\rho^{\tau}_k).
\end{aligned}$$

Replacing the time independent function φ with $\varphi_{(k-1)\tau}$ yields as approximation of I_2

$$I_2 = \sum_{k=2}^{N_T} \Big[\int_\Omega (\frac{3}{2}\rho^{\tau}_k - 2\rho^{\tau}_{k\text{-}1} + \frac{1}{2}\rho^{\tau}_{k\text{-}2})\varphi_{(k-1)\tau}\,\mathrm{d}x + \mathcal{O}(\mathbf{W}_2^2(\rho^{\tau}_{k\text{-}1},\rho^{\tau}_k)) + \mathcal{O}(\mathbf{W}_2^2(\rho^{\tau}_{k\text{-}2},\rho^{\tau}_k))\Big].$$

We rearrange the sum of the first term in I_2 as follows

$$\begin{aligned}
&\sum_{k=2}^{N_T} \int_\Omega (\frac{3}{2}\rho^{\tau}_k - 2\rho^{\tau}_{k\text{-}1} + \frac{1}{2}\rho^{\tau}_{k\text{-}2})\varphi_{(k-1)\tau}\,\mathrm{d}x \\
=&\sum_{k=0}^{N_T} \int_\Omega (\frac{3}{2}\varphi_{(k-1)\tau} - 2\varphi_{k\tau} + \frac{1}{2}\varphi_{(k+1)\tau})\,\rho^{\tau}_k\,\mathrm{d}x - \int_\Omega \frac{3}{2}\varphi_0\rho^{\tau}_1 + (\frac{3}{2}\varphi_{-\tau} - 2\varphi_0)\rho^{\tau}_0\,\mathrm{d}x,
\end{aligned}$$

where we use the convention $\varphi_t \equiv \varphi_0$ for all $t \le 0$. Finally, use the fundamental theorem of calculus and the classical estimate (5.2.10) to bound the second term in I_2, to obtain

$$I_2 = -\int_0^T \int_\Omega (\frac{3}{2}\partial_t\varphi_t - \frac{1}{2}\partial_t\varphi_{t+\tau})\,\overline{\rho}^{\tau}_t\,\mathrm{d}x\,\mathrm{d}t - \int_\Omega \frac{3}{2}\varphi_0\rho^{\tau}_1 + (\frac{3}{2}\varphi_{-\tau} - 2\varphi_0)\rho^{\tau}_0\,\mathrm{d}x + \mathcal{O}(\tau).$$

Indeed, combining the narrow convergence of $\overline{\rho}^{\tau}_t$ with the uniform convergence of $\partial_t\varphi_{t+\tau}$ to $\partial_t\varphi_t$ and with the narrow convergence of the initial data $(\rho^{\tau}_0, \rho^{\tau}_1)$ to ρ_0, the limit of I_2 is given by:

$$\lim_{\tau\searrow 0} I_2 = -\int_0^T \int_\Omega \partial_t\varphi_t\,\rho^*_t\,\mathrm{d}x\,\mathrm{d}t - \int_\Omega \varphi_0\,\rho_0\,\mathrm{d}x.$$

Finally, we can conclude that for arbitrary test functions $\varphi_t \in C_c^\infty([0,\infty)\times\overline{\Omega})$ with $\nabla\varphi_t\cdot\boldsymbol{n} = 0$ on $\partial\Omega$ the limit curve ρ^*_t satisfies:

$$\begin{aligned}
&\int_0^\infty \int_\Omega -\Delta\varphi_t(\rho^*_t)^m + \langle\nabla\varphi_t, \nabla V\rangle\,\rho^*_t + \langle\nabla\varphi_t, \nabla W * \rho^*_t\rangle\,\rho^*_t\,\mathrm{d}x\,\mathrm{d}t \\
=&\int_0^\infty \int_\Omega \partial_t\varphi_t\,\rho^*_t\,\mathrm{d}x\,\mathrm{d}t + \int_\Omega \varphi_0\,\rho_0\,\mathrm{d}x.
\end{aligned}$$

This yields that ρ_* is a solution to the non-linear Fokker-Planck equation (5.2.1) in the weak sense of (5.2.20). □

5.3 Application to Derrida-Lebowitz-Speer-Spohn Equation

This section is concerned with the variational formulation of the second order Backward Differentiation Formula applied to the Derrida-Lebowitz-Speer-Spohn (DLSS) equation with no-flux boundary condition:

$$\partial_t \rho_t = -\operatorname{div}\big(\rho_t \nabla\big(2\frac{\Delta\sqrt{\rho_t}}{\sqrt{\rho_t}}\big)\big) \tag{5.3.1}$$

starting from the initial configuration ρ_0 with non-flux boundary conditions in an open, bounded, and convex domain Ω with Lipschitz-continuous boundary $\partial\Omega$ and normal derivative $\boldsymbol{n}$ or $\Omega = \mathbb{R}^{\mathrm{d}}$. We consider (5.3.1) as an evolutionary equation in the L^2-Wasserstein space $(\mathcal{P}_2(\Omega), \mathbf{W}_2)$. The corresponding free energy functional $\mathcal{E} : \mathcal{P}_2(\Omega) \to \mathbb{R} \cup \{\infty\}$ for the DLSS equation is given by the Fisher information:

$$\mathcal{E}(\mu) := \int_\Omega \|\sqrt{\rho}\|^2 \, \mathrm{d}x$$

provided the measures μ is absolutely continuous with respect to the Lebesgue measure $\mathcal{L}^{\mathrm{d}}$ with $\mu = \rho\,\mathrm{d}\mathcal{L}^{\mathrm{d}}$ and $\sqrt{\rho} \in \mathrm{H}^1(\Omega)$, otherwise we set $\mathcal{E}(\mu) = \infty$. As in the previous section 5.2, the free energy functional $\mathcal{E}$ is (highly) not convex along generalized geodesics in $(\mathcal{P}_2(\Omega), \mathbf{W}_2)$ and hence, the theory developed in the first section 5.1 of this chapter is also not applicable.

Method. The variational formulation of the second order Backward Differentiation Formula applied to the DLSS equation (5.3.1) reads than as follows:

Scheme. For each equidistant partition $\boldsymbol{\tau} = (\tau, 2\tau, 3\tau, \ldots)$ with sufficiently small time step $\tau > 0$, let a pair of initial conditions $(\rho_0^{\boldsymbol{\tau}}, \rho_1^{\boldsymbol{\tau}})$ be given that approximate ρ_0. Then define inductively a discrete solution $(\rho_k^{\boldsymbol{\tau}})_{k\in\mathbb{N}}$ such that each $\rho_{k+1}^{\boldsymbol{\tau}}$ with $k \in \mathbb{N}$ is a minimizer of the following functional,

$$\rho \mapsto \Psi(\tau, \rho_{k-1}^{\boldsymbol{\tau}}, \rho_k^{\boldsymbol{\tau}}; \rho) := \frac{1}{\tau}\mathbf{W}_2^2(\rho_k^{\boldsymbol{\tau}}, \rho) - \frac{1}{4\tau}\mathbf{W}_2^2(\rho_{k-1}^{\boldsymbol{\tau}}, \rho) + \mathcal{E}(\rho).$$

Define the corresponding piecewise constant interpolation in time $\overline{\rho}_t^{\boldsymbol{\tau}} : [0, \infty) \to \mathcal{P}_2(\Omega)$ of the discrete solution $\rho_k^{\boldsymbol{\tau}}$ in time via

$$\overline{\rho}_0^{\boldsymbol{\tau}} = \rho_0^{\boldsymbol{\tau}}, \quad \overline{\rho}_t^{\boldsymbol{\tau}} = \rho_k^{\boldsymbol{\tau}} \quad \text{for } t \in ((k-1)\tau, k\tau] \quad \text{and } k \in \mathbb{N}.$$

Strategy of the Proof. We use the same approach as in the second section 5.2 of this chapter and apply it to this case. Accordingly, we derive in section 5.3.2 the existence of our approximation $(\rho_k^{\boldsymbol{\tau}})_{k\in\mathbb{N}}$ and prove the classical intrinsic properties of the discrete solution in section 5.3.3. Most notably to mention is the different approach to derive the better a priori bounds in section 5.3.3 by means of variations along solutions to the heat equation. In the end, these bounds are sufficient pass to the limit $\tau \to 0$ and prove the convergence of $\overline{\rho}_t^{\boldsymbol{\tau}}$ to a weak solution ρ_t^* of the DLSS equation, see section 5.3.4.

5.3.1 Setup and Assumptions

In this case the *BDF2 penalization* $\Psi : (0, \tau_*) \times (\mathcal{P}_2(\Omega))^3 \to \mathbb{R} \cup \{\infty\}$ with the energy energy functional $\mathcal{E}$ given by the Fisher information reads as

$$\Psi(\tau, \eta, \nu; \cdot) : \mathcal{P}_2(\Omega) \to \mathbb{R} \cup \{\infty\}; \; \Psi(\tau, \eta, \nu; \rho) := \frac{1}{\tau}\mathbf{W}_2^2(\nu, \rho) - \frac{1}{4\tau}\mathbf{W}_2^2(\eta, \rho) + \mathcal{E}(\rho),$$

where we assume an artificial upper bound of the step sizes $\tau_* < 1$. Note, the free energy functional $\mathcal{E}$ satisfies the usual *LSCC*-conditions from [4] with respect to the topology induced by the narrow convergence of measures.

Later, we will need further Assumptions on the approximation $(\rho_0^\tau, \rho_1^\tau)$ of the initial datum ρ_0.

Assumption 5.3.1. There are non-negative constants d_3, d_4 such that for all $\tau \in (0, \tau_*)$:

(I1) $\mathbf{W}_2^2(\rho_0^\tau, \rho_1^\tau) \leq d_3\tau$ and $\mathbf{W}_2^2(\rho_0^\tau, \rho_0) \leq d_3\tau$.

(I2) $\mathcal{I}(\rho_0^\tau) \leq d_4$, $\mathcal{I}(\rho_1^\tau) \leq d_4$, and $\|\rho_1^\tau\|^2_{\mathrm{H}^2(\Omega)} \leq d_4/\tau$.

5.3.2 Basic Properties of the BDF2 Penalization Ψ

Before we prove the well-posedness of the BDF2 scheme applied to the DLSS equation we establish two basic properties of the BDF2 penalization $\Psi(\tau, \eta, \nu; \cdot)$: Boundedness from below and lower semi-continuity with respect to narrow convergence.

Lemma 5.3.2. *For all $\rho, \eta, \nu \in \mathcal{P}_2(\Omega)$ it holds:*

$$\Psi(\tau, \eta, \nu; \rho) \geq \frac{1}{8\tau}\boldsymbol{M}_2(\rho) - \frac{1}{\tau}\boldsymbol{M}_2(\nu) - \frac{3}{4\tau}\boldsymbol{M}_2(\eta).$$

Proof. The proof is similar to the proof of lemma 5.2.3. As long as the Fisher information is non-negative one has:

$$\Psi(\tau, \eta, \nu; \rho) \geq \frac{1}{\tau}\mathbf{W}_2(\nu, \rho) - \frac{1}{4\tau}\mathbf{W}_2(\eta, \rho).$$

Next, to derive a lower bound for the terms comprising the L^2-Wasserstein distance, we use the elementary inequality (5.2.5) given by

$$\boldsymbol{M}_2(\rho) - 2\boldsymbol{M}_2(\nu) \leq 2\mathbf{W}_2^2(\rho, \nu) \leq 3\boldsymbol{M}_2(\rho) + 6\boldsymbol{M}_2(\nu) \qquad \text{for all } \rho, \nu \in \mathcal{P}_2(\Omega).$$

to obtain the lower estimate

$$\Psi(\tau, \eta, \nu; \rho) \geq \frac{1}{8\tau}\boldsymbol{M}_2(\rho) - \frac{1}{\tau}\boldsymbol{M}_2(\nu) - \frac{3}{4\tau}\boldsymbol{M}_2(\eta)$$

which is the desired result. □

Lemma 5.3.3. *For each $\tau > 0$ and for all $\eta, \nu \in \mathcal{P}_2(\Omega)$ the BDF2 penalization $\Psi(\tau, \eta, \nu; \cdot)$ is lower semi-continuous with respect to the narrow convergence.*

Proof. The Fisher information is lower semi-continuous with respect to the narrow convergence, see [47]. By the previous calculations in the proof of lemma 5.2.5 also the auxiliary map $\mathcal{A} : \mathcal{P}_2(\Omega) \to \mathbb{R}$, defined via

$$\mathcal{A}(\rho) := 4\mathbf{W}_2(\nu, \rho) - \mathbf{W}_2(\eta, \rho)$$

is lower semi-continuous with respect to the narrow convergence. Hence, the map $\rho \mapsto \Psi(\tau, \eta, \nu; \rho)$ is lower semi-continuous as sum of lower semi-continuous functions. □

Theorem 5.3.4. *For each $\tau > 0$ and for all $\eta, \nu \in \mathcal{P}_2(\Omega)$, there exists an absolutely continuous minimizer $\rho \in \mathcal{D}(\mathcal{E})$ of the map $\rho \mapsto \Psi(\tau, \eta, \nu; \rho)$.*

Proof. Take a minimizing sequence $(\rho_n)_{n\in\mathbb{N}}$ for the BDF2 penalization $\rho \mapsto \Psi(\tau, \eta, \nu; \rho)$. To extract a convergent subsequence, we use the lower bound of lemma 5.3.2. The pre-factor of the second moment $\boldsymbol{M}_2(\rho)$ in this inequality is positive and therefore $(\boldsymbol{M}_2(\rho_n))_{n\in\mathbb{N}}$ is bounded. Also, the Fisher information of the minimizing sequence is bounded, since

$$\mathcal{E}(\rho_n) \le \Psi(\tau, \eta, \nu; \rho_n) + \frac{1}{4\tau}\mathbf{W}_2(\eta, \rho_n) \le \sup_n \left[\Psi(\tau, \eta, \nu; \rho_n) + C(1 + \boldsymbol{M}_2(\rho_n))\right] < \infty.$$

Hence, the minimizing sequence $(\rho_n)_{n\in\mathbb{N}}$ is contained in some sublevel of the Fisher information $\mathcal{E}$, which is compact with respect to the narrow convergence. So, we can conclude there exists a limit density ρ_* such that $\rho_n \rightharpoonup^* \rho_*$ on a subsequence. Now since the BDF2 penalization is lower semi-continuous with respect to the narrow convergence, the limit density ρ_* is indeed a minimizer. □

5.3.3 Intrinsic Properties of the BDF2 scheme

Given an equidistant partition $\boldsymbol{\tau} = (\tau, 2\tau, 3\tau, \ldots)$ of fixed time step size $\tau > 0$ and a pair of initial data $(\rho_0^{\boldsymbol{\tau}}, \rho_1^{\boldsymbol{\tau}})$ which approximates the initial datum ρ_0. Then, the *discrete solution* $(\rho_k^{\boldsymbol{\tau}})_{k\in\mathbb{N}}$ for $\mathcal{E}$ on $(\mathcal{P}_2(\Omega), \mathbf{W}_2)$ defined in (5.0.3) and equivalently defined by the recursive formula

$$\rho_{k+1}^{\boldsymbol{\tau}} \in \operatorname*{argmin}_{\rho\in\mathcal{P}_2(\Omega)} \Psi(\tau, \rho_{k-1}^{\boldsymbol{\tau}}, \rho_k^{\boldsymbol{\tau}}; \rho) \quad \text{for } k \in \mathbb{N} \tag{5.3.2}$$

is well-posed by theorem 5.3.4. The rest of this section is devoted to derive structural properties of the BDF2 scheme, namely: *Step size independent estimates*, *discrete Euler-Lagrange equations*, *better a priori estimates*.

Step Size Independent Estimates. As in the previous section 5.2 the discrete solution $(\rho_k^\tau)_{k\in\mathbb{N}}$ satisfy the classical estimates on kinetic energy, free energy, and boundedness. Recall, the initial data $(\rho_0^\tau, \rho_1^\tau)$ satisfy (I1)&(I2) from Assumption 5.3.1.

Theorem 5.3.5 (Classical Estimates). *Fix a time horizon $T > 0$. There exists a constant C, depending only on d_1 to d_2 and T, such that the corresponding discrete solutions $(\rho_k^\tau)_{k\in\mathbb{N}}$ satisfy*

$$\sum_{k=1}^{N} \frac{1}{2\tau}\mathbf{W}_2^2(\rho_{k\text{-}1}^\tau, \rho_k^\tau) \le C, \qquad \mathcal{I}(\rho_N^\tau) \le C, \qquad \boldsymbol{M}_2(\rho_N^\tau) \le C, \tag{5.3.3}$$

for all $\tau \in (0, \tau_*)$ and for all $N \in \mathbb{N}$ with $N\tau \le T$.

This proof is almost identical if not simpler – since the Fisher information is non-negative – as the proofs given in the previous two sections, so we shall skip this proof.

Discrete Euler-Lagrange Equations. In the spirit of the JKO-method [54] we derive the discrete Euler-Lagrange equations for the weak formulation of the Derrida-Lebowitz-Speer-Spohn equation.

Theorem 5.3.6 (Discrete Euler-Lagrange Equations). *The discrete solution $(\rho_k^\tau)_{k\in\mathbb{N}}$ obtained by the BDF2 method satisfies for each $k \in \mathbb{N}\backslash\{1\}$ and for all vector fields $\xi \in C_c^\infty(\overline{\Omega}, \mathbb{R}^\mathrm{d})$ with $\xi \cdot \boldsymbol{n} =$ on $\partial\Omega$*

$$\begin{aligned} 0 = &- \int_\Omega \frac{1}{2}\langle \nabla(\operatorname{div}\xi), \nabla\rho_*\rangle + 2\langle \mathrm{D}\xi\nabla\sqrt{\rho_*}, \nabla\sqrt{\rho_*}\rangle \,\mathrm{d}x \\ &+ \frac{2}{\tau}\int_{\Omega^2} \langle \xi(x), x-y\rangle \,\mathrm{d}\boldsymbol{p}_k^\tau(x,y) - \frac{1}{2\tau}\int_{\Omega^2} \langle \xi(x), x-z\rangle \,\mathrm{d}\boldsymbol{q}_k^\tau(x,z), \end{aligned} \tag{5.3.4}$$

where $\boldsymbol{p}_k^\tau \in \Gamma(\rho_k^\tau, \rho_{k\text{-}1}^\tau)$ and $\boldsymbol{q}_\tau^k \in \Gamma(\rho_k^\tau, \rho_{k\text{-}2}^\tau)$ are optimal transport plans.

Proof. Fix $\rho_k^\tau, \rho_{k\text{-}1}^\tau, \rho_{k\text{-}2}^\tau$ and $\xi \in C_c^\infty(\overline{\Omega}, \mathbb{R}^\mathrm{d})$ with $\xi \cdot \boldsymbol{n} =$ on $\partial\Omega$. We consider the perturbation ρ^s of ρ_k^τ as the solution of the transport equation with velocity field ξ starting at ρ_k^τ, i.e., ρ^s is the solution of (2.4.6) as in section 2.4. The first variation of the Fisher information $\mathcal{E}$ along the solution to the transport equation is equal

$$\frac{\mathrm{d}}{\mathrm{d}s}\left[\mathcal{E}(\rho^s)\right]_{s=0} = -\int_\Omega \frac{1}{2}\langle \nabla(\operatorname{div}\xi), \nabla\rho_*\rangle + 2\langle \mathrm{D}\xi\nabla\sqrt{\rho_*}, \nabla\sqrt{\rho_*}\rangle \,\mathrm{d}x.$$

The differentiability of the quadratic L^2-Wasserstein distance $\mathbf{W}_2$ along the solution ρ^s is given by [92, Theorem 8.13], since $\rho_{k\text{-}2}^\tau, \rho_{k\text{-}1}^\tau, \rho_k^\tau, \rho^s$ are all absolutely continuous measures. Hence, we can conclude:

$$\begin{aligned} &\frac{\mathrm{d}}{\mathrm{d}s}\left[4\mathbf{W}_2^2(\rho_{k\text{-}1}^\tau, \rho^s) - \mathbf{W}_2^2(\rho_{k\text{-}2}^\tau, \rho^s)\right]_{s=0} \\ =&8\int_{\Omega^2} \langle \xi(x), x-y\rangle \,\mathrm{d}\boldsymbol{p}_k^\tau(x,y) - 2\int_{\Omega^2} \langle \xi(x), x-z\rangle \,\mathrm{d}\boldsymbol{q}_k^\tau(x,z), \end{aligned}$$

where $\boldsymbol{p}_k^\tau \in \Gamma(\rho_k^\tau, \rho_{k\text{-}1}^\tau)$ and $\boldsymbol{q}_\tau^k \in \Gamma(\rho_k^\tau, \rho_{k\text{-}2}^\tau)$ are optimal transport plans.

Since ρ_k^τ is a minimizer of the BDF2 penalization $\Psi(\tau, \rho_{k\text{-}2}^\tau, \rho_{k\text{-}1}^\tau; \cdot)$ and since the map $s \mapsto \Psi(\tau, \rho_{k\text{-}2}^\tau, \rho_{k\text{-}1}^\tau; \rho^s)$ is differentiable at $s = 0$,

$$\begin{aligned}
0 =& \frac{\mathrm{d}}{\mathrm{d}s}\left[\Psi(\tau, \rho_{k\text{-}2}^\tau, \rho_{k\text{-}1}^\tau; \rho^s)\right]_{s=0} \\
=& \frac{1}{4\tau}\frac{\mathrm{d}}{\mathrm{d}s}\left[4\mathbf{W}_2^2(\rho_{k\text{-}1}^\tau, \rho^s) - \mathbf{W}_2^2(\rho_{k\text{-}2}^\tau, \rho^s)\right]_{s=0} + \frac{\mathrm{d}}{\mathrm{d}s}\left[\mathcal{E}(\rho^s)\right]_{s=0} \\
=& \frac{2}{\tau}\int_{\Omega^2} \langle \xi(x), x - y\rangle \,\mathrm{d}\boldsymbol{p}_k^\tau(x,y) - \frac{1}{2\tau}\int_{\Omega^2} \langle \xi(x), x - z\rangle \,\mathrm{d}\boldsymbol{q}_k^\tau(x,z) \\
&- \int_\Omega \frac{1}{2}\langle \nabla(\operatorname{div}\xi), \nabla\rho_*\rangle + 2\langle \mathrm{D}\xi\nabla\sqrt{\rho_*}, \nabla\sqrt{\rho_*}\rangle \,\mathrm{d}x.
\end{aligned}$$

Indeed, we have the desired equality (5.3.4). □

Refined Regularity. The already obtained regularity results for the interpolated solution $\overline{\rho}_t^\tau$ are not sufficient to pass to the limit in the first integral of the discrete Euler-Lagrange equation (5.3.4). Unfortunately, we are not in the situation as in the Fokker-Planck case, where we could derive from the discrete Euler-Lagrange equations better a priori bounds. To circumvent this issue, we propose a different variation of the discrete solution ρ_k^τ, namely along the heat flow. More precisely, we define as perturbation ρ^s of ρ_k^τ as the solution to the heat equation

$$\partial_s\rho^s = \Delta\rho^s, \qquad \rho^0 = \rho_k^\tau. \tag{5.3.5}$$

Proposition 5.3.7 (Step Size Independent Local $\mathrm{H}^2(\Omega)$-estimates)**.** *Fix a time horizon $T > 0$. There exists a constant C, depending only on d_1, Ω, and T, such that the corresponding discrete solutions $(\rho_k^\tau)_{k\in\mathbb{N}}$ satisfy for all $\tau > 0$ and for all $k \in \mathbb{N}\backslash\{1\}$ with $k\tau \le T$:*

$$\|\sqrt{\rho_k^\tau}\|_{\mathrm{H}^2(\Omega)}^2 \le C\Big(1 + \frac{\mathbf{W}_2(\rho_{k\text{-}1}^\tau, \rho_k^\tau)}{\tau} + \frac{\mathbf{W}_2(\rho_{k\text{-}2}^\tau, \rho_k^\tau)}{\tau}\Big). \tag{5.3.6}$$

Proof. Fix $T > 0$. By definition, the $\mathrm{H}^2(\Omega)$-norm of $\sqrt{\rho_k^\tau}$ comprises three terms:

$$\begin{aligned}
\|\sqrt{\rho_k^\tau}\|_{\mathrm{H}^2(\Omega)}^2 &= \|\sqrt{\rho_k^\tau}\|_{\mathrm{L}^2(\Omega)}^2 + \|\nabla\sqrt{\rho_k^\tau}\|_{\mathrm{L}^2(\Omega;\mathbb{R}^d)}^2 + \|\mathrm{D}^2\sqrt{\rho_k^\tau}\|_{\mathrm{L}^2(\Omega;\mathbb{R}^{d\times d})}^2 \\
&= \|\rho_k^\tau\|_{\mathrm{L}^1(\Omega)}^2 + \mathcal{I}(\rho_k^\tau) + \|\mathrm{D}^2\sqrt{\rho_k^\tau}\|_{\mathrm{L}^2(\Omega;\mathbb{R}^{d\times d})}^2.
\end{aligned}$$

The first term is equal to the L^1-norm of ρ_k^τ and hence equal to one. The second term, i.e., the Fisher information of ρ_k^τ, is bounded by some constant independently of the step size τ by the classical estimates (5.3.3). Hence, it remains to estimate the norm of the second order derivate.

To do so, we use the idea of the flow interchange technique developed by Matthes et al. [68]: we define ρ^s as the perturbation of ρ_k^τ along the heat flow (5.3.5) with $\eta = 1$.

As long as ρ_k^τ is the minimizer of the BDF2 penalization one has

$$
\begin{aligned}
0 \leq & \frac{1}{s}[\Psi(\tau,\rho_{k\text{-}2}^\tau,\rho_{k\text{-}1}^\tau;\rho^s) - \Psi(\tau,\rho_{k\text{-}2}^\tau,\rho_{k\text{-}1}^\tau;\rho_k^\tau)] \\
= & \frac{1}{s\tau}[\mathbf{W}_2^2(\rho_{k\text{-}1}^\tau,\rho^s) - \mathbf{W}_2^2(\rho_{k\text{-}1}^\tau,\rho_k^\tau)] - \frac{1}{4s\tau}[\mathbf{W}_2^2(\rho_{k\text{-}2}^\tau,\rho^s) - \mathbf{W}_2^2(\rho_{k\text{-}2}^\tau,\rho_k^\tau)] \\
& + \frac{1}{s}[\mathcal{E}(\rho^s) - \mathcal{E}(\rho_k^\tau)]
\end{aligned}
$$

for sufficiently small $s > 0$. By passing to the limit $s \to 0$ in each term, we will obtain the better a priori bound for $\mathrm{D}^2\sqrt{\rho_k^\tau}$. The first variation of the Wasserstein distance along the heat Flow ρ^s is given by lemma 2.4.7 and is equal to

$$
\begin{aligned}
& \limsup_{s\to 0} \frac{1}{s\tau}[\mathbf{W}_2^2(\rho_{k\text{-}1}^\tau,\rho^s) - \mathbf{W}_2^2(\rho_{k\text{-}1}^\tau,\rho_k^\tau)] - \frac{1}{4s\tau}[\mathbf{W}_2^2(\rho_{k\text{-}2}^\tau,\rho^s) - \mathbf{W}_2^2(\rho_{k\text{-}2}^\tau,\rho_k^\tau)] \\
\leq & \frac{1}{2\tau}\mathbf{W}_2(\rho_{k\text{-}1}^\tau,\rho_k^\tau)\sqrt{\mathcal{I}(\rho_k^\tau)} + \frac{1}{8\tau}\mathbf{W}_2(\rho_{k\text{-}2}^\tau,\rho_k^\tau)\sqrt{\mathcal{I}(\rho_k^\tau)}.
\end{aligned}
$$

The first variation of the Fisher information along the heat flow ρ^s is given by lemma 2.4.9 and is equal to

$$
\liminf_{s\to 0} \frac{\mathrm{d}}{\mathrm{d}s}[\mathcal{E}(\rho^s)] \leq -C\int_\Omega \left\|\mathrm{D}^2\sqrt{\rho_k^\tau}\right\|^2 \,\mathrm{d}x.
$$

Putting everything together yields

$$
0 \leq \frac{1}{2\tau}\mathbf{W}_2(\rho_{k\text{-}1}^\tau,\rho_k^\tau)\sqrt{\mathcal{I}(\rho_k^\tau)} + \frac{1}{8\tau}\mathbf{W}_2(\rho_{k\text{-}2}^\tau,\rho_k^\tau)\sqrt{\mathcal{I}(\rho_k^\tau)} - C\int_\Omega \left\|\mathrm{D}^2\sqrt{\rho_k^\tau}\right\|^2 \,\mathrm{d}x
$$

which is – after rearranging the inequality and using the classical estimate (5.3.3) on the Fisher information – the desired local a priori estimate (5.3.6). □

Theorem 5.3.8 (Step Size Independent Global $\mathrm{H}^2(\Omega)$-estiamtes)**.** *Fix a time horizon $T > 0$. There exists a constant C, depending only on d_3, d_4, Ω and T, such that the corresponding interpolated solution $\overline{\rho}_t^\tau$ satisfies for each $\tau \in (0,\tau_*)$:*

$$
\left\|\sqrt{\overline{\rho}_t^\tau}\right\|_{\mathrm{L}^2(0,T;\mathrm{H}^2(\Omega))} \leq C. \tag{5.3.7}
$$

Proof. We use the results from Proposition 5.3.7 to estimate the $\mathrm{L}^2(0,T;\mathrm{H}^2(\Omega))$-norm of $\sqrt{\overline{\rho}_t^\tau}$. Let $N_T : \max\{N \in \mathbb{N} \mid N\tau \leq T\}$, then we have with (I2) from Assumption 5.3.1

$$
\begin{aligned}
\left\|\sqrt{\overline{\rho}_t^\tau}\right\|_{\mathrm{L}^2(0,T;\mathrm{H}^2(\Omega))} &\leq \tau\left\|\sqrt{\rho_1^\tau}\right\|_{\mathrm{H}^2(\Omega)}^2 + \sum_{k=2}^{N_T+1}\int_{(k-1)\tau}^{k\tau}\left\|\sqrt{\rho_k^\tau}\right\|_{\mathrm{H}^2(\Omega)}^2 \mathrm{d}t \\
&\leq d_4 + C\sum_{k=2}^{N_T+1}\tau\Big(1 + \frac{\mathbf{W}_2(\rho_{k\text{-}1}^\tau,\rho_k^\tau)}{\tau} + \frac{\mathbf{W}_2(\rho_{k\text{-}2}^\tau,\rho_k^\tau)}{\tau}\Big)^2.
\end{aligned}
$$

By the triangle inequality and a Cauchy type inequality we obtain

$$\begin{aligned}\|\sqrt{\rho_t^{\boldsymbol{\tau}}}\|_{\mathrm{L}^2(0,T;\mathrm{H}^2(\Omega))} &\leq d_4 + C \cdot \sum_{k=2}^{N_T+1} \left(\tau + \frac{\mathbf{W}_2^2(\rho_{k\text{-}1}^{\boldsymbol{\tau}}, \rho_k^{\boldsymbol{\tau}})}{\tau} + \frac{\mathbf{W}_2^2(\rho_{k\text{-}2}^{\boldsymbol{\tau}}, \rho_{k\text{-}1}^{\boldsymbol{\tau}})}{\tau}\right) \\ &\leq d_4 + CT + C \cdot \sum_{k=1}^{N_T+1} \frac{\mathbf{W}_2^2(\rho_{k\text{-}1}^{\boldsymbol{\tau}}, \rho_k^{\boldsymbol{\tau}})}{\tau}.\end{aligned}$$

Finally, we can conclude with the step size independent bounds on the kinetic energy and on the Fisher information from (5.3.3),

$$\|\sqrt{\rho_t^{\boldsymbol{\tau}}}\|_{\mathrm{L}^2(0,T;\mathrm{H}^2(\Omega))} \leq d_4 + CT + C \cdot C \cdot C =: \widetilde{C}$$

for some universal constant $\widetilde{C}$, which depends only on d_3, d_4, Ω and T, but not on the step size $\tau \in (0, \tau_*)$. Hence, we have proven the desired step size independent global $\mathrm{L}^2(0,T;\mathrm{H}^2(\Omega))$-estimate (5.3.7)

□

5.3.4 Convergence

In this section we prove our main theorem, the narrow and strong convergence of the approximation $\overline{\rho}_t^{\boldsymbol{\tau}}$ to the solution ρ_t^* of the Derrida-Lebowitz-Speer-Spohn Equation. Our first weak convergence result follows from the step size independent bounds (5.3.3) and the Arzelà-Ascoli theorem, which can be found in [4, Proposition 3.3.1].

Theorem 5.3.9 (Narrow Convergence in $\mathcal{P}_2(\Omega)$)**.** *Given a sequence of equidistant partitions $(\boldsymbol{\tau}_n)_{n\in\mathbb{N}}$ of vanishing step sizes $\tau_n \in (0, \tau_*)$. Then, there exists a (non-relabelled) subsequence $(\tau_n)_{n\in\mathbb{N}}$ and a limit curve $\rho_t^* \in \mathrm{AC}^2(0,\infty; (\mathcal{P}_2(\Omega), \mathbf{W}_2))$ such that for any $t \geq 0$:*

$$\overline{\rho}_t^{\boldsymbol{\tau}_n} \rightharpoonup \rho_t^* \qquad \textit{narrowly in the space } \mathcal{P}_2(\Omega) \textit{ as } n \to \infty.$$

The proof of this theorem is word by word identical to the proof of theorem 5.2.12 with the only difference, that we use the classical estimates (5.3.3), therefore, we skip this proof.

Next, we state the strong convergence results.

Theorem 5.3.10 (Strong Convergence)**.** *Under the same assumptions as in Theorem 5.3.9 and given the limit curve ρ_* therein, then there exists a further (non-relabelled) subsequence $(\boldsymbol{\tau}_n)_{n\in\mathbb{N}}$ such that for all $T > 0$, for any $p \in [1, \infty)$:*

$$\begin{aligned}\overline{\rho}_t^{\boldsymbol{\tau}_n} &\to \rho_t^* &&\textit{strongly in } \mathrm{L}^p(0,T;\mathrm{L}^1(\Omega)) \textit{ as } n \to \infty, \\ \sqrt{\overline{\rho}_t^{\boldsymbol{\tau}_n}} &\to \sqrt{\rho_t^*} &&\textit{strongly in } \mathrm{L}^2(0,T;\mathrm{H}^1(\Omega)) \textit{ as } n \to \infty, \\ \sqrt{\overline{\rho}_t^{\boldsymbol{\tau}_n}} &\rightharpoonup \sqrt{\rho_t^*} &&\textit{weakly in } \mathrm{L}^2(0,T;\mathrm{H}^2(\Omega)) \textit{ as } n \to \infty.\end{aligned}$$

Proof. Fix $T > 0$. To derive the convergence results we proceed similarly as in the proof of theorem 5.2.13, by applying the extension of Aubin-Lions theorem 2.5.4 once. The other results follow by the Banach-Alaoglu theorem and an interpolation argument.

We seek to apply theorem 2.5.4 for $(u_t^n)_{n\in\mathbb{N}} := (\sqrt{\overline{\rho}_t^{\tau_n}})_{n\in\mathbb{N}}$ with the underlying Banach space $\mathbf{X} = \mathrm{L}^2(\Omega)$. We consider as normal coercive functional $\mathcal{A} : \mathrm{L}^2(\Omega) \to [0,\infty]$ and as pseudo-distance g on $\mathrm{L}^2(\Omega)$:

$$\mathcal{A}(u) := \begin{cases} \|u\|_{\mathrm{H}^1(\Omega)} + \boldsymbol{M}_2(u^2) & \text{if } u \in \mathrm{H}^1(\Omega) \text{ and } u^2 \in \mathcal{P}_2^{\mathrm{ac}}(\Omega), \\ \infty & \text{else,} \end{cases}$$

$$g(f,h) := \begin{cases} \mathbf{W}_2(f^2,h^2) & \text{if } f,h \geq 0 \wedge f^2,h^2 \in \mathcal{P}_2^{\mathrm{ac}}(\Omega), \\ +\infty & \text{else.} \end{cases}$$

Note, with abuse of notation f^2 and h^2 are identified with the corresponding measures $f^2\mathcal{L}^{\mathrm{d}}$ and $h^2\mathcal{L}^{\mathrm{d}}$, respectively.

To prove that $\mathcal{A}$ is a normal coercive integrand, we shall prove all properties. It is clear, that $\mathcal{A}$ is measurable. The lower semi-continuity of the $\mathrm{H}^1(\Omega)$-norm and of the second moment $\boldsymbol{M}_2$ is also trivial. For Ω open, bounded and convex, the compactness of the sublevels of $\mathcal{A}$ follows from the Rellich-Kondrachov theorem. For $\Omega = \mathbb{R}^{\mathrm{d}}$, we use the following observation: given a sequence of measures $(\mu_n)_{n\in\mathbb{N}}$ with $\sup_n \boldsymbol{M}_2(\mu_n) < \infty$ then we can extract by Prokhorov's theorem a subsequence which converges in the weak*-topology of measures to some limit measure μ_*. Hence, every sequence $(u_n)_{n\in\mathbb{N}} \subset \mathrm{L}^2(\Omega)$ contained in a sublevel of $\mathcal{A}$ satisfies $\sup_n \boldsymbol{M}_2(\mu_n) < \infty$ and therefore $(u_n^2)_{n\in\mathbb{N}}$ converges to some limit measure μ_* in the weak*-topology of measures. Therefore, we can use Lemma 2.2 from [68] which yields the strong convergence of $(u_t^n)_{n\in\mathbb{N}}$ in the $\mathrm{L}^2(\Omega)$-topology. Additionally, the pseudo-distance g is indeed a lower semi-continuous pseudo-distance, cf. [66, Proposition 7.6].

Next, we have to verify, that $(u_t^n)_{n\in\mathbb{N}}$ satisfies the hypothesis of theorem 2.5.4. The tightness of $(u_t^n)_{n\in\mathbb{N}}$ with respect to $\mathcal{A}_t$ follows directly from the classical estimates (5.3.3). By the same calculations as in theorem 5.2.13, $(u_t^n)_{n\in\mathbb{N}}$ satisfies the weak integral equi-continuity condition, since $g(u_t^n, u_{t+h}^n) = \mathbf{W}_2(\overline{\rho}_t^{\tau_n}, \overline{\rho}_{t+h}^{\tau_n})$. We can conclude by theorem 2.5.4 that (on a subsequence) $(u_t^n)_{n\in\mathbb{N}}$ converges to some u_t^* in $\mathcal{M}(0,T;\mathrm{L}^2(\Omega))$. We have also uniform $\mathrm{L}^\infty(0,T;\mathrm{L}^2(\Omega))$ bounds. So we can use Remark 2.1.1 to conclude the strong convergence result of $(u_t^n)_{n\in\mathbb{N}}$ to some u_t^* in $\mathrm{L}^p(0,T;\mathrm{L}^2(\Omega))$.

It follows directly by the ε-independent $\mathrm{L}^2(0,T;\mathrm{H}^2(\Omega))$-estimates (5.3.7) that u_t^n converges weakly in the $\mathrm{L}^2(0,T;\mathrm{H}^2(\Omega))$-topology. Hence, the weak convergence of $(u_t^n)_{n\in\mathbb{N}}$ to u_t^* in $\mathrm{L}^2(0,T;\mathrm{H}^2(\Omega))$ and the strong convergence in $\mathrm{L}^2(0,T;\mathrm{L}^2(\Omega))$ immediately yields strong convergence in $\mathrm{L}^2(0,T;\mathrm{H}^1(\Omega))$ by an interpolation argument.

Lastly, we have to verify $\rho_t^* = (u_t^*)^2$ for almost every $x \in \Omega$. After possibly extracting further subsequences, we can ensure that $(\rho_t^{\varepsilon_n})_{n\in\mathbb{N}}$ and $(u_t^n)_{n\in\mathbb{N}}$ converge almost everywhere on $[0,T]\times\Omega$. Clearly, we have by continuity of the square $\rho_t^* = (u_t^*)^2$ for almost everywhere on $[0,T]\times\Omega$. A diagonal argument in T yields the final result. □

Lastly, we have to verify that the limit function ρ_t^* of theorems 5.3.9&5.3.10 are indeed a solution to the Derrida-Lebowitz-Speer-Spohn Equation.

Theorem 5.3.11 (Solution to the Derrida-Lebowitz-Speer-Spohn Equation)**.** *Under the same assumptions as in Theorem 5.3.10 and given the limit curve ρ_t^* from there. The limit curve ρ_t^* is a solution to the Derrida-Lebowitz-Speer-Spohn equation with no-flux boundary condition* (5.3.1) *in the following weak sense: For each test function $\varphi_t \in C_c^\infty([0,\infty)\times\overline{\Omega})$ with $\nabla\varphi_t\cdot\boldsymbol{n} =$ on $\partial\Omega$ the curve ρ_t^* satisfies:*

$$\begin{aligned} &-\int_0^\infty\int_\Omega \frac{1}{2}\langle\nabla(\Delta\varphi_t),\nabla\rho_t^*\rangle + 2\langle \operatorname{Hess}\varphi_t\nabla\sqrt{\rho_t^*},\nabla\sqrt{\rho_t^*}\rangle\,\mathrm{d}x\,\mathrm{d}t \\ &=\int_0^\infty\int_\Omega \rho_t^*\partial_t\varphi_t\,\mathrm{d}x\,\mathrm{d}t + \int_\Omega \varphi_0\,\rho_0\,\mathrm{d}x. \end{aligned} \tag{5.3.8}$$

Proof. For simplicity, we drop the index n and write for the step size only τ and $\tau\to 0$. Fix $\varphi_t\in C_c^\infty([0,\infty)\times\overline{\Omega})$ with $\nabla\varphi_t\cdot\boldsymbol{n}=0$ on $\partial\Omega$ and let $T>0$ be such that $\operatorname{supp}\varphi_t\subset[0,T]\times\overline{\Omega}$. Define as in the previous section the piecewise constant interpolation $\overline{\varphi}_t^{\boldsymbol{\tau}}$ of φ_t by

$$\overline{\varphi}_0^{\boldsymbol{\tau}} = \varphi_0,\quad \overline{\varphi}_t^{\boldsymbol{\tau}} = \varphi_{k\tau}\quad \text{for } t\in((k-1)\tau,k\tau]\quad\text{and } k\in\mathbb{N}.$$

For each $k\in\mathbb{N}\backslash\{1\}$ insert the smooth function $x\mapsto\nabla\varphi_{(k-1)\tau}$ in the discrete Euler-Lagrange equation (5.3.4) for the vector field ξ. Summing the resulting equations from $k=2$ to N_T+1 and multiplying with τ yields:

$$\begin{aligned} 0 = &-\int_\tau^\infty\int_\Omega \frac{1}{2}\langle\nabla(\Delta\overline{\varphi}_t^{\boldsymbol{\tau}}),\nabla\overline{\rho}_t^{\boldsymbol{\tau}}\rangle + 2\langle\operatorname{Hess}\overline{\varphi}_t^{\boldsymbol{\tau}}\nabla\sqrt{\overline{\rho}_t^{\boldsymbol{\tau}}},\nabla\sqrt{\overline{\rho}_t^{\boldsymbol{\tau}}}\rangle\,\mathrm{d}x \\ &+\sum_{k=2}^{N_T}\Big[2\int_{\Omega^2}\langle\nabla\varphi_{(k-1)\tau}(x),x-y\rangle\,\mathrm{d}\boldsymbol{p}_k^{\boldsymbol{\tau}}(x,y) - \frac{1}{2}\int_{\Omega^2}\langle\nabla\varphi_{(k-1)\tau}(x),x-z\rangle\,\mathrm{d}\boldsymbol{q}_k^{\boldsymbol{\tau}}(x,z)\Big] \\ &=: I_1 + I_2. \end{aligned}$$

Due to the strong convergence in $\mathrm{L}^2(0,T;\mathrm{H}^1(\Omega))$ of $\sqrt{\overline{\rho}_t^{\boldsymbol{\tau}}}$ to $\sqrt{\rho_t^*}$ and due to the uniform convergence of $\Delta\overline{\varphi}_t^{\boldsymbol{\tau}}$ to $\Delta\varphi_t$

$$\lim_{\tau\searrow 0} I_1 = -\int_0^\infty\int_\Omega\frac{1}{2}\langle\nabla(\Delta\varphi_t),\nabla\rho_t^*\rangle + 2\langle\operatorname{Hess}\varphi_t\nabla\sqrt{\rho_t^*},\nabla\sqrt{\rho_t^*}\rangle\,\mathrm{d}x.$$

The limit of I_2 is given by the same calculations as in the proof of theorem 5.2.14, i.e., by the classical estimates (5.3.3) and by the assumptions in the initial data $(\rho_0^{\boldsymbol{\tau}},\rho_1^{\boldsymbol{\tau}})$:

$$\lim_{\tau\searrow 0} I_2 = -\int_0^T\int_\Omega\partial_t\varphi_t\,\rho_t^*\,\mathrm{d}x\,\mathrm{d}t - \int_\Omega\varphi_0\,\rho_0\,\mathrm{d}x.$$

Finally, we can conclude that the limit curve ρ_t^* satisfies the Derrida-Lebowitz-Speer-Spohn equation (5.3.1) in the weak sense of (5.3.8). □

5.4 Illustration by Numerical Experiments

In this section, we illustrate the convergence of our variational BDF2 method in comparison to the implicit Euler scheme in several numerical experiments. As examples, we have chosen a flow on the two-dimensional sphere $\mathbb{S}^2$, a reaction-diffusion equation as flow on the Hilbert space $\mathrm{L}^2(0,1)$, and an aggregation-diffusion equation as a flow in the space $\mathcal{P}(\Omega)$ of probability measures on $\Omega = [-1,1]$, equipped with the L^2-Wasserstein distance $\mathbf{W}_2$. We observe that the order of convergence is indeed very close to two in each of our simulations. This underlines our philosophy that one reaches the optimal order in "typical" problems, despite the fact that our main Theorem 5.1.11 only provides order one-half, and that there are specific counter-examples with sub-optimal converge rates, as in the introduction of this chapter.

Method. In each of the examples below, we compare the numerical results for the implicit Euler scheme and for the BDF2 method at various moderately small time steps $\tau > 0$ to a reference solution that is obtained by the BDF2 method with a very small time step τ_{ref}. The approximation with the implicit Euler method of step size $\tau > 0$ — see chapter 3 for details — is denoted by $(\overline{u}_t^{\tau})^{(1)}$, and the approximation with BDF2 by $(\overline{u}_t^{\tau})^{(2)}$, respectively. For the time-discrete initial data, we choose the original datum u_0 for both schemes at $t = 0$, and for the second initial datum (at $t = \tau$) of the BDF2 method, we use the result of the first step of the implicit Euler scheme.

Remark 5.4.1. This choice ensures in the ODE setting the enhanced convergence rate of order two since the startup calculation with one step of the implicit Euler scheme is of order two, cf. [11, Theorem 7.23]. More precise, let u_1^{τ} be the first minimizer obtained by the implicit Euler scheme and let $u_t^* \in \mathcal{C}^2(0,T)$ be the solution of the gradient flow equation $\dot{u}_t^* = -\nabla\mathcal{E}(\rho_t^*)$. Then, use the definition of u_1^{τ}, Taylor's formula applied to u_t^*, and the Lipschitz-continuity of the map $x \mapsto \nabla\mathcal{E}(x)$ to get

$$\|u_\tau^* - u_1^\tau\| = \big\| - u_0 - \tau\nabla\mathcal{E}(u_\tau^*) + \frac{1}{2}\tau\ddot{u}_t^* + u_0 - \tau\nabla\mathcal{E}(u_1^\tau)\big\| \leq \tau L\big\|u_\tau^* - u_1^\tau\big\| + \frac{1}{2}\tau^2\big\|\ddot{u}_t^*\big\|_\infty.$$

Assume $\tau L < 1/2$, then a kick-back argument yields

$$\|u_\tau^* - u_1^\tau\| \leq \frac{\big\|\ddot{u}_t^*\big\|_\infty}{2(1-\tau L)}\tau^2 \leq C\tau^2.$$

The numerical rate of convergence is then computed as follows. In addition to the very small reference time step τ_{ref}, we choose a moderately large time step τ_{coarse} that is an integer multiple of τ_{ref}. Then, we calculate $(\overline{u}_t^{\tau})^{(1)}$ and $(\overline{u}_t^{\tau})^{(2)}$ for several intermediate time steps $\tau \in (\tau_{\mathrm{ref}}, \tau_{\mathrm{coarse}})$ that are chosen such that τ is an integer multiple of τ_{ref}, and τ_{coarse} is an integer multiple of τ. For each such choice of τ, the respective solutions $(\overline{u}_t^{\tau})^{(1)}$ and $(\overline{u}_t^{\tau})^{(2)}$ are compared to the reference solution $(\overline{u}_t^{\mathrm{ref}})^{(2)}$: specifically, we calculate a mean numerical error by taking the average of the distances $\boldsymbol{d}((\overline{u}_{t_k^\tau}^{\tau})^{(i)}, (u_{t_k^\tau}^{\mathrm{ref}})^{(2)})$ at times $t_k^\tau = k\tau_{\mathrm{coarse}} \in [0,T]$ on the coarsest grid.

All simulations have been performed with `MATLAB`. Both variational schemes are implemented by solving the sequence of variational problems using the built-in method `fmincon`.

Gradient Flow on the Sphere $\mathbb{S}^2$. The first test problem is placed on the unit 2-sphere $\mathbb{S}^2 := \{u \in \mathbb{R}^3 \,|\, u_1^2 + u_2^2 + u_3^2 = 1\} \subset \mathbb{R}^3$ equipped with the intrinsic (great-circle) distance $\boldsymbol{d}_{\mathbb{S}^2}$, defined by $\boldsymbol{d}_{\mathbb{S}^2}(u, v) = \arccos(u_1 v_1 + u_2 v_2 + u_3 v_3)$ for $u, v \in \mathbb{S}^2$. For the potential $\mathcal{E} : \mathbb{S}^2 \to \mathbb{R}$, we choose the restriction of

$$\widetilde{\mathcal{E}}(u) = \sum_{i=1}^{3} \left(u_i - \frac{1}{2}\right)\left(u_i + \frac{1}{2}\right)^2.$$

The corresponding gradient flow satisfies the ODE

$$\dot{u} = -\nabla_{\mathbb{S}^2}\mathcal{E}(u) = \Pi_u\big[-\nabla\widetilde{\mathcal{E}}(u)\big],$$

where $\Pi_u[v] = v - u^T v$ is the projection of a vector v to the tangent space of $\mathbb{S}^2$ at u. Its flow lines are sketched in Figure 5.1 (left). The example falls into the class of gradient flows on Riemannian manifolds that are covered by Theorem 5.1.5.

A series of simulations has been performed for the initial datum

$$u_0 = \frac{1}{\sqrt{30}}(1, 2, 5)$$

and the reference step size $\tau_{ref} = 10^{-5}$. The observed numerical convergence rates are 1.00 for the implicit Euler method, and 2.06 for BDF2, see Figure 5.1 (right). Further experiments with different initial data and other potentials yield very similar results. In this smooth, finite-dimensional setting, second-order convergence of the BDF2 method was naturally expected, as the solution curve and $\nabla_{\mathbb{S}^2}\mathcal{E}$ are smooth in the ambient space.

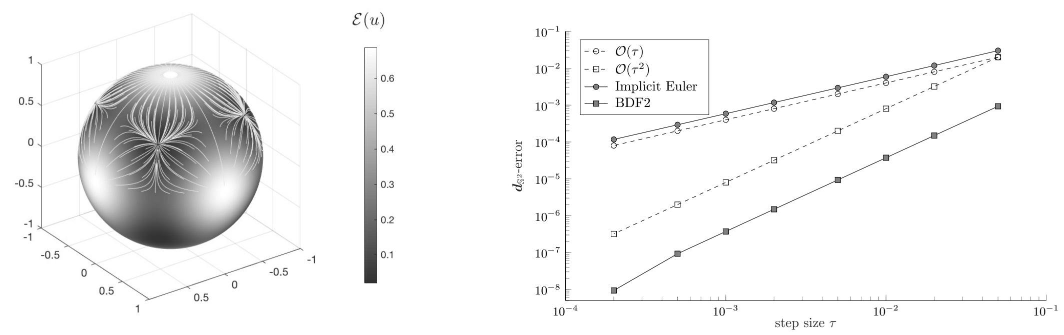


Figure 5.1: *Gradient flow on the Sphere* $\mathbb{S}^2$. Left: the values of $\mathcal{E}$ are color-coded by gray scale. The white lines are sample trajectories of the gradient flow generated by $\mathcal{E}$. Right: the $\boldsymbol{d}_{\mathbb{S}^2}$-error plot of $(\overline{u}_t^{\boldsymbol{\tau}})^{(i)}$ compared with $\overline{u}_t^{\boldsymbol{\tau}_{ref}}$.

Reaction-Diffusion Equation with Obstacle. Next, we consider the constrained reaction-diffusion equation

$$\partial_t u = \Delta u + 60u^3 \quad \text{subject to} \quad |u| \leq 1$$

on $\Omega = [0,1]$, subject to homogeneous Neumann boundary conditions. This PDE constitutes a gradient flow on the Hilbert space $\mathrm{L}^2(0,1)$ for the energy

$$\mathcal{E}(u) = \begin{cases} \frac{1}{2}\int_0^1 \big(\partial_x u(x)\big)^2 \,\mathrm{d}x - 15\int_0^1 u(x)^4 \,\mathrm{d}x, & \text{for } u \in \mathrm{H}^1(0,1),\ |u| \leq 1, \\ +\infty, & \text{otherwise.} \end{cases}$$

The second variation of $\mathcal{E}$ amounts to

$$\mathrm{d}^2\mathcal{E}(u)[\varphi]^2 = \int_0^1 \big(\partial_x \varphi(x)\big)^2 \,\mathrm{d}x - 180\int_0^1 u(x)^2\varphi(x)^2 \,\mathrm{d}x \geq -180\|\varphi\|^2_{\mathrm{L}^2(0,1)},$$

since $0 \leq u(x)^2 \leq 1$. Hence $\mathcal{E}$ is uniformly semi-convex of modulus $\lambda = -90$.

For the numerical approximation, we first perform the implicit Euler or BDF2 method for discretization in time, then we apply a spatial discretization of the PDE, using central finite differences. The qualitative behavior of the approximate solution for the initial condition

$$u_0(x) = \frac{1}{2}\sin(2\pi x) + \frac{1}{4}$$

has been plotted in Figure 5.2 (left). Notice that the upper barrier is hit after a short transient time. The reference step size is $\tau_{\mathrm{ref}} = 10^{-6}$. Since we are interested in the convergence rate of the temporal discretization for the PDE, we need to estimate the influence of the additional spatial discretization on the numerical error. For that reason, the experiment is carried out with different choices of the spatial resolution, using $K = 50, 100, 250, 500, 1000$ grid points.

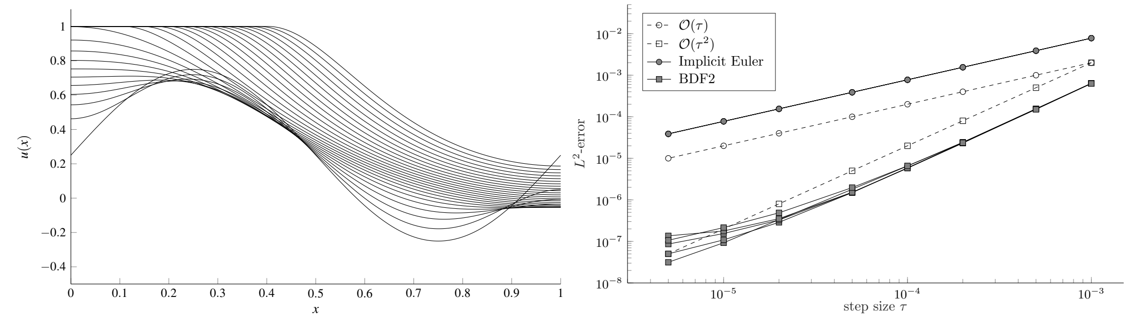


Figure 5.2: *Reaction-diffusion equation with obstacle.* Evolution of the reference solution $\overline{u}_{\tau_{ref}}$ (left). The L^2-error plot of $(\overline{u}_t^{\boldsymbol{\tau}})^{(i)}$ compared with $\overline{u}_t^{\boldsymbol{\tau}_{ref}}$ for different K (right).

Our results on the numerical error are given in Figure 5.2 (right). The error curve for the implicit Euler scheme is proportional to τ, as expected. For time steps $\tau > 10^{-5}$, the error curve for the BDF2 scheme is almost perfectly proportional to τ^2, and there is no significant dependence on the spatial discretization. For very small steps $\tau \leq 10^{-5}$, there is apparently an additional contribution to the numerical error due to the spatial discretization, however as K is increased, the error curve extends its approximate proportionality to τ^2 also into that regime. This is a strong indication that for a purely temporal discretization by BDF2, the order of convergence is indeed quadratic in τ. We performed further experiments with different initial data, and with variants of the energy functional. The results remain approximately the same.

Aggregation-Diffusion Equations. In our last example, we study discretizations of the following aggregation-diffusion equation,

$$\partial_t \rho_t = \Delta \rho_t + \partial_x(\rho_t W' * \rho_t) \tag{5.4.1}$$

on $\Omega = [-1, 1]$, with no-flux boundary condition, i.e., $\partial_x \rho_t + W' * \rho_t = 0$ at $x = \pm 1$. For the interaction kernel, we use $W(x) = 2x^4 - x^2$. Weak solutions to (5.4.1) conserve mass and positivity, so we restrict attention to solutions u that are probability densities. Under this restriction, solutions to (5.4.1) correspond to the gradient flow on the space $\mathbf{X} = \mathcal{P}([-1, 1])$ of probability measures μ with respect to the L^2-Wasserstein distance $\boldsymbol{d} = \mathbf{W}_2$ for the energy functional

$$\mathcal{E}(\mu) := \int_\Omega \rho \log(\rho)\,\mathrm{d}x + \frac{1}{2}\int_{\Omega\times\Omega} \rho(x) W(x-y)\rho(y)\,\mathrm{d}y\,\mathrm{d}x,$$

if $\mu \in \mathcal{P}_2^{\mathrm{ac}}(\Omega)$ with density $\rho \in \mathrm{L}^1(\Omega)$ and otherwise we set $\mathcal{E}(\mu) = +\infty$. For numerical simulation, we employ the isometry of the Wasserstein space $(\mathcal{P}(\Omega), \mathbf{W}_2)$ and the space $\widetilde{\mathbf{X}}$ of non-decreasing càdlàg functions $X : [0, 1] \to \Omega$, equipped with the $\mathrm{L}^2(0, 1)$-norm. This isometry is realized by assigning to each μ its inverse distribution function $X_\mu : [0, 1] \to \Omega$, i.e., $\mu\big([-1, X_\mu(\xi)]\big) = \xi$ for all $\xi \in [0, 1]$. Accordingly, the Wasserstein gradient flow transforms into an $\mathrm{L}^2(0, 1)$-gradient flow on $\widetilde{\mathbf{X}}$ with the energy functional

$$\widetilde{\mathcal{E}}(X) := -\int_{[0,1]} \log(\partial_\xi X(\xi))\,\mathrm{d}\xi + \frac{1}{2}\iint_{[0,1]\times[0,1]} W(X(\xi) - X(\eta))\,\mathrm{d}\xi\,\mathrm{d}\eta.$$

Remark 5.4.2. Note that the change of coordinates has transformed the original gradient flow, that had been posed on the metric Wasserstein space, into a gradient flow on (a closed subset of) the Hilbert space $\mathrm{L}^2(0, 1)$. This example is thus rather another illustration for flows on Hilbert spaces. The combination of the BDF2 time discretization with one of the spatial discretizations for Wasserstein gradient flows in multiple space dimensions — where a transformation into a Hilbert space flow is not possible anymore — is currently under investigation.

In the numerical experiments, we prescribe an initial datum u_0 via its inverse distribution function

$$X_{u_0}(\xi) := 2\xi - 1 + \frac{1}{8\pi}\sin(8\pi\xi)\cdot(10(\xi(\xi-0.5)(x-1))+1)\,.$$

Concerning the discretization in space, we proceed as in the previous example, using central finite differences with $K = 50, 100, 250, 500, 1000$ spatial grid points. The qualitative behavior of the reference solution (in original variables with $\tau_{ref} = 10^{-6}$, and $K = 1000$) is sketched in Figure 5.3 (left).

Our results on the numerical error are given in Figure 5.3 (right). The error curves for the implicit Euler and the BDF2 schemes, respectively, are almost perfectly proportional to τ and τ^2. The results are comparable to (and even better than in) the previous example; we do not observe any significant effect of the spatial discretization, even for very small time steps. This indicates that the purely temporal discretization of the original PDE with BDF2 leads an approximation error of τ^2.

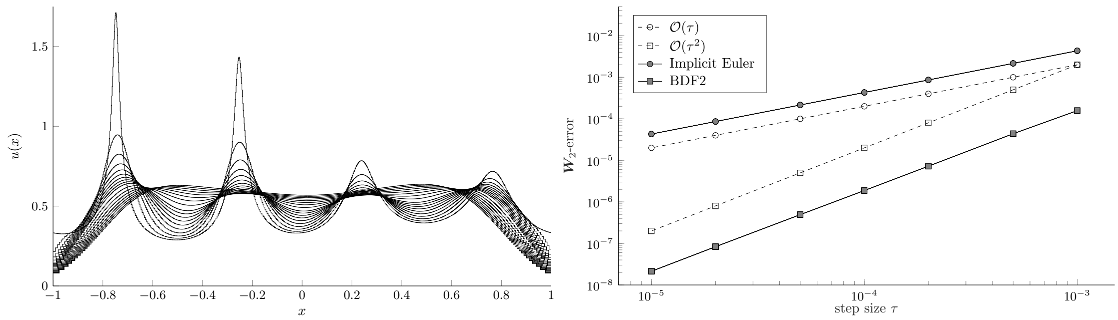


Figure 5.3: *Aggregation-diffusion equation.* Evolution of the reference solution $\overline{u}_t^{\boldsymbol{\tau}_{ref}}$ (left) and the $\mathbf{W}_2$-error plot of $(\overline{u}_t^{\boldsymbol{\tau}})^{(i)}$ compared with $\overline{u}_t^{\boldsymbol{\tau}_{ref}}$ for different K (right).

6 Weighted Energy-Dissipation Principle

This chapter is based on the joint work with S. Lisini, D. Matthes, and G. Savaré [65]. We are concerned with approximating by means of the *Weighted Energy-Dissipation Principle* non-negative solutions of two families of diffusion equations, namely: the second-order, linear Fokker-Planck (FP) equation:

$$\partial_t \rho_t = \Delta \rho_t + \operatorname{div}(\rho_t \nabla V) + \operatorname{div}(\rho_t (\nabla W * \rho_t)); \tag{6.0.1}$$

and the fourth-order Derrida-Lebowitz-Speer-Spohn (DLSS) equation:

$$\partial_t \rho_t = -\operatorname{div}\big(\rho_t \nabla\big(2\frac{\Delta\sqrt{\rho_t}}{\sqrt{\rho_t}}\big)\big). \tag{6.0.2}$$

Both PDEs shall start from the initial configuration ρ_0 in the domain $\Omega = \mathbb{R}^{\mathrm{d}}$. Unfortunately, the method developed in this chapter is not applicable in the case Ω is an open, bounded, and convex subset of $\mathbb{R}^{\mathrm{d}}$. Our method fails in the derivation of the refined a priori bounds (6.2.15) and (6.3.3). However, we formulated as much theorems as possible also for this case since the remaining arguments go through. It is known, if these equations are initialized with $\rho_0 \,\mathrm{d}\mathcal{L}^{\mathrm{d}} \in \mathcal{P}_2(\Omega)$, then there exists a weak solution $\rho_t^* : [0,\infty) \times \Omega \to \mathbb{R} \cup \{\infty\}$ with initial configuration $\rho_0^* = \rho_0$ and for each $t \geq 0$ the measure ρ_t^* is absolutely continuous with respect to the Lebesgue measure.

Gradient Flow Structure. As before, the guiding principle is to exploit the Gradient Flow structure of these drift-diffusion equations, for reference see [4, 54, 87] in the second-order case and [46, 47, 68, 78] in the fourth-order case. The underlying metric space is the space of probability measures $\mathcal{P}_2(\Omega)$ equipped with the L^2-Wasserstein distance $\mathbf{W}_2$ and the corresponding free energy functionals $\mathcal{E}$ are defined by

$$\mathcal{E}(\mu) := \int_\Omega \rho \log(\rho) + \rho V + \rho(W * \rho)\,\mathrm{d}x \qquad \text{or} \qquad \mathcal{E}(\rho) := \int_\Omega \|\nabla(\sqrt{\rho})\|^2 \,\mathrm{d}x \tag{6.0.3}$$

if the measure $\mu = \rho \,\mathrm{d}\mathcal{L}^{\mathrm{d}}$ is absolutely continuous and the integrals on the right-hand side are well-defined, otherwise we set $\mathcal{E}(\mu) = \infty$, see section 2.4 for more details. Then, in the L^2-Wasserstein framework (6.0.1) and (6.0.2) are equivalent to the coupling of the continuity equation with a sort of Darcy's law where the pressure is given by the variational derivative of the free energy functional $\mathcal{E}$:

$$\partial_t \rho_t + \operatorname{div}(\mathbf{w}_t) = 0, \qquad \mathbf{w}_t = -\rho_t \mathrm{D}\frac{\delta \mathcal{E}(\rho_t)}{\delta \rho}. \tag{6.0.4}$$

Weighted Energy-Dissipation Principle. Compared to the modern approach to PDEs with gradient flow structure, that is the well-developed time-discrete theory of *Minimizing Movements* [4, 29, 30], an alternative *time continuous* approach has been proposed recently in [75, 76, 86, 88]. The main idea here is to perturb the gradient flow equation (6.0.4) for an arbitrary free energy functional $\mathcal{E}$ by a elliptic regularization in time

$$-\varepsilon\partial_t^2\rho_t^\varepsilon + \partial_t\rho_t^\varepsilon + \operatorname{div}(\mathbf{w}_t^\varepsilon) = 0, \qquad \mathbf{w}_t^\varepsilon = -\rho_t^\varepsilon \mathrm{D}\frac{\delta\mathcal{E}(\rho_t^\varepsilon)}{\delta\rho}. \tag{6.0.5}$$

Even though one loses the gradient flow structure at first glance, the solutions ρ_t^ε satisfy another crucial variational principle. In particular, it has been shown that solutions ρ_t^ε of (6.0.5) are the minimizer of a global-in-time minimization of the parameter-dependent Weighted-Energy-Dissipation (WED) functional Φ_ε, given by

$$\rho_t^\varepsilon = \operatorname*{argmin}_{\rho_t} \int_0^\infty \frac{e^{-t/\varepsilon}}{\varepsilon}\left(\frac{\varepsilon}{2}\left|\rho_t'\right|^2 + \mathcal{E}(\rho_t)\right)\,\mathrm{d}t. \tag{6.0.6}$$

Here, $|\rho_t'|$ denotes the metric slope of the curve ρ_t and the minimization is performed over the class of L^2-absolutely continuous curves emanating from ρ_0.

It has been proven, that for each $\varepsilon \in (0, \varepsilon_*)$ there exists at least one minimizer ρ_t^ε and in the limit $\varepsilon \to 0$ the approximations ρ_t^ε converge to a limit function ρ_t^* which solves (6.0.4) in the sense of the energy dissipation equality (EDE). Note, it turns out in the analysis of this problem that this perturbed system (6.0.5) possesses a gradient flow structure with respect to the value functional V_ε which is defined as the minimal value in (6.0.6) for a given initial datum ρ_0. I.e., the WED-approximation ρ_t^ε satisfies

$$\frac{\mathrm{d}}{\mathrm{d}t}V_\varepsilon(\rho_t^\varepsilon) = -|(\rho_t^\varepsilon)'|^2 \qquad \text{for a.e. } t.$$

Contribution. The disadvantage of the previously developed WED theory posed in abstract metric spaces is the lack of a classical solution concept. To prove that the limit function ρ_t^* of the WED-approximation ρ_t^ε is indeed a classical distributional solution to the corresponding PDE an adapted notion of convexity is needed. In the context of L^2-Wasserstein gradient flows, convexity along generalized geodesics is a suitable notion. Whilst the free energy functional $\mathcal{E}$ for the Fokker-Planck equation (6.0.1) falls into this class of λ-convex functionals, the energy functional $\mathcal{E}$ for the DLSS equation (6.0.2) is not convex along generalized geodesics. Therefore, one can not deduce that the limit function ρ_t^* of the WED-approximation ρ_t^ε with respect to $\mathcal{E}$ is a weak solution of (6.0.2).

For this reason, the main objective of this part is to apply and cast the WED method for (6.0.2) in a hands-on way and to derive directly a distributional solution of (6.0.2). Even though the case of the Fokker-Planck case is already covered in [86], we apply our WED method also to (6.0.2) as doability check. It turns out the analysis in this "easier" case is of interest by itself, since the lack of regularity, compared to the analysis for $\mathcal{E}$ with respect to the DLSS equation, requires finer arguments and estimates.

Method. The key ingredient of our approach is the equivalent dynamical reformulation of (6.0.6), where we utilize the characterization of L^2-absolutely continuous curves as solutions to the continuity equation, see section 2.2. In particular, the solutions ρ_t^ε of the elliptic regularization of the gradient flow (6.0.5) are also minimizer of our WED functional Ψ_ε:

$$(\rho_t^\varepsilon, \mathbf{w}_t^\varepsilon) = \underset{(\rho_t, \mathbf{w}_t) \in \mathcal{C}(\rho_0)}{\operatorname{argmin}} \int_0^\infty \frac{e^{-t/\varepsilon}}{\varepsilon} \Big(\frac{\varepsilon}{2} \mathcal{K}(\rho_t, \mathbf{w}_t) + \mathcal{E}(\rho_t) \Big) \, \mathrm{d}t \tag{6.0.7}$$

where $\mathcal{C}(\rho_0)$ and $\mathcal{K}$ are as defined in section 2.2. The advantage of this new variational formulation (6.0.7) is the additional degree of freedom in the flux variable $\mathbf{w}_t$. However, in the minimization, the flux variable $\mathbf{w}_t$ is coupled to the density ρ_t via the continuity equation, so this additional degree of freedom is purely virtual. Nevertheless, this formulation of the WED approach encodes the behavior of the metric slope in a better way, which allows us to derive additional properties of the approximated solutions.

A Variation Along the Transport Equation. The first perturbation ρ_t^s of the WED-approximation ρ_t^ε that we consider to obtain the Euler-Lagrange equation is the continuous variation along solutions of the transport equation in the auxiliary time s:

$$\partial_s \rho_t^s + \operatorname{div}(\rho_t^s \cdot \xi_t) = 0 \qquad \rho_t^0 = \rho_t^\varepsilon,$$

where ξ_t is a time-dependent smooth vector field. Note, in order to have a feasible competitor for $(\rho_t^\varepsilon, \mathbf{w}_t^\varepsilon)$, we have to define also perturbations $\mathbf{w}_t^s$ of the associated flux $\mathbf{w}_t^\varepsilon$ due to the coupling of ρ_t^s and $\mathbf{w}_t^s$ through the continuity equation (2.2.5). The key ingredient in the calculation is that we have for both the perturbed density ρ_t^s and the perturbed flux $\mathbf{w}_t^s$ an explicit representation given by

$$(\rho_t^s \circ \mathrm{X}_t^s) \cdot \det(\mathrm{DX}_t^s) = \rho_t^\varepsilon, \qquad (\mathbf{w}^s \circ \mathrm{X}_t^s) \cdot \det(\mathrm{DX}_t^s) = \mathrm{DX}_t^s \mathbf{w}_t^\varepsilon + \rho_t^\varepsilon \cdot \partial_t \mathrm{X}_t^s\,,$$

where X_t^s is the flow map corresponding to the vector field ξ_t, see section 2.4.

B Variation Along the Heat Equation. To derive the refined regularity estimates we adapt the time-discrete flow interchange technique developed by Matthes et al. [68] and transfer it to the time-continuous setting of the Weighted Energy-Dissipation principle. Therefore, we define continuously in the time t the perturbation ρ_t^s as the solution to the heat equation in the auxiliary time s:

$$\partial_s \rho_t^s = \eta_t \Delta(\rho_t^s), \qquad \rho_t^0 = \rho_t^\varepsilon.$$

As in the case before, the key ingredient in variation along the heat flow is the explicit representation of the perturbed WED-approximation $(\rho_t^s, \mathbf{w}_t^s)$

$$\rho_t^s := \mathcal{T}_{\mathfrak{G}_t^s}[\rho_t^\varepsilon], \qquad \mathbf{w}_t^s := \vec{\mathcal{T}}_{\mathfrak{G}_t^s}[\mathbf{w}_t^\varepsilon] - s \partial_t \eta_t \nabla \mathcal{T}_{\mathfrak{G}_t^s}[\rho_t^\varepsilon]$$

in terms of the corresponding Greens function $\mathfrak{G}_t^s$ for the heat equation. This allows us to exploit the monotonicity property of the kinetic energy $\mathcal{K}$ given by lemma 2.5.6 and derive an estimate for the derivative of the kinetic energy $\mathcal{K}$ along solutions to the heat equation.

Main Results. Our main results concerning the well-posedness and the limit behavior as $\varepsilon \searrow 0$ of the WED-approximation ρ_t^ε are stated in the following two theorems. First, our convergence result for the linear Fokker-Planck equation.

Theorem 6.0.1 (Main Result: Linear Fokker-Planck Equation). *Let $\Omega = \mathbb{R}^{\mathrm{d}}$ and let V, W satisfy Assumption 6.2.1 and define the corresponding free energy functional $\mathcal{E}$ for the linear Fokker-Planck equation* (6.0.1).

a) ***Existence.*** *For each $\varepsilon \in (0, \varepsilon_*)$ and each $\rho_0 \in \mathcal{D}(\mathcal{E})$ there exists an approximated solution ρ_t^ε with respect to $\mathcal{E}$.*

b) ***Convergence.*** *Given a vanishing sequence $(\varepsilon_n)_{n\in\mathbb{N}}$ with $\varepsilon_n \in (0, \varepsilon_*)$. Then, there exits a (non-relabeld) subsequence ε_n and a limit function $\rho_t^* : [0,\infty) \times \Omega \to \mathbb{R}$ with $\rho_t^* \in \mathcal{P}_2^{ac}(\Omega)$ such that for each $p \in [1, \infty)$ and for all $T > 0$ one has:*

$$\rho_t^{\varepsilon_n} \to \rho_t^* \qquad \text{strongly in } \mathrm{L}^p(0,T;\mathrm{L}^1(\Omega)).$$

c) ***Solution.*** *The limit function ρ_t^* from b) is a solution of the linear Fokker-Planck equation* (6.0.1) *in the weak sense of* (6.2.22).

Second, our main result about the existence and convergence of the Weighted Energy-Dissipation principle applied to the Derrida-Lebowitz-Speer-Spohn equation is given as follows.

Theorem 6.0.2 (Main Result: Derrida-Lebowitz-Speer-Spohn Equation). *Let $\Omega = \mathbb{R}^{\mathrm{d}}$ and let $\mathcal{E}$ be the corresponding free energy functional for the Derrida-Lebowitz-Speer-Spohn equation* (6.0.2).

a) ***Existence.*** *For each $\varepsilon \in (0, \varepsilon_*)$ and each $\rho_0 \in \mathcal{D}(\mathcal{E})$ there exists an approximated solution ρ_t^ε with respect to $\mathcal{E}$.*

b) ***Convergence.*** *Given a vanishing sequence $(\varepsilon_n)_{n\in\mathbb{N}}$ with $\varepsilon_n \in (0, \varepsilon_*)$. Then, there exits a (non-relabeld) subsequence ε_n and a limit function $\rho_t^* : [0,\infty) \times \Omega \to \mathbb{R}$ with $\rho_t^* \in \mathcal{P}_2^{ac}(\Omega)$ such that for each $p \in [1, \infty)$ and for all $T > 0$ one has:*

$$\begin{aligned} \rho_t^{\varepsilon_n} &\to \rho_t^* && \text{strongly in } \mathrm{L}^p(0,T;\mathrm{L}^1(\Omega)), \\ \sqrt{\rho_t^{\varepsilon_n}} &\to \sqrt{\rho_t^*} && \text{strongly in } \mathrm{L}^2(0,T;\mathrm{H}^1(\Omega)), \\ \sqrt{\rho_t^{\varepsilon_n}} &\rightharpoonup \sqrt{\rho_t^*} && \text{weakly in } \mathrm{L}^2(0,T;\mathrm{H}^2(\Omega)). \end{aligned}$$

c) ***Solution.*** *The limit function ρ_t^* from b) is a solution of the Derrida-Lebowitz-Speer-Spohn equation* (6.0.2) *in the weak sense of* (6.3.6).

6.1 Introduction to the WED-Principle in Metric Spaces

In [86] Rossi et al. investigated in the framework of abstract metric gradient flows the WED principle (6.0.6). This method focuses on the minimization of the parameter-dependent global-in-time functional of trajectories

$$\Phi_\varepsilon(\rho_t) := \int_0^\infty \frac{e^{-t/\varepsilon}}{\varepsilon} \Big(\frac{\varepsilon}{2}\,|\rho_t'|^2 + \mathcal{E}(\rho_t)\Big)\,\mathrm{d}t$$

featuring the weighted sum of energetic and dissipative terms. In particular, the theory therein covers the case of both free energy functionals from (6.0.3), since these energies satisfy the *lower semicontinuity-coercivity-compactness* (LSCC) conditions, cf. [4]. Contrary to [86], we investigate in this chapter the minimization of the parameter-dependent global-in-time functional Ψ_ε which depends additionally on the flux $\mathbf{w}_t$. So Ψ_ε reads

$$\Psi_\varepsilon(\rho_t, \mathbf{w}_t) := \int_0^\infty \frac{e^{-t/\varepsilon}}{\varepsilon} \Big(\frac{\varepsilon}{2}\mathcal{K}(\rho_t, \mathbf{w}_t) + \mathcal{E}(\rho_t)\Big)\,\mathrm{d}t.$$

But in addition, in the minimization of Ψ_ε the density-flux pair $(\rho_t, \mathbf{w}_t)$ is coupled through the continuity equation (2.2.5). The link between the two minimization problems (6.0.6) and (6.0.7) is given by the characterization of L^2-absolutely continuous curves via the continuity equation, see Theorem 2.2.1. So it is clear that $\Phi_\varepsilon(\rho_t) \le \Psi_\varepsilon(\rho_t, \mathbf{w}_t)$ for all $(\rho_t, \mathbf{w}_t) \in \mathcal{C}(\rho^0)$. Visa versa, there exists always some density-flux pair $(\rho_t, \mathbf{w}_t) \in \mathcal{C}(\rho_0)$ such that $\Psi_\varepsilon(\rho_t, \mathbf{w}_t) \le \Phi_\varepsilon(\rho_t)$. Hence, the minimization problems (6.0.6) and (6.0.7) are equivalent and therefore, we recover the main results in [86] also for $(\rho_t^\varepsilon, \mathbf{w}_t^\varepsilon)$, namely: *lower boundedness of* Ψ_ε, *existence of minimizer* ρ_t^ε, *inner variation of the approximation* ρ_t^ε, *fundamental identity of the value function* V_ε, and ε-*independent bounds of the approximation* $(\rho_t^\varepsilon, \mathbf{w}_t^\varepsilon)$.

Setup and Assumptions. In the following, we want to briefly recap the aforementioned results, which are formulated with respect to our framework and for some arbitrary free energy functional $\mathcal{E}$ which includes (6.0.3). We emphasize, that we state the idea of a proof rather than giving actually detailed and riguours proofs, which all can be found in [86]. Therefore, we shall assume in the sequel the standard *LSCC*-conditions on the free energy functional $\mathcal{E}$.

Assumption 6.1.1. The free energy functional $\mathcal{E} : \mathcal{P}_2(\Omega) \to \mathbb{R} \cup \{\infty\}$ is proper and satisfy the following regularity conditions:

(E1) **Lower Semi-continuity.** $\mathcal{E}$ is sequentially lower semi-continuous with respect to weak* convergence of measures, i.e.,

$$\rho_n \rightharpoonup^* \rho_* \qquad \Rightarrow \qquad \mathcal{E}(\rho_*) \le \liminf_{n\to\infty} \mathcal{E}(\rho_n).$$

(E2) **Coercivity.** There exists $C > 0$ s.t. $\mathcal{E}(\rho) \ge -C(1 + \boldsymbol{M}_2(\rho))$ for all $\rho \in \mathcal{P}_2(\Omega)$.

(E3) **Compactness.** For each $C > 0$ the set $\{\rho \in \mathcal{P}_2(\Omega) \mid \mathcal{E}(\rho) \le C\}$ is sequentially compact in the topology induced by the weak* convergence of measures.

Lower Bounds. Recall that assumptions (E0)–(E3) for the free energy functional $\mathcal{E}$ hold and that we assume from now on $\varepsilon_* < 1/(32C)$. The first result is an auxiliary calculation which is essential in the proof of the lower boundedness of the WED-functional Ψ_ε.

Lemma 6.1.2. *Let $\rho_t \in \mathrm{AC}^2(0,T;(\mathcal{P}_2(\Omega),\mathbf{W}_2))$ be an absolutely continuous curve with integrable metric derivative $|\rho_t'| \in \mathrm{L}^1(0,\infty)$. Then*

$$\frac{1}{4\varepsilon^2}\int_0^\infty \frac{e^{-t/\varepsilon}}{\varepsilon}\Big(\int_0^t |\rho_r'|\,\mathrm{d}r\Big)^2\,\mathrm{d}t \le \int_0^\infty \frac{e^{-t/\varepsilon}}{\varepsilon}\,|\rho_t'|^2\,\mathrm{d}t. \tag{6.1.1}$$

Sketch of Proof. This results follows from the binomial formula and an integration by parts. Define the auxiliary function $L_t := \int_0^t |\rho_r'|\,\mathrm{d}r$ such that $L_t' = |\rho_t'|$. Then, one can compute

$$0 \le \int_0^\infty \frac{e^{-t/\varepsilon}}{\varepsilon}\Big(L_t' - \frac{1}{2\varepsilon}L_t\Big)^2\,\mathrm{d}t = \int_0^\infty \frac{e^{-t/\varepsilon}}{\varepsilon}\big((L_t')^2 + \frac{1}{4\varepsilon^2}(L_t)^2\big)\,\mathrm{d}t - \int_0^\infty \frac{e^{-t/\varepsilon}}{\varepsilon^2}L_t'L_t\,\mathrm{d}t.$$

Note, the second integral can be rewritten as

$$-\int_0^\infty \frac{e^{-t/\varepsilon}}{\varepsilon^2}L_t'L_t\,\mathrm{d}t = -\int_0^\infty \frac{e^{-t/\varepsilon}}{\varepsilon^2}\frac{\mathrm{d}}{\mathrm{d}t}\big[\frac{1}{2}(L_t)^2\big]\,\mathrm{d}t = -\int_0^\infty \frac{e^{-t/\varepsilon}}{\varepsilon^3}\frac{1}{2}(L_t)^2\,\mathrm{d}t$$

thanks to the integration by parts formula and the integrability of the metric velocity $|\rho_t'| \in \mathrm{L}^1(0,\infty)$. Insert this result in the first inequality and after a rearrangement this is the desired result. □

Lemma 6.1.3 (Lower Bound)**.** *Let $\rho_t \in \mathrm{AC}^2(0,T;(\mathcal{P}_2(\Omega),\mathbf{W}_2))$ with integrable metric velocity $|\rho_t'| \in \mathrm{L}^1(0,\infty)$. For each $\varepsilon \in (0,\varepsilon_*)$ one has*

$$\Psi_\varepsilon(\rho_t,\mathbf{w}_t) \ge \int_0^\infty \frac{e^{-t/\varepsilon}}{4}\mathcal{K}(\rho_t,\mathbf{w}_t)\,\mathrm{d}t - C\big(1+2\boldsymbol{M}_2(\rho_0)\big). \tag{6.1.2}$$

Sketch of Proof. Use the notation as above, i.e., $L_t := \int_0^t |\rho_r'|\,\mathrm{d}r$. Then, by characterization of L^2-absolutely continuous curves we have $\mathcal{K}(\rho_t,\mathbf{w}_t) \ge (L_t')^2$. Using this and the coercivity of the free energy functional $\mathcal{E}$ we can estimate Ψ_ε from below as follows

$$\Psi_\varepsilon(\rho_t,\mathbf{w}_t) \ge \int_0^\infty \frac{e^{-t/\varepsilon}}{\varepsilon}\Big(\frac{\varepsilon}{4}\mathcal{K}(\rho_t,\mathbf{w}_t) + \frac{\varepsilon}{4}(L_t')^2 - C(1+\boldsymbol{M}_2(\rho_t)\Big)\,\mathrm{d}t. \tag{6.1.3}$$

Due to the triangle inequality, we can estimate the second moment of ρ_t as follows

$$\boldsymbol{M}_2(\rho_t) \le 2\boldsymbol{M}_2(\rho_0) + 2\mathbf{W}_2^2(\rho_0,\rho_t) \le 2\boldsymbol{M}_2(\rho_0) + 2(L_t)^2.$$

Insert this into (6.1.3) to obtain

$$\Psi_\varepsilon(\rho_t,\mathbf{w}_t) \ge \int_0^\infty \frac{e^{-t/\varepsilon}}{4}\mathcal{K}(\rho_t,\mathbf{w}_t)\,\mathrm{d}t - C\big(1+2\boldsymbol{M}_2(\rho_0)\big) + \int_0^\infty \frac{e^{-t/\varepsilon}}{\varepsilon}\big(\frac{\varepsilon}{4}(L_t')^2 - 2C(L_t)^2\big)\,\mathrm{d}t.$$

Lastly, since $\varepsilon < \varepsilon_* < 1/(32C)$ the second integral is non-negative by (6.1.3) and hence we obtain the desired lower bound (6.1.2). □

Existence of Minimizers. Recall that assumptions (E0)–(E3) for the free energy functional $\mathcal{E}$ hold. Then, the existence of a minimizer of the WED-approximation follows by the extension of the Aubin-Lions Theorem.

Theorem 6.1.4 (Existence of Minimizers). *Let $\varepsilon \in (0, \varepsilon_*)$. For each $\rho_0 \in \mathcal{D}(\mathcal{E})$ the minimization problem* (6.0.7) *has at least one minimizer $(\rho_t^\varepsilon, \mathbf{w}_t^\varepsilon)$ in the class $\mathcal{C}(\rho^0)$.*

Sketch of Proof. Let $((\rho_t^n, \mathbf{w}_t^n))_{n\in\mathbb{N}}$ be an infimizing sequence for Ψ_ε of density-flux pairs $(\rho_t^n, \mathbf{w}_t^n) \in \mathcal{C}(\rho_0)$ and fix some $T > 0$. Since the density-flux pair $(\rho^0, \vec{0})$ is feasible in (6.0.7) and since the free energy functional $\mathcal{E}$ is coercive (E2), we can assume that with out loss of generality there exits a non-negative constant C such that

$$\sup_n \int_0^T \mathcal{K}(\rho_t^n, \mathbf{w}_t^n)\,\mathrm{d}t < Ce^{T/\varepsilon} \qquad \text{and} \qquad \sup_n \int_0^T \mathcal{E}(\rho_t^n)\,\mathrm{d}t < Ce^{T/\varepsilon} \tag{6.1.4}$$

thanks to (6.1.2). To prove the existence of minimizers for (6.0.7) we seek to apply the extension of Aubin-Lions Lemma for metric spaces theorem 2.5.4 to $(\rho_t^n)_{n\in\mathbb{N}}$ with the underlying metric space $\mathbf{X} = \mathcal{P}_2(\Omega)$ with $\boldsymbol{d}$ induced by the weak*-topology of measures. We choose as normal coercive functional $\mathcal{A}_t = \mathcal{E} + C\boldsymbol{M}_2$, where C is the constant from (E2), and as pseudo distance $g(\mu, \nu) = \mathbf{W}_2(\mu, \nu)$.

Due to (6.1.4), the sequence $(\rho_t^n)_{n\in\mathbb{N}}$ is tight with respect to $\mathcal{A}$. The weak integral equi-continuity condition for $(\rho_t^n)_{n\in\mathbb{N}}$ follows by the Benamou-Brenier formula. The density flux pair $(\rho_{t+sh}^n, h\mathbf{w}_{t+sh}^n)$ solves the continuity equation and connects ρ_t and ρ_{t+h} and therefore we can estimate

$$\mathbf{W}_2^2(\rho_t^n, \rho_{t+h}^n) \le \int_0^1 \mathcal{K}(\rho_{t+sh}^n, h\mathbf{w}_{t+sh}^n)\,\mathrm{d}t = \int_t^{t+h} h\mathcal{K}(\rho_s^n, \mathbf{w}_s^n)\,\mathrm{d}s.$$

With this estimate at hand one can deduce that $(\rho_t^n)_{n\in\mathbb{N}}$ satisfies the equi-continuity condition and therefore Theorem 2.5.4 implies (on a subsequence) that $(\rho_t^n)_{n\in\mathbb{N}}$ converges to some ρ_t^* in $\mathcal{M}(0, T; (\mathcal{P}_2(\Omega), \boldsymbol{d}))$. After possibly extracting another subsequence, we can conclude that $\rho_t^n \rightharpoonup^* \rho_t^*$ for almost every $t \in [0, \infty)$.

The compactness of $(\mathbf{w}_t^n)_{n\in\mathbb{N}}$ can be derived as follows. By Hölder's inequality we have

$$\int_0^T \int_\Omega \|\mathbf{w}_t^n\|\,\mathrm{d}x\,\mathrm{d}t \le \Big(\int_0^T \mathcal{K}(\rho_t^n, \mathbf{w}_t^n)\,\mathrm{d}t\Big)^{1/2} \Big(\int_0^T \int_\Omega \rho_t^n\,\mathrm{d}x\,\mathrm{d}t\Big)^{1/2} \le Ce^{T/2\varepsilon}T^{1/2}.$$

Hence, the sequence $(\mathbf{w}_t^n)_{n\in\mathbb{N}}$ is bounded in $\mathrm{L}^1([0, T] \times \Omega)$. Since the map $t \mapsto \boldsymbol{M}_2(\rho_t^n)$ is uniformly integrable by (6.1.4), we can conclude by Hölder's inequality that the map $(t, x) \mapsto \|x\|\|\mathbf{w}_t^n\|$ is also uniformly bounded. As long as the map $(t, x) \mapsto \|x\|$ has compact sublevels in $[0, T] \times \Omega$ the sequence $(\mathbf{w}_t^n)_{n\in\mathbb{N}}$ is tight. This implies by Prokhorov's compactness theorem that $\mathbf{w}_t^n \rightharpoonup^* \mathbf{w}_t^*$ for every $t \in [0, T]$ and with a diagonal argument we can extend the weak*-convergence for almost all $t \in [0, \infty)$.

Clearly, the continuity equation is stable with respect to weak*-convergence of measures and therefore $(\rho_t^*, \mathbf{w}_t^*) \in \mathcal{C}(\rho_0)$. Since $\mathcal{K}$ and $\mathcal{E}$ are lower semi-continuous, it follows by Fatou's Lemma that the limit $(\rho_t^*, \mathbf{w}_t^*)$ of the infimizing sequence $((\rho_t^n, \mathbf{w}_t^n))_{n\in\mathbb{N}}$ is a minimizer of the minimization problem (6.0.7). □

Variational Properties of the Value Function V_ε. By the previous part, the WED-approximation $(\rho_t^\varepsilon, \mathbf{w}_t^\varepsilon)$ is well-defined for each $\varepsilon \in (0, \varepsilon_*)$ and for each initial datum $\rho_0 \in \mathcal{D}(\mathcal{E})$. So similar to the Moreau-Yosida approximation ϕ of the Minimizing Movement scheme, the value function V_ε defined by

$$V_\varepsilon(\rho_0) := \inf_{(\rho_t, \mathbf{w}_t) \in \mathcal{C}(\rho_0)} \Psi_\varepsilon(\rho_t, \mathbf{w}_t) = \Psi_\varepsilon(\rho_t^\varepsilon, \mathbf{w}_t^\varepsilon)$$

plays a crucial role in the analysis of the Weighted Energy-Dissipation approach. It turns out that the value function V_ε encodes the dissipation of the WED-approximation $(\rho_t^\varepsilon, \mathbf{w}_t^\varepsilon)$, i.e.,

$$\frac{\mathrm{d}}{\mathrm{d}t} V_\varepsilon(\rho_t^\varepsilon) = -|(\rho_t^\varepsilon)'|^2 \qquad \text{for a.e. } t. \tag{6.1.5}$$

Hence, the WED principle can be understood as a perturbation of the free energy functional $\mathcal{E}$ by the value function V_ε which on the one hand destroys the physical time-causality, but which preserves the gradient flow structure. To derive (6.1.5) we need the following two theorems.

Theorem 6.1.5 (Inner Variation)**.** *The map* $\mathfrak{V}_t^\varepsilon := -\frac{\varepsilon}{2}\mathcal{K}(\rho_t^\varepsilon, \mathbf{w}_t^\varepsilon) + \mathcal{E}(\rho_t^\varepsilon)$ *belongs to* $W^{1,1}(0,T)$ *and it fullfills*

$$\frac{\mathrm{d}}{\mathrm{d}t} \mathfrak{V}_t^\varepsilon = -\mathcal{K}(\rho_t^\varepsilon, \mathbf{w}_t^\varepsilon) \qquad \text{in } \mathcal{D}'(0,T). \tag{6.1.6}$$

Sketch of Proof. Given some arbitrary $\eta_t \in \mathcal{C}_c^\infty((0,\infty))$, define the family of smooth diffeomorpishms of $(0,\infty)$ by $S_t^s := t + s\eta_t$ and denote by T_t^s the inverse of the map $t \mapsto S_t^s$. Further, we define the perturbation $(\rho_t^s, \mathbf{w}_t^s)$ of the WED-approximation $(\rho_t^\varepsilon, \mathbf{w}_t^\varepsilon)$ via $\rho_t^s := \rho_{T_t^s}^\varepsilon$ and $\mathbf{w}_t^s := \partial_t T_t^s \cdot \mathbf{w}_{T_t^s}^\varepsilon$. By this definition the density-flux pair $(\rho_t^s, \mathbf{w}_t^s)$ satisfies for each s the continuity equation (2.2.5). Due to $\eta_0 = 0$, we have $\rho_0^s = \rho_0$ and hence $(\rho_t^s, \mathbf{w}_t^s) \in \mathcal{C}(\rho_0)$. With the change of variables $t = S_r^s$ we can rewrite $\Psi_\varepsilon(\rho_t^s, \mathbf{w}_t^s)$ as follows

$$\Psi_\varepsilon(\rho_t^s, \mathbf{w}_t^s) = \int_0^\infty \frac{e^{-S_r^s/\varepsilon}}{\varepsilon} \Big[\frac{\varepsilon}{2}\frac{1}{\partial_r S_r^s}\mathcal{K}(\rho_r^\varepsilon, \mathbf{w}_r^\varepsilon) + \mathcal{E}(\rho_r^\varepsilon)\partial_r S_r^s\Big]\,\mathrm{d}r.$$

By taking the derivative with respect to s at the minimum point $s = 0$ one obtains

$$\begin{aligned} 0 =& \frac{\mathrm{d}}{\mathrm{d}s}\big[\Psi_\varepsilon(\rho_t^s, \mathbf{w}_t^s)\big]_{s=0} \\ =& \int_0^\infty \frac{\mathrm{d}}{\mathrm{d}s}\Big[\frac{e^{-S_r^s/\varepsilon}}{\varepsilon}\frac{\varepsilon}{2}\frac{1}{\partial_r S_r^s}\Big]_{s=0}\mathcal{K}(\rho_r^\varepsilon, \mathbf{w}_r^\varepsilon) + \frac{\mathrm{d}}{\mathrm{d}s}\Big[\frac{e^{-S_r^s/\varepsilon}}{\varepsilon}\partial_r S_r^s\Big]_{s=0}\mathcal{E}(\rho_r^\varepsilon)\,\mathrm{d}r \\ =& \int_0^\infty \frac{e^{-t/\varepsilon}}{\varepsilon}\Big[-\frac{1}{2}\eta_r - \frac{\varepsilon}{2}\partial_r\eta_r\Big]\mathcal{K}(\rho_r^\varepsilon, \mathbf{w}_r^\varepsilon) + \frac{e^{-t/\varepsilon}}{\varepsilon}\Big[-\frac{1}{\varepsilon}\eta_r + \partial_r\eta_r\Big]\mathcal{E}(\rho_r^\varepsilon)\,\mathrm{d}r. \end{aligned} \tag{6.1.7}$$

Note, we used in the last step the identity $\frac{\mathrm{d}}{\mathrm{d}s}S_r^s = \eta_r$. Lastly, choose $\eta_t = \varepsilon e^{t/\varepsilon}\varphi_t$ for a test-function $\varphi_t \in \mathcal{C}_c^\infty((0,\infty))$ and simplify the right hand side of (6.1.7):

$$0 = \int_0^\infty (-\varphi_r - \frac{\varepsilon}{2}\partial_r\varphi_r)\mathcal{K}(\rho_r^\varepsilon, \mathbf{w}_r^\varepsilon) + \partial_r\varphi_r\mathcal{E}(\rho_r^\varepsilon)\,\mathrm{d}r = \int_0^\infty -\varphi_r\mathcal{K}(\rho_r^\varepsilon, \mathbf{w}_r^\varepsilon) + \partial_r\varphi_r\mathfrak{V}_t^\varepsilon\,\mathrm{d}r. \quad \square$$

Theorem 6.1.6 (Fundamental Identity)**.** *Given the WED-approximation* $(\rho_t^\varepsilon, \mathbf{w}_t^\varepsilon)$*, the value function* V_ε*, and the map* $\mathfrak{V}_\varepsilon(t)$ *defined in theorem 6.1.5. Then,*

$$V_\varepsilon(\rho_t^\varepsilon) = \mathfrak{V}_t^\varepsilon \qquad \textit{for a.e. } t \in (0,\infty). \tag{6.1.8}$$

Sketch of Proof. Every minimizer $(\rho_t^\varepsilon, \mathbf{w}_t^\varepsilon)$ of Ψ_ε satisfies

$$V_\varepsilon(\rho_t^\varepsilon) = \int_0^\infty \frac{e^{-s/\varepsilon}}{\varepsilon} \Big(\frac{\varepsilon}{2}\mathcal{K}(\rho_{t+s}^\varepsilon, \mathbf{w}_{t+s}^\varepsilon) + \mathcal{E}(\rho_{t+s}^\varepsilon)\Big)\,\mathrm{d}s \tag{6.1.9}$$

since by the dynamical programming principle the value function can be also defined as

$$V_\varepsilon(\rho_0) = \inf_{(\rho_t,\mathbf{w}_t)\in\mathcal{C}(\rho_0)} \Big[\int_0^T \frac{e^{-t/\varepsilon}}{\varepsilon}\Big(\frac{\varepsilon}{2}\mathcal{K}(\rho_t,\mathbf{w}_t) + \mathcal{E}(\rho_t)\Big)\,\mathrm{d}t + e^{-T/\varepsilon}V_\varepsilon(\rho_T)\Big],$$

see also [6] for further reference. If we denote $\mathfrak{V}_t^\varepsilon$ by the absolutely continuous representation of the map $t \mapsto \mathfrak{V}_t^\varepsilon$, we can rewrite (6.1.9) due to (6.1.6) to

$$\begin{aligned} V_\varepsilon(\rho_t^\varepsilon) &= \int_0^\infty \frac{e^{-s/\varepsilon}}{\varepsilon}\Big(\big[-\frac{\varepsilon}{2}+\varepsilon\big]\mathcal{K}^2(\rho_{t+s}^\varepsilon, \mathbf{w}_{t+s}^\varepsilon) + \mathcal{E}(\rho_{t+s}^\varepsilon)\Big)\,\mathrm{d}s \\ &= \int_0^\infty \frac{e^{-s/\varepsilon}}{\varepsilon}\Big(\mathfrak{V}_{t+s}^\varepsilon - \varepsilon\frac{\mathrm{d}}{\mathrm{d}t}[\mathfrak{V}_{t+s}^\varepsilon]\Big)\,\mathrm{d}s = \int_0^\infty \frac{e^{-s/\varepsilon}}{\varepsilon}\Big(\mathfrak{V}_{t+s}^\varepsilon - \varepsilon\frac{\mathrm{d}}{\mathrm{d}s}[\mathfrak{V}_{t+s}^\varepsilon]\Big)\,\mathrm{d}s. \end{aligned}$$

Applying integration by parts to the second integrand yields (6.1.8). □

ε-independent Bounds. Lastly, we prove the surrogate of the classical estimates form the Minimizing Movement scheme. Here, we exploit the hidden gradient flow structure (6.1.5) of the WED principle.

Theorem 6.1.7 (ε-independent Bounds)**.** *There exist* ε*-independent and a non-negative constant* C *which depends on* T *such that*

$$\int_0^T \mathcal{K}(\rho_t^\varepsilon, \mathbf{w}_t^\varepsilon)\,\mathrm{d}t \le C, \qquad \textit{and} \qquad \int_0^T \mathcal{E}(\rho_t^\varepsilon)\,\mathrm{d}t \le C. \tag{6.1.10}$$

Sketch of Proof. Differentiate the fundamental identity (6.1.8) with respect to t and insert the inner variation formula (6.1.6) to get

$$\frac{\mathrm{d}}{\mathrm{d}t}V_\varepsilon(\rho_t^\varepsilon) = \frac{\mathrm{d}}{\mathrm{d}t}\mathfrak{V}_\varepsilon(t) = -\mathcal{K}(\rho_t^\varepsilon, \mathbf{w}_t^\varepsilon) \qquad \text{for a.e. } t \in (0,\infty).$$

This shows, that the map $t \mapsto V_\varepsilon(\rho_t^\varepsilon)$ is monotonically decreasing. Therefore, we have

$$\int_0^T \mathcal{K}(\rho_t^\varepsilon, \mathbf{w}_t^\varepsilon)\,\mathrm{d}t = V_\varepsilon(\rho_0) - V_\varepsilon(\rho_\varepsilon^T) \le \mathcal{E}(\rho_0) + C(1 + \boldsymbol{M}_2(\rho_0))$$

where the estimate follows from $V_\varepsilon(\rho_0) \le \mathcal{E}(\rho_0)$ and from (6.1.2). Likewise it follows from the fundamental identity (6.1.8) and the monotonicity of the map $t \mapsto V_\varepsilon(\rho_t^\varepsilon)$ that

$$\mathcal{E}(\rho_t^\varepsilon) = V_\varepsilon(\rho_t^\varepsilon) + \frac{\varepsilon}{2}\mathcal{K}(\rho_t^\varepsilon, \mathbf{w}_t^\varepsilon) \le \mathcal{E}(\rho^0) + \frac{\varepsilon}{2}\mathcal{K}(\rho_t^\varepsilon, \mathbf{w}_t^\varepsilon).$$

A further integration over $(0,T)$ yields the second estimate from (6.1.10). □

6.2 Application to Linear Fokker-Planck Equation

In this section, we want to apply the Weighted Energy-Dissipation principle to the linear Fokker-Planck equation, given by

$$\partial_t \rho_t = \Delta\rho_t + \operatorname{div}(\rho_t \nabla V) + \operatorname{div}(\rho_t(\nabla W * \rho_t)) \tag{6.2.1}$$

starting from the initial configuration ρ_0 on $\Omega = \mathbb{R}^{\mathrm{d}}$. Recall, that most of the proofs work also in the case of $\Omega \subset \mathbb{R}^{\mathrm{d}}$ which is open, bounded, and convex. However our method fails in the derivation of the refined a priori bounds, see section 6.2.3 for further explanaition. Recall, the corresponding free energy functional $\mathcal{E}$ in the L^2-Wasserstein framework is given by

$$\mathcal{E}(\mu) := \int_\Omega \rho\log(\rho) + \rho V + \rho(W * \rho)\,\mathrm{d}x$$

if the measure $\mu = \rho\,\mathrm{d}\mathcal{L}^{\mathrm{d}} \in \mathcal{P}_2(\Omega)$ is absolutely continuous and $\rho \in \mathrm{L}^1\log(\mathrm{L}^1)(\Omega)$, otherwise we set $\mathcal{E}(\mu) = \infty$.

Method. Our approximation of the solution to the linear Fokker-Planck equation (6.2.1) by means of the Weighted Energy-Dissipation Principle reads as follows

Scheme. Given the free energy functional $\mathcal{E}$ and an initial configuration ρ_0, define the WED-approximation $(\rho_t^\varepsilon, \mathbf{w}_t^\varepsilon)$ for a given perturbation $\varepsilon \in (0, \varepsilon_*)$ as the minimizer of the WED-functional Ψ_ε, i.e.

$$(\rho_t^\varepsilon, \mathbf{w}_t^\varepsilon) \in \operatorname*{argmin}_{(\rho_t, \mathbf{w}_t)\in\mathcal{C}(\rho_0)} \int_0^\infty \frac{e^{-t/\varepsilon}}{\varepsilon} \int_\Omega \frac{\varepsilon}{2}\frac{\|\mathbf{w}_t\|^2}{\rho_t} + \rho_t\log(\rho_t) + \rho_t V + \rho_t(W * \rho_t)\,\mathrm{d}x\,\mathrm{d}t.$$

If we assume the proper assumptions 6.2.1 the free energy functional $\mathcal{E}$ satisfies (E1)–(E3) and we can use the results from the previous section, specifically the ε-independent bounds (6.1.10) from theorem 6.1.7.

Strategy of the Proof. The structure of the convergence prove of this scheme is done in three steps. Firstly, we derive the time-continuous Euler-Lagrange equations in section 6.2.2. This is done, in the same manner as compared to the original method by Jordan et al. [54], by defining a time-continuous perturbation ρ_t^s of the WED-approximation $(\rho_t^\varepsilon, \mathbf{w}_t^\varepsilon)$ along solutions to the Transport equation with vector field ξ_t. Secondly, we use in section 6.2.3 another perturbation of the WED-approximation $(\rho_t^\varepsilon, \mathbf{w}_t^\varepsilon)$ along solutions to the heat equation to derive refined regularity bounds. Note, this perturbation is a to the time continuous WED-setting adapted version of the flow-interchange theorem developed in [68]. Lastly, these refined regularity bounds are then sufficient to pass in the Euler-Lagrange equation to the limit $\varepsilon \to 0$ by means of the extension of Aubin-Lions compactness theorem for Banach spaces 2.5.4, see therefore section 6.2.4.

6.2.1 Assumptions

Our assumption on the confinement potential V and the interaction kernel W reads as follows:

Assumption 6.2.1. The confinement potential V and the interaction kernel W satisfy:

(F1) $V \in \mathcal{C}^2(\Omega)$, $W \in \mathcal{C}^2(\mathbb{R}^{\mathrm{d}})$, and W is symmetric.

(F2) There exists some non-negative constant d_1 such that

$$|V(x)|\,,\ |W(x)|\,,\ \|\nabla V(x)\|\,,\ \|\nabla W(x)\| \le d_1\big(1+\|x\|^2\big).$$

(F3) There exits some non-negative constant d_2 such that

$$|\Delta V(x)|\,, |\Delta W(x)| \le d_2.$$

6.2.2 Time-Continuous Euler-Lagrange Equations

Next, we want to derive the corresponding Euler-Lagrange equations of the WED-approximation $(\rho_t^\varepsilon, \mathbf{w}_t^\varepsilon)$. The idea is to exploit the differential structure of the L^2-Wasserstein space as in the original paper [54] where the JKO-method was developed. For this instance, we have to transfer the method from the time-discrete setting of the Minimizing Movement scheme to the time-continuous setting of the Weighted Energy-Dissipation principle.

Definition and Justification of the Perturbation. To do so, we perturb our approximation ρ_t^ε, with $\rho_t^\varepsilon \in \mathcal{P}_2^{\mathrm{ac}}(\Omega)$, at each time $t \ge 0$ along the Transport equation in the auxiliary time s

$$\partial_s \rho_t^s + \operatorname{div}(\rho_t^s\, \xi_t) = 0 \qquad \rho_t^0 = \rho_t^\varepsilon. \tag{6.2.2}$$

with time-dependent velocity field $\xi_t \in \mathcal{C}_c^\infty([0,\infty)\times\overline{\Omega}, \mathbb{R}^{\mathrm{d}})$ with $\xi_0 = 0$ and $\xi_t \cdot \boldsymbol{n} = 0$ on $\partial\Omega$. It is clear from section 2.4 that the perturbed density ρ_t^s has the explicit representation $\rho_t^s := (\mathrm{X}_t^s)_\# \rho_t^\varepsilon$, where X_t^s is the flow map corresponding to the velocity field ξ_t, i.e., X_t^s is the solution to of the initial value problem

$$\tfrac{\mathrm{d}}{\mathrm{d}s}\mathrm{X}_t^s = \xi_t(\mathrm{X}_t^s), \qquad \text{with} \qquad \mathrm{X}_t^0(x) = x,$$

The flow map X_t^s exists, is jointly continuous in (s,t,x), and for fixed (s,t) the flow map X_t^s is a diffeomorphism on Ω. Further more, we define by $\mathfrak{X}_t^s := \det(\mathrm{D}\mathrm{X}_t^s) > 0$ the volume distortion of the flow map X_t^s. With this notation at hand, we can define the corresponding perturbation $(\rho_t^s, \mathbf{w}_t^s)$ via

$$(\rho_t^s \circ \mathrm{X}_t^s)\cdot \mathfrak{X}_t^s = \rho_t^\varepsilon, \qquad\qquad (\mathbf{w}^s \circ \mathrm{X}_t^s)\cdot \mathfrak{X}_t^s = \mathrm{D}\mathrm{X}_t^s \cdot \mathbf{w}_t^\varepsilon + \rho_t^\varepsilon \cdot \partial_t \mathrm{X}_t^s\,. \tag{6.2.3}$$

This representation follows from the following calculations: Assume that our perturbation in the density variable is given by $\rho_t^s := (\mathrm{X}_t^s)_\# \rho_t^\varepsilon$. Furthermore, given an arbitrary test function $\varphi_t \in \mathcal{C}_c^\infty((0,\infty)\times\overline{\Omega})$. Then, insert in the weak formulation of the continuity equation for the density-flux pair $(\rho_t^\varepsilon, \mathbf{w}_t^\varepsilon)$ the new test function $\psi_t^s := \varphi_t \circ \mathrm{X}_t^s$ and expand every term containing φ_t:

$$\begin{aligned} 0 &= \int_0^\infty \int_\Omega \partial_t(\varphi_t \circ \mathrm{X}_t^s)\cdot\rho_t^\varepsilon + \langle \nabla(\varphi_t\circ \mathrm{X}_t^s), \mathbf{w}_t^\varepsilon\rangle \,\mathrm{d}x\,\mathrm{d}t \\ &= \int_0^\infty \int_\Omega \big[(\partial_t\varphi_t)\circ \mathrm{X}_t^s) + \langle(\nabla\varphi_t)\circ \mathrm{X}_t^s, \partial_t \mathrm{X}_t^s\rangle\big]\rho_t^\varepsilon + \langle (\mathrm{DX}_t^s)^\top\cdot(\nabla(\varphi_t)\circ \mathrm{X}_t^s), \mathbf{w}_t^\varepsilon\rangle\,\mathrm{d}x\,\mathrm{d}t. \end{aligned}$$

A change of variables $x = \mathrm{X}_t^s(y)$ yields

$$0 = \int_0^\infty \int_\Omega \partial_t\varphi_t\,\rho_t^s + \langle \nabla\varphi_t, \big[\partial_t\mathrm{X}_t^s\rho_t^\varepsilon + \mathrm{DX}_t^s\cdot\mathbf{w}_t^\varepsilon\big]\circ(\mathrm{X}_t^s)^{-1}\rangle \det(\mathrm{D}(\mathrm{X}_t^s)^{-1})\,\mathrm{d}x\,\mathrm{d}t.$$

Hence, if the perturbed density-flux pair $(\rho_t^s, \mathbf{w}_t^s)$ shall solve the continuity equation, the perturbed flux $\mathbf{w}_t^s$ shall obey the representation

$$\mathbf{w}_t^s = \big[\partial_t\mathrm{X}_t^s\rho_t^\varepsilon + \mathrm{DX}_t^s\cdot\mathbf{w}_t^\varepsilon\big]\circ(\mathrm{X}_t^s)^{-1}\det(\mathrm{D}(\mathrm{X}_t^s)^{-1})$$

which is equivalent to (6.2.3). Indeed, as long as $(\rho_t^\varepsilon, \mathbf{w}_t^\varepsilon)$ solves the continuity equation also $(\rho_t^s, \mathbf{w}_t^s)$ solves the continuity equation. Since $\xi_0 = 0$, one has $\rho_0^s = \rho_0$ and we can conclude that $(\rho_t^s, \mathbf{w}_t^s) \in \mathcal{C}(\rho_0)$ is a feasible competitor for $(\rho_t^\varepsilon, \mathbf{w}_t^\varepsilon)$.

L^2-subdifferential of the Kinetic Energy $\mathcal{K}$. Before we prove the Euler-Lagrange equation of the WED-approximation $(\rho_t^\varepsilon, \mathbf{w}_t^\varepsilon)$, we compute the first variation of the kinetic energy $\mathcal{K}$ along the solutions of the Transport equation (6.2.2).

Lemma 6.2.2. *Given a density-flux pair $(\rho_t, \mathbf{w}_t) \in \mathcal{C}(\rho_0)$ such that $\rho_t \in \mathcal{P}_2^{ac}(\Omega)$ for each t and such that the map $t \mapsto e^{-t/\varepsilon}\mathcal{K}(\rho_t, \mathbf{w}_t) \in \mathrm{L}^1(0,\infty)$. For each vector field $\xi_t \in \mathcal{C}_c^\infty([0,\infty)\times\overline{\Omega})$ with $\xi_0 = 0$ and $\xi_t\cdot\boldsymbol{n} = 0$ on $\partial\Omega$ define the perturbation $(\rho_t^s, \mathbf{w}_t^s)$ via (6.2.3). Then,*

$$\frac{\mathrm{d}}{\mathrm{d}s}\Big[\int_0^\infty \frac{e^{-t/\varepsilon}}{2}\mathcal{K}(\rho_t^s, \mathbf{w}_t^s)\,\mathrm{d}t\Big]_{s=0} = \int_0^\infty e^{-t/\varepsilon}\int_\Omega \frac{\langle \mathbf{w}_t, \mathrm{D}\xi_t\mathbf{w}_t\rangle}{\rho_t} + \langle\partial_t\xi_t, \mathbf{w}_t\rangle\,\mathrm{d}x\,\mathrm{d}t. \quad (6.2.4)$$

Note, the integrand is well-defined for a.e. (t,x) since $t\mapsto e^{-t/\varepsilon}\mathcal{K}(\rho_t,\mathbf{w}_t) \in \mathrm{L}^1(0,\infty)$.

Proof. Fix ξ_t and $t > 0$. Using the explicit representation formula (6.2.3) and the change of variables $x = \mathrm{X}_t^s(y)$ we can rewrite $\mathcal{K}(\rho_t^s, \mathbf{w}_t^s)$ as follows

$$\mathcal{K}(\rho_t^s, \mathbf{w}_t^s) = \int_\Omega \frac{\|\mathrm{DX}_t^s\mathbf{w}_t\|^2}{\rho_t}\,\mathrm{d}y + 2\langle \mathrm{DX}_t^s\mathbf{w}_t, \partial_t\mathrm{X}_t^s\rangle + \rho_t\,\|\partial_t\mathrm{X}_t^s\|^2\,\mathrm{d}y.$$

Taking into account that X_t^0 is the identity, $\rho_t \in \mathcal{P}_2^{\mathrm{ac}}(\Omega)$, and $t\mapsto e^{-t/\varepsilon}\mathcal{K}(\rho_t,\mathbf{w}_t)$ is locally integrable, it follows by a dominated convergence argument

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}s}\Big|_{s=0}\int_0^\infty \frac{e^{-t/\varepsilon}}{2}\mathcal{K}(\rho_t^s,\mathbf{w}_t^s)\,\mathrm{d}t &= \int_0^\infty \frac{e^{-t/\varepsilon}}{2}\int_\Omega \frac{\mathrm{d}}{\mathrm{d}s}\Big[\frac{\|\mathrm{DX}_t^s\mathbf{w}_t\|^2}{\rho_t}\Big]_{s=0}\mathrm{d}y\,\mathrm{d}t \\ &+ \int_0^\infty \frac{e^{-t/\varepsilon}}{2}\int_\Omega \frac{\mathrm{d}}{\mathrm{d}s}\Big[2\langle\mathrm{DX}_t^s\mathbf{w}_t, \partial_t\mathrm{X}_t^s\rangle\Big]_{s=0}\mathrm{d}y\,\mathrm{d}t \\ &+ \int_0^\infty \frac{e^{-t/\varepsilon}}{2}\int_\Omega \frac{\mathrm{d}}{\mathrm{d}s}\Big[\rho_t\,\|\partial_t\mathrm{X}_t^s\|^2\Big]_{s=0}\mathrm{d}y\,\mathrm{d}t. \end{aligned}$$

In order to evaluate the s-derivative at $s = 0$, we make use of the following identities

$$\mathrm{DX}_t^0 = \mathbb{1}_d, \qquad \frac{\mathrm{d}}{\mathrm{d}s}\big[\mathrm{DX}_t^s\big]_{s=0} = \mathrm{D}\xi_t, \qquad \partial_t \mathrm{X}_t^0 = 0, \quad \frac{\mathrm{d}}{\mathrm{d}s}\big[\partial_t \mathrm{X}_t^s\big]_{s=0} = \partial_t \xi_t. \tag{6.2.5}$$

Hence, using these identities 6.2.5 the different derivatives are as follows: The derivative of the first integrand is given by

$$\frac{\mathrm{d}}{\mathrm{d}s}\Big[\frac{\|\mathrm{DX}_t^s \mathbf{w}_t\|^2}{\rho_t}\Big]_{s=0} = \frac{2\langle \mathrm{DX}_t^0 \mathbf{w}_t, \frac{\mathrm{d}}{\mathrm{d}s}\big[\mathrm{DX}_t^s\big]_{s=0}\mathbf{w}_t\rangle}{\rho_t} = \frac{\langle \mathbf{w}_t, \mathrm{D}\xi_t \mathbf{w}_t\rangle}{\rho_t}.$$

The derivative of the second integrand is given by

$$\frac{\mathrm{d}}{\mathrm{d}s}\Big[2\langle \mathrm{DX}_t^s \mathbf{w}_t, \partial_t \mathrm{X}_t^s\rangle\Big]_{s=0} = 2\langle \frac{\mathrm{d}}{\mathrm{d}s}[\mathrm{DX}_t^s]_{s=0}\mathbf{w}_t, \partial_t \mathrm{X}_t^0\rangle + 2\langle \mathrm{DX}_t^0 \mathbf{w}_t, \frac{\mathrm{d}}{\mathrm{d}s}[\partial_t \mathrm{X}_t^s]_{s=0}\rangle = \langle \mathbf{w}_t, \partial_t \xi_t\rangle.$$

Lastly, the derivative of the third integrand is given by

$$\frac{\mathrm{d}}{\mathrm{d}s}\Big[\rho_t \,\|\partial_t \mathrm{X}_t^s\|^2\Big]_{s=0} = 2\rho_t \langle \partial_t \mathrm{X}_t^0, \frac{\mathrm{d}}{\mathrm{d}s}[\partial_t \mathrm{X}_t^s]_{s=0}\rangle = 0.$$

Inserting these pointwise derivatives yields the desired result (6.2.4).

□

Time-continuous Euler-Lagrange Equations. Having the first variation of the kinetic energy $\mathcal{K}$ along the Transport equation at hand, we are able to proof the time-continuous Euler-Lagrange equations

Theorem 6.2.3 (Time-continuous Euler-Lagrange Equations)**.** *Let $\varepsilon \in (0, \varepsilon_*)$. Then, the WED-approximation $(\rho_t^\varepsilon, \mathbf{w}_t^\varepsilon)$ satisfies for each test function $\varphi_t \in \mathcal{C}_c^\infty([0,\infty) \times \overline{\Omega})$ with $\nabla\varphi_t \cdot \boldsymbol{n} = 0$ on $\partial\Omega$:*

$$\begin{aligned} 0 = &\int_0^\infty \int_\Omega \langle \nabla\varphi_t, \mathbf{w}_t^\varepsilon\rangle \,\mathrm{d}x\,\mathrm{d}t \\ &+ \varepsilon \int_0^\infty (1 - e^{-t/\varepsilon}) \int_\Omega \frac{\langle \mathbf{w}_t^\varepsilon, \mathrm{Hess}\,\varphi_t \mathbf{w}_t^\varepsilon\rangle}{\rho_\varepsilon^t} \,\mathrm{d}x\,\mathrm{d}t \\ &+ \varepsilon \int_0^\infty (1 - e^{-t/\varepsilon}) \int_\Omega \langle \partial_t \nabla\varphi_t, \mathbf{w}_t^\varepsilon\rangle \,\mathrm{d}x\,\mathrm{d}t \\ &- \int_0^\infty (1 - e^{-t/\varepsilon}) \int_\Omega \Delta\varphi_t \rho_t^\varepsilon - \langle \nabla\varphi_t, \nabla V + \nabla W * \rho_t^\varepsilon\rangle \rho_t^\varepsilon \,\mathrm{d}x\,\mathrm{d}t. \end{aligned} \tag{6.2.6}$$

Proof. Given the WED-approximation $(\rho_t^\varepsilon, \mathbf{w}_t^\varepsilon)$ define the perturbation $(\rho_t^s, \mathbf{w}_t^s)$ via (6.2.3) for a smooth vector field $\xi_t \in \mathcal{C}_c^\infty([0,\infty) \times \overline{\Omega}, \mathbb{R}^\mathrm{d})$ with $\xi_0 = 0$ and $\xi_t \cdot \boldsymbol{n} = 0$ on $\partial\Omega$. Due to the minimality property of $(\rho_t^\varepsilon, \mathbf{w}_t^\varepsilon)$ and since $(\rho_t^s, \mathbf{w}_t^s) \in \mathcal{C}(\rho_0)$, one has:

$$0 \le \frac{\mathrm{d}}{\mathrm{d}s}\big[\Psi_\varepsilon(\rho_t^s, \mathbf{w}_t^s)\big]_{s=0} = \frac{\mathrm{d}}{\mathrm{d}s}\Big[\int_0^\infty \frac{e^{-t/\varepsilon}}{\varepsilon}\big(\frac{\varepsilon}{2}\mathcal{K}(\rho_t^s, \mathbf{w}_t^s) + \mathcal{E}(\rho_t^s)\big)\,\mathrm{d}t\Big]_{s=0}. \tag{6.2.7}$$

The time-continuous Euler-Lagrange equation (6.2.6) follows by evaluating the s-derivatives on the right-hand side of (6.2.7). To actual compute the first variation of the free energy function $\mathcal{E}$ we proceed as in the time-discrete theory, cf. [4, Chapter 10.]. The difference quotient of the part containing the Boltzman entropy can be simplified using the explicit representation of ρ_t^s and the change of variables $x = \mathrm{X}_t^s(y)$:

$$\int_0^\infty \frac{e^{-t/\varepsilon}}{\varepsilon}\frac{1}{s}\big(\mathcal{H}(\rho_t^s) - \mathcal{H}(\rho_t^\varepsilon)\big)\,\mathrm{d}t = -\int_0^\infty \int_\Omega \frac{e^{-t/\varepsilon}}{\varepsilon}\log(\det(\mathrm{DX}_t^s))\rho_t^\varepsilon \,\mathrm{d}x\,\mathrm{d}t.$$

The pointwise limit of the integrand is given by (6.2.5). Since ξ_t has only compact support in time and space, $\mathrm{X}_t^s(x) = x$ outside this set. Hence, the effective domain of the integral above is compact. Further, the Jacobian of the flow map DX_t^s depends continuously on (s,t,x) and therefore we can conclude with a dominated convergence argument

$$\frac{\mathrm{d}}{\mathrm{d}s}\Big|_{s=0}\int_0^\infty \frac{e^{-t/\varepsilon}}{\varepsilon}\mathcal{H}(\rho_t^s)\,\mathrm{d}t = -\int_0^\infty \frac{e^{-t/\varepsilon}}{\varepsilon}\int_\Omega \mathrm{div}(\xi_t)\,\rho_t\,\mathrm{d}x\,\mathrm{d}t.$$

Similarly, one can compute the first variation of the confinement energy $\mathcal{V}$ and of the interaction energy $\mathcal{W}$ such that the first variation of the free energy functional $\mathcal{E}$ along the flow ρ_t^s amounts to

$$\frac{\mathrm{d}}{\mathrm{d}s}\Big[\int_0^\infty \frac{e^{-t/\varepsilon}}{\varepsilon}\mathcal{E}(\rho_t^s)\,\mathrm{d}t\Big]_{s=0} = -\int_0^\infty \frac{e^{-t/\varepsilon}}{\varepsilon}\int_\Omega \mathrm{div}(\xi_t)\rho_t^\varepsilon - \langle \xi_t, \nabla V + \nabla W * \rho_t^\varepsilon\rangle\rho_t^\varepsilon\,\mathrm{d}x\,\mathrm{d}t.$$

Due to (6.1.10) the map $t \mapsto e^{-t/\varepsilon}\mathcal{K}(\rho_t^\varepsilon, \mathbf{w}_t^\varepsilon)$ is locally integrable and we can apply Lemma 6.2.2 to get

$$\frac{\mathrm{d}}{\mathrm{d}s}\Big[\int_0^\infty \frac{e^{-t/\varepsilon}}{2}\mathcal{K}(\rho_t^s, \mathbf{w}_t^s)\,\mathrm{d}t\Big]_{s=0} = \int_0^\infty e^{-t/\varepsilon}\int_\Omega \frac{\langle \mathbf{w}_t, \mathrm{D}\xi_t\mathbf{w}_t\rangle}{\rho_t} + \langle \partial_t\xi_t, \mathbf{w}_t\rangle\,\mathrm{d}x\,\mathrm{d}t.$$

Inserting these two equations in (6.2.7) yields

$$\begin{aligned} 0 \le &\int_0^\infty e^{-t/\varepsilon}\int_\Omega \frac{\langle \mathbf{w}_t^\varepsilon, \mathrm{D}\xi_t\mathbf{w}_t^\varepsilon\rangle}{\rho_t^\varepsilon} + \langle \partial_t\xi_t, \mathbf{w}_t^\varepsilon\rangle\,\mathrm{d}x\,\mathrm{d}t \\ &- \int_0^\infty \frac{e^{-t/\varepsilon}}{\varepsilon}\int_\Omega \mathrm{div}(\xi_t)\rho_t^\varepsilon - \langle \xi_t, \nabla V + \nabla W * \rho_t^\varepsilon\rangle\rho_t^\varepsilon\,\mathrm{d}x\,\mathrm{d}t. \end{aligned}$$

Since the equation above is linear with respect to ξ_t, repeating the calculations above for $-\xi_t$ in place of ξ_t yields the converse inequality and hence we have equality. Lastly, we choose as specific test functions $\xi_t = \varepsilon(e^{t/\varepsilon} - 1)\nabla\varphi_t$ where $\varphi_t \in \mathcal{C}_c^\infty([0,\infty)\times\overline{\Omega})$ with $\nabla\varphi_t \cdot \boldsymbol{n} = 0$ on $\partial\Omega$ which yields the desired time-continuous Euler-Lagrange equations (6.2.6). □

6.2.3 Better A Priori Bounds by Continuous Variations

The ε-independent bounds (6.1.10) are not sufficient to pass in the Euler-Lagrange equations (6.2.6) to the limit $\varepsilon \to 0$. Inspired by the flow interchange technique [68] to deduce better a priori bounds in the time-discrete JKO-method, we use suitable perturbations of the minimizers to improve the ε-independent bounds. In particular, we consider as *time-continuous* perturbation ρ_t^s of the approximated solution ρ_t^ε the solution to the heat equation

$$\partial_s \rho_t^s = \eta_t \Delta(\rho_t^s), \qquad \rho_t^0 = \rho_t^\varepsilon \tag{6.2.8}$$

with no-flux boundary condition and for some given time-dependent diffusivity parameter $\eta_t \in C_c^\infty(0,\infty)$ with $\eta_0 = 0$ and $\partial_t \eta_0 = 0$. By the parabolic regularity theory, it is clear, that for the heat equation with diffusivity parameter η_t and no-flux boundary condition there exists a smooth and non-negative solution ρ_t^s. To have a feasible density-flux pair $(\rho_t^s, \mathbf{w}_t^s)$ we define the perturbed flux $\mathbf{w}_t^s$ as the solution to the inhomogeneous vectorial heat equation

$$\partial_s \mathbf{w}_t^s = \eta_t \Delta(\mathbf{w}_t^s) - \partial_t \eta_t \nabla \rho_t^s, \qquad \mathbf{w}_t^0 = \mathbf{w}_t^\varepsilon \tag{6.2.9}$$

with Dirichlet boundary condition. Note, the vectorial heat equation (6.2.9) is meant to be understood componentwise. Furthermore, this equation is also well-posed for $t = 0$ since we assumed $\eta_0 = 0$ and $\partial_t \eta_0 = 0$. Again, it follows by the parabolic regularity theory, that for the inhomogeneous vectorial heat equation (6.2.9) with Dirichlet boundary condition there exists a smooth solution $\mathbf{w}_t^s$.

Motivation. Before we elaborate on the feasibility of the competitor $(\rho_t^s, \mathbf{w}_t^s)$, their explicit representations, and the actual proof of the refined a priori bounds, we motivate our approach with a heuristic calculation. By the regularizing effects of the heat equation, the map $s \mapsto \mathcal{K}(\rho_t^s, \mathbf{w}_t^s)$ is differentiable at each $s > 0$ with derivative given by

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}s} \mathcal{K}(\rho_t^s, \mathbf{w}_t^s) &= \int_\Omega \frac{2\langle \mathbf{w}_t^s, \partial_s \mathbf{w}_t^s \rangle}{\rho_t^s} - \frac{\|\mathbf{w}_t^s\|^2}{(\rho_t^s)^2} \partial_s \rho_t^s \, \mathrm{d}x \\ &= \int_\Omega 2\langle \frac{\mathbf{w}_t^s}{\rho_t^s}, \eta_t \Delta(\mathbf{w}_t^s) - \partial_t \eta_t \nabla \rho_t^s \rangle - \eta_t \Big\| \frac{\mathbf{w}_t^s}{\rho_t^s} \Big\|^2 \Delta \rho_t^s \, \mathrm{d}x. \end{aligned}$$

Write $\mathbf{v}_t^s = \mathbf{w}_t^s / \rho_t^s$ and apply Green's identity to simplify the first integrand as follows

$$\begin{aligned} \int_\Omega 2\langle \mathbf{v}_t^s, \eta_t \Delta(\mathbf{w}_t^s) \rangle \, \mathrm{d}x &= 2\eta_t \sum_{i=1}^d \int_\Omega (\mathbf{v}_t^s)_i \Delta(\mathbf{w}_t^s)_i \, \mathrm{d}x \\ &= 2\eta_t \sum_{i=1}^d \Big[\int_{\partial\Omega} \frac{1}{\rho_t^s} (\mathbf{w}_t^s)_i (\nabla(\mathbf{w}_t^s)_i \cdot \boldsymbol{n}) \, \mathrm{dS} - \int_\Omega \langle \nabla(\mathbf{v}_t^s)_i, \nabla(\mathbf{w}_t^s)_i \rangle \, \mathrm{d}x \Big]. \end{aligned}$$

Note, the surface integral is equal to zero if each component of the perturbed flux $\mathbf{w}_t^s$ satisfies either Dirichlet boundary conditions or no-flux boundary conditions. However, $\mathbf{w}_t^s$ has to satisfy also the boundary condition from the continuity equation, i.e. $\mathbf{w}_t^s \cdot \boldsymbol{n} = 0$ on $\partial\Omega$. Hence, this motivates the Dirichlet boundary conditions for $(\mathbf{w}_t^s)_i$ in (6.2.9).

Continuing the calculation, we simplify the right hand side further as follows

$$-2\eta_t \sum_{i=1}^{d} \int_\Omega \langle \nabla(\mathbf{v}_t^s)_i, \nabla(\mathbf{w}_t^s)_i \rangle \,\mathrm{d}x = -2\eta_t \sum_{i=1}^{d} \int_\Omega \left\| \nabla\big(\mathbf{v}_t^s\big)_i \right\|^2 \rho_t^s + \langle \nabla\big(\mathbf{v}_t^s\big)_i, \big(\mathbf{v}_t^s\big)_i \nabla \rho_t^s \rangle \,\mathrm{d}x$$
$$= -2\eta_t \int_\Omega \left\| \mathrm{D}\mathbf{v}_t^s \right\|_{\mathrm{HS}}^2 \rho_t^s + \frac{1}{2} \langle \nabla \left\| \mathbf{v}_t^s \right\|^2, \nabla \rho_t^s \rangle \,\mathrm{d}x.$$

Similarly, we can simplify the third integrand with Green's identity to obtain

$$-\eta_t \int_\Omega \left\| \mathbf{v}_t^s \right\|^2 \Delta \rho_t^s \,\mathrm{d}x = -\eta_t \int_{\partial\Omega} \left\| \mathbf{v}_t^s \right\|^2 (\nabla \rho_t^s \cdot \boldsymbol{n}) \,\mathrm{dS} + \eta_t \int_\Omega \langle \nabla \left\| \mathbf{v}_t^s \right\|^2, \nabla \rho_t^s \rangle \,\mathrm{d}x.$$

Since ρ_t^s satisfies the no-flux boundary condition, the surface integral vanishes. In conclusion, we get as variation of the kinetic energy $\mathcal{K}$:

$$\frac{\mathrm{d}}{\mathrm{d}s} \mathcal{K}(\rho_t^s, \mathbf{w}_t^s) = -\eta_t \int_\Omega \rho_t^s \left\| \mathrm{D}\mathbf{v}_t^s \right\|_{\mathrm{HS}}^2 \,\mathrm{d}x - 2\partial_t \eta_t \int_\Omega \langle \mathbf{w}_t^s, \nabla \log(\rho_t^s) \rangle \,\mathrm{d}x. \tag{6.2.10}$$

The first term is always negative and the second integral is formally equal to the time derivative of the map $t \mapsto \mathcal{H}(\rho_t^s)$. Integrating (6.2.10) with respect to t and taking the first variation of the free energy functional $\mathcal{E}$ and the minimality of $(\rho_t^\varepsilon, \mathbf{w}_t^\varepsilon)$ into account, we end up with the following estimate:

$$0 \le \frac{\mathrm{d}}{\mathrm{d}s} [\Psi_\varepsilon(\rho_t^s, \mathbf{w}_t^s)]_{s=0} \le \int_0^\infty \frac{e^{-t/\varepsilon}}{\varepsilon} \Big[-2\varepsilon \partial_t \eta_t \frac{\mathrm{d}}{\mathrm{d}t} \mathcal{H}(\rho_t^\varepsilon) + \frac{\mathrm{d}}{\mathrm{d}s} [\mathcal{E}(\rho_t^s)]_{s=0} \Big] \,\mathrm{d}t.$$

We emphasize that this is exactly the time-continuous analogue of the time-discrete flow interchange estimate. However, without having additional information on the regularity of $t \mapsto \rho_t^\varepsilon$ this computation and the passage to the limit $s \to 0$ is a priori not justified.

With another approach, one could also estimate the second term in (6.2.10) with Hölders inequality by

$$\frac{\mathrm{d}}{\mathrm{d}s} \mathcal{K}(\rho_t^s, \mathbf{w}_t^s) \le 2 \left| \partial_t \eta_t \right| \sqrt{\mathcal{K}(\rho_t^s, \mathbf{w}_t^s)} \sqrt{\mathcal{I}(\rho_t^s)}.$$

Since the map $s \mapsto \mathcal{I}(\rho_t^s)$ is monotonically decreasing, information about the differentiability of $\mathcal{K}$ at $s = 0$ can be only retrieved if the Fisher-Information $\mathcal{I}(\rho_t^\varepsilon)$ is finite. To circumvent this issue, we exploit the following observation: Computing simultaneously the first variation of $\mathcal{K}$ and the Boltzmann entropy $\mathcal{H}$ yields

$$\frac{\mathrm{d}}{\mathrm{d}s} \left[\frac{\varepsilon}{2} \mathcal{K}(\rho_t^s, \mathbf{w}_t^s) + \mathcal{H}(\rho_t^s) \right] \le \varepsilon \left| \partial_t \eta_t \right| \sqrt{\mathcal{K}(\rho_t^s, \mathbf{w}_t^s)} \sqrt{\mathcal{I}(\rho_t^s)} - \eta_t \mathcal{I}(\rho_t^s).$$

Optimizing over $\mathcal{I}(\rho_t^s)$, i.e., using the elementary inequality $a\sqrt{x} - bx \le \frac{a^2}{4b}$ with $a = \varepsilon \left| \partial_t \eta_t \right| \sqrt{\mathcal{K}(\rho_t^s, \mathbf{w}_t^s)}$, $b = \frac{1}{2}\eta_t$, and $x = \mathcal{I}(\rho_t^s)$, yields

$$\frac{\mathrm{d}}{\mathrm{d}s} \left[\frac{\varepsilon}{2} \mathcal{K}(\rho_t^s, \mathbf{w}_t^s) + \mathcal{H}(\rho_t^s) \right] \le \varepsilon^2 \frac{\left| \partial_t \eta_t \right|^2}{\eta_t} \mathcal{K}(\rho_t^s, \mathbf{w}_t^s) - \frac{1}{2} \eta_t \mathcal{I}(\rho_t^s). \tag{6.2.11}$$

Hence, as long quotient $(\partial_t \eta_t)^2 / \eta_t$ is bounded from above by positive constant C we can derive a differentiability result for the map $s \mapsto \frac{\varepsilon}{2} \mathcal{K}(\rho_t^s, \mathbf{w}_t^s) + \mathcal{H}(\rho_t^s)$ at $s = 0$ by means of a Gronwall type argument. However, at this point point we have the issue of the continuity of the map $s \mapsto \mathcal{K}(\rho_t^s, \mathbf{w}_t^s)$ which is a priori not clear.

Definition and Justification of the Perturbation Prior to the justification of the feasibility of the perturbation $(\rho_t^s, \mathbf{w}_t^s)$, we define a specific diffusivity parameter η_t via

$$\eta_t = \begin{cases} \varepsilon(e^{t/2\varepsilon} - 1)^2 & \text{for } t \in [0, T], \\ -\varepsilon(e^{(2T-t)/2\varepsilon} - 1)^2 + 2\varepsilon(e^{T/2\varepsilon} - 1)^2 & \text{for } t \in [T, 2T], \\ -\varepsilon(e^{(t-2T)/2\varepsilon} - 1)^2 + 2\varepsilon(e^{T/2\varepsilon} - 1)^2 & \text{for } t \in [2T, 3T], \\ \varepsilon(e^{(4T-t)/2\varepsilon} - 1)^2 & \text{for } t \in [3T, 4T], \\ 0 & \text{for } t \in [4T, \infty). \end{cases} \tag{6.2.12}$$

With this definition, the diffusivity parameter η_t is non-negative, with compact support, continuously differentiable, with $\eta_0 = 0$, and with $\partial_t \eta_0 = 0$. Further more, we have the following growth bounds on η_t and $\partial_t \eta_t$:

$$\varepsilon e^{-t/\varepsilon}(\partial_t \eta_t)^2 \leq \eta_t \leq \varepsilon e^{t/\varepsilon} \quad \forall t \in [0, 4T]. \tag{6.2.13}$$

Hence, by the previous remark η_t is suitable in (6.2.11).

The next lemma shows the feasibility of the perturbation $(\rho_t^s, \mathbf{w}_t^s)$.

Lemma 6.2.4 (Feasible Competitor)**.** *Given a time-dependent diffusivity parameter* $\eta_t \in C_c^\infty(0, \infty)$ *with* $\eta_0 = 0$ *and* $\partial_t \eta_0 = 0$. *Define the perturbed density-flux pair* $(\rho_t^s, \mathbf{w}_t^s)$ *viz* (6.2.8) *and* (6.2.9), *respectively. Then,* $(\rho_t^s, \mathbf{w}_t^s) \in \mathcal{C}(\rho_0)$ *for each* $s \geq 0$.

Proof. To check, that this perturbation $(\rho_t^s, \mathbf{w}_t^s)$ is a feasible competitor for $(\rho_t^\varepsilon, \mathbf{w}_t^\varepsilon)$ we have to check three properties:

1. **Initial condition.** Due to assumption $\eta_0 = 0$ and $\partial_t \eta_0 = 0$ we have $(\rho_0^s, \mathbf{w}_0^s) = (\rho_0, \mathbf{w}_0^\varepsilon)$ and therefor the density-flux pair $(\rho_t^s, \mathbf{w}_t^s)$ satisfies the initial condition.

2. **No-flux Boundary Condition.** Since $\mathbf{w}_t^s$ satisfies the Dirichlet boundary condition in each component, i.e., $(\mathbf{w}_t^s)_i = 0$ on $\partial\Omega$ for each $s > 0$ and every $i \in \{1, \ldots, d\}$, we trivially satisfy the no-flux boundary condition of the continuity equation, i.e., $\mathbf{w}_t^s \cdot \boldsymbol{n} = 0$ on $\partial\Omega$ for each $s \geq 0$.

3. **Continuity Equation.** To verify that the density-flux pair $(\rho_t^s, \mathbf{w}_t^s)$ satisfy the continuity equation, we compute the partial s-derivative of the map $(s, x) \mapsto \partial_t \rho_t^s + \operatorname{div} \mathbf{w}_t^s$ for each fixed t:

$$\partial_s(\partial_t \rho_t^s + \operatorname{div} \mathbf{w}_t^s) = \partial_t(\eta_t \Delta \rho_t^s) + \operatorname{div}(\eta_t \Delta \mathbf{w}_t^s - \partial_t \eta_t \nabla \rho_t^s) = \eta_t \Delta(\partial_t \rho_t^s + \operatorname{div} \mathbf{w}_t^s).$$

   Hence, the map $(s, x) \mapsto \partial_t \rho_t^s + \operatorname{div} \mathbf{w}_t^s$ solves the heat equation with initial datum $\partial_t \rho_t^0 + \operatorname{div} \mathbf{w}_t^0 = \partial_t \rho_t^\varepsilon + \operatorname{div} \mathbf{w}_t^\varepsilon$ Indeed, as long as $(\rho_t^\varepsilon, \mathbf{w}_t^\varepsilon)$ solves the continuity equation also $(\rho_t^s, \mathbf{w}_t^s)$ solves the continuity equation. □

Explicit Representation. If $\Omega = \mathbb{R}^{\mathrm{d}}$ is the whole space there is an explicit representation for the perturbed density-flux pair $(\rho_t^s, \mathbf{w}_t^s)$ in terms of the corresponding Greens function $\mathfrak{G}_t^s$ of the heat equation (6.2.8) which is given by

$$\mathfrak{G}_t^s(x,y) = (4\pi\eta_t s)^{-d/2} \exp\Big[-\frac{\|x-y\|^2}{4\eta_t s} \Big].$$

With the Greens function $\mathfrak{G}_t^s$ at hand we define the perturbation ρ_t^s of the density ρ_t^ε by $\rho_t^s := \mathcal{T}_{\mathfrak{G}_t^s}[\rho_t^\varepsilon]$. The explicit representation formula for $\mathbf{w}_t^s$ is given by Duhamel's principle:

$$\mathbf{w}_t^s = \vec{\mathcal{T}}_{\mathfrak{G}_t^s}[\mathbf{w}_t^\varepsilon] - \partial_t\eta_t \int_0^s \mathcal{T}_{\mathfrak{G}_t^{s-r}}[\nabla \mathcal{T}_{\mathfrak{G}_t^r}[\rho_t^\varepsilon]]\,\mathrm{d}r = \vec{\mathcal{T}}_{\mathfrak{G}_t^s}[\mathbf{w}_t^\varepsilon] - \partial_t\eta_t \nabla \int_0^s \mathcal{T}_{\mathfrak{G}_t^{s-r}}[\mathcal{T}_{\mathfrak{G}_t^r}[\rho_t^\varepsilon]]\,\mathrm{d}r.$$

Thanks to the semigroup property of the heat kernel we can simplify the second term:

$$\int_0^s \mathcal{T}_{\mathfrak{G}_t^{s-r}}[\mathcal{T}_{\mathfrak{G}_t^r}[\rho_t^\varepsilon]]\,\mathrm{d}r = \int_0^s \mathcal{T}_{\mathfrak{G}_t^s}[\rho_t^\varepsilon]\,\mathrm{d}r = s\mathcal{T}_{\mathfrak{G}_t^s}[\rho_t^\varepsilon].$$

In conclusion, we end up with the explicit representation formulas

$$\rho_t^s := \mathcal{T}_{\mathfrak{G}_t^s}[\rho_t^\varepsilon] \qquad \text{and} \qquad \mathbf{w}_t^s := \vec{\mathcal{T}}_{\mathfrak{G}_t^s}[\mathbf{w}_t^\varepsilon] - s\partial_t\eta_t\nabla\mathcal{T}_{\mathfrak{G}_t^s}[\rho_t^\varepsilon]. \tag{6.2.14}$$

In the case $\Omega \subset \mathbb{R}^{\mathrm{d}}$ is an open, bounded and convex set we don't have such an explicit representation as (6.2.14). The reason is as follows. Let $\mathfrak{D}_t^s$ and $\mathfrak{N}_t^s$ be the greens functions of the heat equation on Ω with Dirichlet boundary conditions or no-flux boundary conditions, respectively. Then, ρ_t^s and $\mathbf{w}_t^s$ are given by

$$\rho_t^s = \mathcal{T}_{\mathfrak{N}_t^s}[\rho_t^\varepsilon], \qquad \text{and} \qquad \mathbf{w}_t^s = \mathcal{T}_{\mathfrak{D}_t^s}[\rho_t^\varepsilon] - \partial_t\eta_t\nabla \int_0^s \mathcal{T}_{\mathfrak{D}_t^{s-r}}[\mathcal{T}_{\mathfrak{N}_t^r}[\rho_t^\varepsilon]]\,\mathrm{d}s.$$

In this case, we cannot invoke the semigroup property of the two different heat kernels and therefor we cannot end up with representation formulas as (6.2.14). One might suggest to require no-flux boundary conditions in the heat equation (6.2.9) such that we could exploit the semi-group property of the heat kernel $\mathfrak{N}_t^s$ and end up with

$$\rho_t^s := \mathcal{T}_{\mathfrak{N}_t^s}[\rho_t^\varepsilon] \qquad \text{and} \qquad \mathbf{w}_t^s := \vec{\mathcal{T}}_{\mathfrak{N}_t^s}[\mathbf{w}_t^\varepsilon] - s\partial_t\eta_t\nabla\mathcal{T}_{\mathfrak{N}_t^s}[\rho_t^\varepsilon].$$

Still, the motivating calculation holds true and one can derive the differential inequality (6.2.10). Nevertheless, with no-flux boundary conditions for (6.2.9) it is not clear, that the perturbed flux $\mathbf{w}_t^s$ satisfy the boundary condition $\mathbf{w}_t^s \cdot \boldsymbol{n} = 0$ on $\partial\Omega$ of the continuity equation. Since $\mathcal{T}_{\mathfrak{N}_t^s}[\rho_t^\varepsilon]$ satisfies the no-flux boundary condition, a simple calculation shows $\mathbf{w}_t^s \cdot \boldsymbol{n} = \vec{\mathcal{T}}_{\mathfrak{N}_t^s}[\mathbf{w}_t^\varepsilon] \cdot \boldsymbol{n}$ on $\partial\Omega$. However, a priori we don't have any information on $\vec{\mathcal{T}}_{\mathfrak{N}_t^s}[\mathbf{w}_t^\varepsilon] \cdot \boldsymbol{n}$ on $\partial\Omega$ except that $\mathrm{D}\vec{\mathcal{T}}_{\mathfrak{N}_t^s}[\mathbf{w}_t^\varepsilon] \cdot \boldsymbol{n} = 0$ on $\partial\Omega$. So this approach is not feasible for our purpose. Another approach is to define $\mathbf{w}_t^s$ directly by

$$\mathbf{w}_t^s := \vec{\mathcal{T}}_{\mathfrak{D}_t^s}[\mathbf{w}_t^\varepsilon] - s\partial_t\eta_t\nabla\mathcal{T}_{\mathfrak{N}_t^s}[\rho_t^\varepsilon].$$

The by this formula defined perturbation $\mathbf{w}_t^s$ satisfies the heat equation though with no boundary condition. Still, $\mathbf{w}_t^s$ satisfies the boundary condition of the continuity equation. However, the motivating calculation is invalid and also our method in the proof of the refined a priori bounds does not work, since our method heavily rely on the monotonicity of $\mathcal{K}$ which requires the same integral kernel for ρ_t^s and $\mathbf{w}_t^s$, see lemma 2.5.6.

Refined A Priori Bounds.

Theorem 6.2.5 (Better A Priori Bounds). *Let $\varepsilon \in (0, \varepsilon_*)$. There exists non-negative constant C, which depends only on d_1, d_2, T and the initial datum ρ_0 such that the WED-approximation ρ_t^ε with respect to $\mathcal{E}$ emanating from $\rho_0 \in \mathcal{D}(\mathcal{E})$ satisfies*

$$\int_0^T (1 - e^{-t/2})^2 \int_\Omega \|\nabla(\sqrt{\rho_t^\varepsilon})\|^2 \,\mathrm{d}x \,\mathrm{d}t \le C. \tag{6.2.15}$$

Proof of Theorem 6.2.5. Fix $T > 0$. Given the WED-approximation $(\rho_t^\varepsilon, \mathbf{w}_t^\varepsilon)$ with respect to $\mathcal{E}$, let $(\rho_t^s, \mathbf{w}_t^s)$ be the perturbation defined by (6.2.14) with η_t given by (6.2.12). Due to the minimality property of $(\rho_t^\varepsilon, \mathbf{w}_t^\varepsilon)$ and due to $(\rho_t^s, \mathbf{w}_t^s) \in \mathcal{C}(\rho_0)$, we have

$$\begin{aligned} 0 \le &\int_0^\infty \frac{e^{-t/\varepsilon}}{2} \frac{1}{s} \left[\frac{\varepsilon}{2} \mathcal{K}(\rho_t^s, \mathbf{w}_t^s) + \mathcal{H}(\rho_t^s) - \frac{\varepsilon}{2} \mathcal{K}(\rho_t^\varepsilon, \mathbf{w}_t^\varepsilon) - \mathcal{H}(\rho_t^\varepsilon) \right] \mathrm{d}t \\ &+ \int_0^\infty \frac{e^{-t/\varepsilon}}{\varepsilon} \frac{1}{s} \left[\mathcal{V}(\rho_t^s) + \mathcal{W}(\rho_t^s) - \mathcal{V}(\rho_t^\varepsilon) - \mathcal{W}(\rho_t^\varepsilon) \right] \mathrm{d}t =: I_1 + I_2. \end{aligned} \tag{6.2.16}$$

for sufficiently small $s > 0$. By passing to the limit $s \to 0$ in each term of (6.2.16), we will obtain the better a priori bound (6.2.15).

First, we estimate I_1, i.e., the difference quotient of $\mathcal{K} + \mathcal{H}$ along the heat flow. With the representation (6.2.14) we obtain by Minkowski's inequality for the weighted integral with weight function $(1/\mathcal{T}_{\mathfrak{G}_t^s}[\rho_t^\varepsilon])$ a first estimate for $\sqrt{\mathcal{K}}$:

$$\begin{aligned} \sqrt{\mathcal{K}(\rho_t^s, \mathbf{w}_t^s)} =& \Big(\int_\Omega \left\| \vec{\mathcal{T}}_{\mathfrak{G}_t^s}[\mathbf{w}_t^\varepsilon] - s \partial_t \eta_t \nabla \mathcal{T}_{\mathfrak{G}_t^s}[\rho_t^\varepsilon] \right\|^2 \frac{1}{\mathcal{T}_{\mathfrak{G}_t^s}[\rho_t^\varepsilon]} \,\mathrm{d}x \Big)^{1/2} \\ \le& \Big(\int_\Omega \left\| \vec{\mathcal{T}}_{\mathfrak{G}_t^s}[\mathbf{w}_t^\varepsilon] \right\|^2 \frac{1}{\mathcal{T}_{\mathfrak{G}_t^s}[\rho_t^\varepsilon]} \,\mathrm{d}x \Big)^{1/2} + \Big(\int_\Omega \left\| s \partial_t \eta_t \nabla \mathcal{T}_{\mathfrak{G}_t^s}[\rho_t^\varepsilon] \right\|^2 \frac{1}{\mathcal{T}_{\mathfrak{G}_t^s}[\rho_t^\varepsilon]} \,\mathrm{d}x \Big)^{1/2} \\ =& \sqrt{\mathcal{K}\big(\mathcal{T}_{\mathfrak{G}_t^s}[\rho_t], \vec{\mathcal{T}}_{\mathfrak{G}_t^s}[\mathbf{w}_t]\big)} + s \left| \partial_t \eta_t \right| \sqrt{\mathcal{K}\big(\mathcal{T}_{\mathfrak{G}_t^s}[\rho_t^\varepsilon], \nabla \mathcal{T}_{\mathfrak{G}_t^s}[\rho_t^\varepsilon]\big)}. \end{aligned}$$

By classic parabolic regularity theory the Greens function $\mathfrak{G}_t^s$ satisfies the assumption of Lemma 2.5.6 and we can conclude

$$\sqrt{\mathcal{K}(\rho_t^s, \mathbf{w}_t^s)} \le \sqrt{\mathcal{K}(\rho_t, \mathbf{w}_t)} + s \left| \partial_t \eta_t \right| \sqrt{\mathcal{K}\left(\rho_t^s, \nabla \rho_t^s\right)} = \sqrt{\mathcal{K}(\rho_t, \mathbf{w}_t)} + 2s \left| \partial_t \eta_t \right| \sqrt{\mathcal{I}(\rho_t^s)},$$

where we used in the last step $\mathcal{K}\left(\rho, \nabla \rho\right) = 4\mathcal{I}(\rho)$. Taking the square yields

$$\frac{1}{s} \left[\mathcal{K}(\rho_t^s, \mathbf{w}_t^s) - \mathcal{K}(\rho_t, \mathbf{w}_t) \right] \le 4 \sqrt{\mathcal{K}(\rho_t, \mathbf{w}_t)} \left| \partial_t \eta_t \right| \sqrt{\mathcal{I}(\rho_t^s)} + 4s \left| \partial_t \eta_t \right|^2 \mathcal{I}(\rho_t^s). \tag{6.2.17}$$

The difference quotient of the entropy functional $\mathcal{H}$ can be estimated by using the properties of the heat flow, cf. [4, Chapter 11.]:

$$\frac{1}{s} \left[\mathcal{H}(\rho_t^s) - \mathcal{H}(\rho_t) \right] = -\eta_t \frac{1}{s} \int_0^s \mathcal{I}(\rho_t^r) \,\mathrm{d}r \le -\eta_t \mathcal{I}(\rho_t^s). \tag{6.2.18}$$

Combining (6.2.17) and (6.2.18) yields for the combined difference quotient of $\mathcal{K}+\mathcal{H}$:

$$\begin{aligned}&\frac{1}{s}\left[\frac{\varepsilon}{2}\mathcal{K}(\rho_t^s,\mathbf{w}_t^s)+\mathcal{H}(\rho_t^s)-\frac{\varepsilon}{2}\mathcal{K}(\rho_t,\mathbf{w}_t)+\mathcal{H}(\rho_t)\right]\\ \leq&2\varepsilon\sqrt{\mathcal{K}(\rho_t,\mathbf{w}_t)}\,|\partial_t\eta_t|\,\sqrt{\mathcal{I}(\rho_t^s)}+2s\varepsilon\,|\partial_t\eta_t|^2\,\mathcal{I}(\rho_t^s)-\eta_t\mathcal{I}(\rho_t^s).\end{aligned}$$

Using the elementary inequality $a\sqrt{x}-bx\leq\frac{a^2}{4b}$ with $a=2\varepsilon\sqrt{\mathcal{K}(\rho_t,\mathbf{w}_t)}\,|\partial_t\eta_t|$, $b=\frac{1}{2}\eta_t$, and $x=\mathcal{I}(\rho_t^s)$, yields for I_1:

$$\begin{aligned}I_1\leq&2\int_0^\infty\varepsilon e^{-t/\varepsilon}\frac{(\partial_t\eta_t)^2}{\eta_t}\mathcal{K}(\rho_t^\varepsilon,\mathbf{w}_t^\varepsilon)\,\mathrm{d}t-\frac{1}{2}\int_0^\infty\frac{e^{-t/\varepsilon}}{\varepsilon}\eta_t\mathcal{I}(\rho_t^s)\,\mathrm{d}t\\ &+2\int_0^\infty e^{-t/\varepsilon}(\partial_t\eta_t)^2 s\mathcal{I}(\rho_t^s)\,\mathrm{d}t=:I_3+I_4+I_5.\end{aligned}\tag{6.2.19}$$

We pass in (6.2.19) to the limit $s\to0$. The first integral I_3 on the right hand side of (6.2.19) is estimated as follows by (6.2.13) and (6.1.10)

$$I_3:=2\int_0^\infty\varepsilon e^{-t/\varepsilon}\frac{(\partial_t\eta_t)^2}{\eta_t}\mathcal{K}(\rho_t^\varepsilon,\mathbf{w}_t^\varepsilon)\,\mathrm{d}t\leq8\int_0^{4T}\mathcal{K}(\rho_t^\varepsilon,\mathbf{w}_t^\varepsilon)\,\mathrm{d}t\leq C.$$

The integrand of the second integral I_4 of the right-hand side of (6.2.19) is positive, so we estimate I_4 with (6.2.13) to get

$$I_4:=-\frac{1}{2}\int_0^\infty\frac{e^{-t/\varepsilon}}{\varepsilon}\eta_t\mathcal{I}(\rho_t^s)\,\mathrm{d}t\leq-\frac{1}{2}\int_0^T\frac{e^{-t/\varepsilon}}{\varepsilon}\eta_t\mathcal{I}(\rho_t^s)\,\mathrm{d}t=-\frac{1}{2}\int_0^T(1-e^{-t/2\varepsilon})^2\mathcal{I}(\rho_t^s)\,\mathrm{d}t.$$

By Fatou's Lemma and the lower semi-continuity of the Fisher-information $\mathcal{I}$ we get

$$\limsup_{s\searrow0}I_4=\limsup_{s\searrow0}-\frac{1}{2}\int_0^\infty\frac{e^{-t/\varepsilon}}{\varepsilon}\eta_t\mathcal{I}(\rho_t^s)\,\mathrm{d}t\leq-\frac{1}{2}\int_0^T(1-e^{-t/2})^2\mathcal{I}(\rho_t^\varepsilon)\,\mathrm{d}t.$$

The limit $s\to0$ of the third integral I_5 on the right-hand side of (6.2.19) can be derived as follows. By (6.2.18) the integrand is bounded from above by

$$I_5:=2\int_0^\infty e^{-t/\varepsilon}(\partial_t\eta_t)^2 s\mathcal{I}(\rho_t^s)\,\mathrm{d}t\leq2\int_0^{4T}e^{-t/\varepsilon}\frac{(\partial_t\eta_t)^2}{\eta_t}\left[\mathcal{H}(\rho_t^\varepsilon)-\mathcal{H}(\rho_t^s)\right]\mathrm{d}t.$$

Hence, by the lower semi-continuity of $\mathcal{H}$, the point-wise limit of the integrand of the third integral in (6.2.19) is less than zero. By the Carleman estimate, the entropy $\mathcal{H}$ is bounded from below in terms of the second moment $\boldsymbol{M}_2$. Furthermore, the dissipation of the second moment $\boldsymbol{M}_2$ along the heat flow takes place with rate $2\mathrm{d}\eta_t$. Hence,

$$-\mathcal{H}(\rho_t^s)\leq C(1+\boldsymbol{M}_2(\rho_t^s))=C(1+\boldsymbol{M}_2(\rho_t^\varepsilon)+s2\mathrm{d}\eta_t).$$

Note, due to the definition of the second moment and the dynamic formulation of the L^2-Wasserstein distance one has

$$\boldsymbol{M}_2(\rho_t^\varepsilon)\leq2\boldsymbol{M}_2^2(\rho_0)+2\mathbf{W}_2^2(\rho_0,\rho_t^\varepsilon)\leq2\boldsymbol{M}_2^2(\rho_0)+2t\int_0^t\mathcal{K}(\rho_r^\varepsilon,\mathbf{w}_r^\varepsilon)\,\mathrm{d}r.$$

Hence, it follows with (6.2.13) that the integrand of I_6 is bounded from above by a map, which is locally integrable due to the ε-independent bounds (6.1.10). So we can conclude with Fatou's Lemma

$$\limsup_{s\searrow 0} I_5 \leq \limsup_{s\searrow 0} 2\int_0^{4T} \frac{e^{-t/\varepsilon}}{\varepsilon}\frac{(\partial_t\eta_t)^2}{\eta_t}\left[\mathcal{H}(\rho_t^\varepsilon) - \mathcal{H}(\rho_t^s)\right]\,\mathrm{d}t \leq 0.$$

Putting everything together, yields

$$\limsup_{s\searrow 0} I_1 \leq C - \frac{1}{2}\int_0^T (1-e^{-t/2})^2\mathcal{I}(\rho_t^\varepsilon)\,\mathrm{d}t. \tag{6.2.20}$$

Next, we estimate I_2. The first variations for $\mathcal{V}$ and $\mathcal{W}$ are given by the classical theory of the flow-interchange lemma:

$$\limsup_{s\searrow 0} \frac{1}{s}\left[\mathcal{V}(\rho_t^s) + \mathcal{W}(\rho_t^s) - \mathcal{V}(\rho_t^\varepsilon) - \mathcal{W}(\rho_t^\varepsilon)\right] = \eta_t \int_\Omega (\Delta V + \Delta W * \rho_t^\varepsilon)\rho_t^\varepsilon\,\mathrm{d}x.$$

So we get by Fatou's Lemma, by the uniform bounds of ΔV and ΔW, and with (6.2.13)

$$\limsup_{s\searrow 0} I_2 \leq d_2\int_0^\infty \frac{e^{-t/\varepsilon}}{\varepsilon}\eta_t\,\mathrm{d}t \leq 4d_2T. \tag{6.2.21}$$

Insert (6.2.20) and (6.2.21) into (6.2.16) to obtain the desired result (6.2.15). □

6.2.4 Convergence

With the better a priori bounds of theorem 6.2.5 at hand we can prove convergence. The cornerstone of the proof is the extension of Aubin-Lions theorem for Banach spaces [85].

Theorem 6.2.6 (Convergence). *Given a vanishing sequence $(\varepsilon_n)_{n\in\mathbb{N}}$ with $\varepsilon_n \in (0,\varepsilon_*)$ and let $(\rho_t^{\varepsilon_n})_{n\in\mathbb{N}}$ be the WED-approximation the with respect to $\mathcal{E}$ emanating from $\rho_0 \in \mathcal{D}(\mathcal{E})$. Then, there exits a (non-relabeld) subsequence ε_n and a limit function $\rho_t^* : [0,\infty)\times\Omega \to \mathbb{R}$ with $\rho_t^* \in \mathcal{P}_2^{ac}(\Omega)$ for each $t > 0$ such that for all $p \in [1,\infty)$ and for all $T > 0$:*

$$\rho_t^{\varepsilon_n} \to \rho_t^* \qquad \text{strongly in } \mathrm{L}^p(0,T;\mathrm{L}^1(\Omega)).$$

This limit function ρ_t^* is a solution of the linear Fokker-Planck equation (6.0.1) *in the weak sense of* (6.2.22).

Proof. Fix $T > 0$. We seek to apply Theorem 2.5.4 for $(\rho_t^{\varepsilon_n})_{n\in\mathbb{N}}$ with the underlying Banach space $\mathbf{X} = \mathrm{L}^1(\Omega)$. We consider as normal coercive functional $\mathcal{A}_t : (0,T)\times\mathrm{L}^1(\Omega) \to [0,\infty]$ and as pseudo-distance g on $\mathrm{L}^1(\Omega)$:

$$\mathcal{A}(\rho) := \begin{cases} \left\|\sqrt{\rho}\right\|_{\mathrm{H}^1(\Omega)} + \boldsymbol{M}_2(\rho) & \text{if } \sqrt{\rho} \in \mathrm{H}^1(\Omega), \\ +\infty & \text{else,} \end{cases}$$

$$g(f,h) := \begin{cases} \mathbf{W}_2(f,h) & \text{if } f,h \in \mathcal{P}_2^{\mathrm{ac}}(\Omega), \\ +\infty & \text{else.} \end{cases}$$

The functional $\mathcal{A}_t$ is normal coercive integrand and g is a lower semi-continuous pseudo-distance in the sense of definition 2.5.1, cf. lemma 2.5.3.

Due to the ε-independent bounds (6.1.10), the sequence $(\rho_t^{\varepsilon_n})_{n\in\mathbb{N}}$ is tight with respect to $\mathcal{A}_t$. To be more precise, we use

$$\boldsymbol{M}_2(\rho_t^{\varepsilon_n}) \leq 2\boldsymbol{M}_2(\rho_0) + 2\mathbf{W}_2^2(\rho_0, \rho_t^{\varepsilon_n}) \leq 2\boldsymbol{M}_2(\rho_0) + 2t\int_0^t \mathcal{K}(\rho_r^{\varepsilon_n}, \mathbf{w}_r^{\varepsilon_n})\,\mathrm{d}r$$

since $(\rho_{st}^{\varepsilon_n}, t\mathbf{w}_{st}^{\varepsilon_n})_{s\in[0,1]}$ is a feasible competitor in the Benamou-Brennier formula for ρ_0 and $\rho_t^{\varepsilon_n}$. Hence, integrating this with respect to t yields the uniform integrability of the map $t \mapsto \boldsymbol{M}_2(\rho_t^{\varepsilon_n})$. The uniform integrability of $t \mapsto \|\sqrt{\rho_t^{\varepsilon_n}}\|_{\mathrm{H}^1(\Omega)}$ follows by the integrability of the Fisher information, see (6.1.10). Hence, the sequence $(\rho_t^{\varepsilon_n})_{n\in\mathbb{N}}$ is tight with respect to $\mathcal{A}$.

The weak integral equi-continuity condition for $(\rho_t^{\varepsilon_n})_{n\in\mathbb{N}}$ follows also by (6.1.10). To be more precise, we utilize the dynamic Benamou-Brenier formulation of the L^2-Wasserstein distance. The density-flux pair $(\rho_{t+sh}^{\varepsilon}, h\mathbf{w}_{t+sh}^{\varepsilon})$ solves the continuity equation and connects ρ_t and ρ_{t+h} and therefore we can estimate

$$\mathbf{W}_2^2(\rho_t^{\varepsilon_n}, \rho_{t+h}^{\varepsilon_n}) \leq \int_0^1 \mathcal{K}(\rho_{t+sh}^{\varepsilon_n}, h\mathbf{w}_{t+sh}^{\varepsilon_n})\,\mathrm{d}t = \int_t^{t+h} h\mathcal{K}(\rho_s^{\varepsilon_n}, \mathbf{w}_s^{\varepsilon_n})\,\mathrm{d}s \leq hCe^{BT}$$

where the constants B and C are independent of ε. Hence, there is a ε-independent constant C such that

$$\int_0^{T-h} g(\rho_t^{\varepsilon_n}, \rho_{t+h}^{\varepsilon_n})\,\mathrm{d}t = \int_0^{T-h} \mathbf{W}_2(\rho_t^{\varepsilon_n}, \rho_{t+h}^{\varepsilon_n})\,\mathrm{d}t \leq C\,h^{1/2}.$$

Therefore, theorem 2.5.4 implies that (on a subsequence) $(\rho_t^{\varepsilon_n})_{n\in\mathbb{N}}$ converges to some ρ_t^* in $\mathcal{M}(0,T;\mathrm{L}^1(\Omega))$. Since $\rho_t^\varepsilon \in \mathcal{P}_2(\Omega)$ for each t and ε, we have uniform $\mathrm{L}^p(0,T;\mathrm{L}^1(\Omega))$ bounds. So we can use Remark 2.1.1 to conclude the strong convergence of $(\rho_t^{\varepsilon_n})_{n\in\mathbb{N}}$ to ρ_t^* in the $\mathrm{L}^p(0,T;\mathrm{L}^1(\Omega))$-topology. A diagonal argument in $T\to\infty$ yields the result.

It remains to proof, that ρ_t^* is a solution of the linear Fokker-Planck equation. Therefore, for a given test functions $\varphi_t \in \mathcal{C}_c^\infty([0,\infty)\times\overline{\Omega})$ with $\nabla\varphi_t\cdot\boldsymbol{n} = 0$ on $\partial\Omega$ we pass in the Euler-Lagrange equations (6.2.6) to the limit $\varepsilon_n \to 0$. First, define $T > 0$ such that $\varphi_t = 0$ for all $t \geq T$ and split up the different parts of (6.2.6):

$$\begin{aligned}
0 =& \int_0^\infty \int_\Omega \langle \mathbf{w}_t^{\varepsilon_n}, \nabla\varphi_t\rangle \,\mathrm{d}x\,\mathrm{d}t \\
&- \int_0^\infty (1-e^{-t/\varepsilon_n}) \int_\Omega \rho_t^{\varepsilon_n}\Delta\varphi_t - \rho_t^{\varepsilon_n}\langle \nabla V + \nabla W * \rho_t^{\varepsilon_n}, \nabla\varphi_t\rangle \,\mathrm{d}x\,\mathrm{d}t \\
&+ \varepsilon_n \int_0^\infty (1-e^{-t/\varepsilon_n}) \int_\Omega \langle \mathbf{w}_t^{\varepsilon_n}, \partial_t\nabla\varphi_t\rangle \,\mathrm{d}x\,\mathrm{d}t \\
&+ \varepsilon_n \int_0^\infty (1-e^{-t/\varepsilon_n}) \int_\Omega \frac{\langle \mathbf{w}_t^{\varepsilon_n}, \mathrm{Hess}\,\varphi_t \mathbf{w}_t^{\varepsilon_n}\rangle}{\rho_t^{\varepsilon_n}} \,\mathrm{d}x\,\mathrm{d}t \\
=:& I_1 - I_2 + \varepsilon_n(I_3 + I_4).
\end{aligned}$$

We simplify I_1 by using the fact that each density-flux pair $(\rho_t^\varepsilon, \mathbf{w}_t^\varepsilon)$ solves the continuity equation to obtain in the limit $\varepsilon_n \to 0$:

$$\lim_{n\to\infty} I_1 = \lim_{n\to\infty} -\int_0^\infty \int_\Omega \rho_t^{\varepsilon_n} \partial_t \varphi_t \,\mathrm{d}x\,\mathrm{d}t - \int_\Omega \varphi_0\rho_0 \,\mathrm{d}x = -\int_0^\infty \int_\Omega \rho_t^* \partial_t \varphi_t \,\mathrm{d}x\,\mathrm{d}t - \int_\Omega \varphi_0\rho_0 \,\mathrm{d}x$$

thanks to the strong $\mathrm{L}^1(0,T;\mathrm{L}^1(\Omega))$-convergence of $(\rho_t^{\varepsilon_n})_{n\in\mathbb{N}}$ to ρ_t^*. The limit of I_2 is given by

$$\begin{aligned}\lim_{n\to\infty} I_2 &= \lim_{n\to\infty} \int_0^\infty (1-e^{-t/\varepsilon_n}) \int_\Omega \rho_t^{\varepsilon_n} \Delta\varphi_t - \rho_t^{\varepsilon_n} \langle \nabla V + \nabla W * \rho_t^{\varepsilon_n}, \nabla\varphi_t\rangle \,\mathrm{d}x\,\mathrm{d}t \\ &= \int_0^\infty \int_\Omega \rho_t^* \Delta\varphi_t - \rho_t^* \langle \nabla V + \nabla W * \rho_t^*, \nabla\varphi_t\rangle \,\mathrm{d}x\,\mathrm{d}t\end{aligned}$$

due to the strong convergence of $(\rho_t^{\varepsilon_n})_{n\in\mathbb{N}}$ in the $\mathrm{L}^1(0,T;\mathrm{L}^1(\Omega))$-topology and the point-wise convergence of $t \mapsto (1-e^{-t/\varepsilon_n})$. We estimate I_3 using the Cauchy-Schwartz inequality, Young's inequality, and (6.1.10) to get

$$|I_3| \le \|\partial_t \nabla\varphi_t\|_\infty \Big(\int_0^T \mathcal{K}(\rho_t^{\varepsilon_n}, \mathbf{w}_t^{\varepsilon_n}) \,\mathrm{d}t + \int_0^T \int_\Omega \rho_t^{\varepsilon_n} \,\mathrm{d}x\,\mathrm{d}t \Big) \le Ce^{BT}.$$

We estimate I_4 with (6.1.10) such that

$$|I_4| \le \|\mathrm{Hess}\,\varphi_t\|_\infty \int_0^T \mathcal{K}(\rho_t^{\varepsilon_n}, \mathbf{w}_t^{\varepsilon_n}) \,\mathrm{d}t \le Ce^{BT}.$$

Hence, we have shown $\lim_{n\to\infty} \varepsilon_n(I_3+I_4) = 0$. In conclusion, the limit function ρ_t^* solves

$$\begin{aligned} 0 = &\int_0^\infty \int_\Omega \rho_t^* \partial_t\varphi_t \,\mathrm{d}x\,\mathrm{d}t + \int_\Omega \varphi_0\rho_0 \,\mathrm{d}x \\ &- \int_0^\infty \int_\Omega \rho_t^* \Delta\varphi_t - \rho_t^* \langle \nabla V + \nabla W * \rho_t^*, \nabla\varphi_t\rangle \,\mathrm{d}x\,\mathrm{d}t \end{aligned} \tag{6.2.22}$$

yielding that ρ_t^* is a distributional solution to the linear Fokker-Planck equation (6.0.1). □

6.3 Application to Derrida-Lebowitz-Speer-Spohn Equation

Lastly, we investigate the Weighted Energy-Dissipation principle applied to the Derrida-Lebowitz-Speer-Spohn equation, given by

$$\partial_t \rho_t = -\operatorname{div}\big(\rho_t \nabla\big(2\frac{\Delta\sqrt{\rho_t}}{\sqrt{\rho_t}}\big)\big) \tag{6.3.1}$$

starting from the initial configuration ρ_0 on $\Omega = \mathbb{R}^d$. We use the techniques and methods developed in the previous section 6.2 to prove the existence and convergence of the WED-approximation. Recall, the corresponding free energy functional $\mathcal{E}$ in the L^2-Wasserstein framework is given by the Fisher information, defined by

$$\mathcal{E}(\mu) := \int_\Omega \|\nabla(\sqrt{\rho})\|^2 \,\mathrm{d}x$$

if the measure $\mu = \rho\,\mathrm{d}\mathcal{L}^d$ is absolutely continuous and $\sqrt{\rho} \in H^1(\Omega)$, otherwise we set $\mathcal{E}(\mu) = \infty$.

Method. In the case of the Derrida-Lebowitz-Speer-Spohn equation driven by the Fisher Information our approximation of the L^2-Wasserstein gradient flow reads:

Scheme. Given the free energy functional $\mathcal{E}$ and an initial configuration ρ_0, define the WED-approximation $(\rho_t^\varepsilon, \mathbf{w}_t^\varepsilon)$ for a given perturbation $\varepsilon \in (0, \varepsilon_*)$ as the minimizer of the WED-functional Ψ_ε, i.e.

$$(\rho_t^\varepsilon, \mathbf{w}_t^\varepsilon) \in \operatorname*{argmin}_{(\rho_t, \mathbf{w}_t) \in \mathcal{C}(\rho_0)} \int_0^\infty \frac{e^{-t/\varepsilon}}{\varepsilon} \int_\Omega \frac{\varepsilon}{2}\frac{\|\mathbf{w}_t\|^2}{\rho_t} + \|\nabla(\sqrt{\rho_t})\|^2 \,\mathrm{d}x\,\mathrm{d}t.$$

Since Ω is bounded the free energy functional $\mathcal{E}$ satisfies (E1)–(E3) and we can use the results from the previous section.

Strategy of the Proof. We use the same techniques and results as in the previous section 6.2 to derive the convergence of the WED-approximation $(\rho_t^\varepsilon, \mathbf{w}_t^\varepsilon)$ applied to the DLSS equation (6.3.1). Firstly, we derive in section 6.3.1 in a similar way the time-continuous Euler-Lagrange equations of the WED-approximation $(\rho_t^\varepsilon, \mathbf{w}_t^\varepsilon)$ and, secondly, we derive in section 6.3.2 the refined regularity estimates by means of the time-continuous flow interchange lemma. It is worthwhile to mention that due to the stronger, intrinsic ε-independent bounds of the approximation ρ_t^ε the calculations in the time-continuous flow interchange lemma are easier to justify than in the previous case. Still, we cannot extend our result to open, bounded, and convex domains Ω because of the lack of a suitable explicit representation formula fo the WED-approximation $(\rho_t^\varepsilon, \mathbf{w}_t^\varepsilon)$. In section 6.3.3, we can prove due to the better a priori estimates the convergence of the WED-approximation $(\rho_t^\varepsilon, \mathbf{w}_t^\varepsilon)$ to a weak solution of the Derrida-Lebowitz-Speer-Spohn equation (6.3.1) by means of the extension of Aubin-Lions compactness theorem 2.5.4.

6.3.1 Time-Continuous Euler-Lagrange Equations

As before, we want to derive the Euler-Lagrange equations by means of perturbing the WED-approximation $(\rho_t^\varepsilon, \mathbf{w}_t^\varepsilon)$ along the Transport equation. Hence, with the same notation as in the previous section, we define explicitly the perturbation $(\rho_t^s, \mathbf{w}_t^s)$ by

$$(\rho_t^s \circ \mathrm{X}_t^s) \cdot \mathfrak{X}_t^s = \rho_t^\varepsilon, \qquad (\mathbf{w}^s \circ \mathrm{X}_t^s) \cdot \mathfrak{X}_t^s = \mathrm{DX}_t^s \mathbf{w}_t^\varepsilon + \rho_t^\varepsilon \cdot \partial_t \mathrm{X}_t^s$$

for a fixed velocity field $\xi_t \in \mathcal{C}_c^\infty([0,\infty) \times \overline{\Omega}, \mathbb{R}^{\mathrm{d}})$ with $\xi_0 = 0$ and with $\xi_t \cdot \boldsymbol{n} = 0$ on $\partial\Omega$. We denote by X_t^s the flow map with respect to ξ_t.

The Euler-Lagrange equation reads than as follows.

Theorem 6.3.1 (Time-Continuous Euler-Lagrange Equations)**.** *Let $\varepsilon \in (0, \varepsilon_*)$. Then, the WED-approximation $(\rho_t^\varepsilon, \mathbf{w}_t^\varepsilon)$ with respect to $\mathcal{E}$ emanating from $\rho^0 \in \mathcal{D}(\mathcal{E})$ satisfies for each test function $\varphi_t \in \mathcal{C}_c^\infty([0,\infty) \times \overline{\Omega})$ with $\nabla\varphi_t \cdot \boldsymbol{n} = 0$ on $\partial\Omega$:*

$$\begin{aligned} 0 = &\int_0^\infty \int_\Omega \langle \nabla\varphi_t, \mathbf{w}_t^\varepsilon \rangle \,\mathrm{d}x \,\mathrm{d}t + \varepsilon \int_0^\infty (1 - e^{-t/\varepsilon}) \int_\Omega \frac{\langle \mathbf{w}_t^\varepsilon, \operatorname{Hess}\varphi_t \mathbf{w}_t^\varepsilon \rangle}{\rho_\varepsilon^t} \,\mathrm{d}x \,\mathrm{d}t \\ &+ \varepsilon \int_0^\infty (1 - e^{-t/\varepsilon}) \int_\Omega \langle \partial_t \nabla\varphi_t, \mathbf{w}_t^\varepsilon \rangle \,\mathrm{d}x \,\mathrm{d}t \\ &- \int_0^\infty (1 - e^{-t/\varepsilon}) \int_\Omega \frac{1}{2} \langle \nabla(\Delta\varphi_t), \nabla\rho_t^\varepsilon \rangle + 2 \langle \operatorname{Hess}\varphi_t \nabla\sqrt{\rho_t^\varepsilon}, \nabla\sqrt{\rho_t^\varepsilon} \rangle \,\mathrm{d}x \,\mathrm{d}t. \end{aligned} \tag{6.3.2}$$

Proof. We proceed as in the proof of Theorem 6.2.3 with the only difference that we have to compute the first variation of the Fisher-information $\mathcal{I}$ along solutions of the Transport equation (6.2.2). This has already been computed in [47, Thm 4.2] for Ω open, bounded and convex or [68, Lemma 2.5.] for $\Omega = \mathbb{R}^{\mathrm{d}}$ and is given by

$$\begin{aligned} &\frac{\mathrm{d}}{\mathrm{d}s} \Big[\int_0^\infty \frac{e^{-t/\varepsilon}}{\varepsilon} \mathcal{I}(\rho_t^s) \,\mathrm{d}t \Big]_{s=0} \\ &= - \int_0^\infty \frac{e^{-t/\varepsilon}}{\varepsilon} \int_\Omega \frac{1}{2} \langle \nabla(\operatorname{div}\xi_t), \nabla\rho_t^\varepsilon \rangle + 2 \langle \mathrm{D}\xi_t \nabla\sqrt{\rho_t^\varepsilon}, \nabla\sqrt{\rho_t^\varepsilon} \rangle \,\mathrm{d}x \,\mathrm{d}t. \end{aligned}$$

Combing this with the first variation of the kinetic energy $\mathcal{K}$ yields the desired result. □

6.3.2 Better A Priori Bounds by Continuous Variations

Also in this case of the DLSS equation, the ε-independent bounds (6.1.10) are not sufficient to pass in the Euler-Lagrange equations (6.3.2) to the limit $\varepsilon \to 0$. Analogously to the section before, we consider as perturbation ρ_t^s of the approximated solution ρ_t^ε the solution of the Heat equation

$$\partial_s \rho_t^s = \eta_t \Delta(\rho_t^s), \qquad \rho_t^0 = \rho_t^\varepsilon$$

with no-flux boundary condition and for some given smooth diffusivity parameter $\eta_t \in C_c^\infty(0,\infty)$ with $\eta_0 = 0$. If we denote by $\mathfrak{G}_t^s$ the greens function for the heat equation on $\Omega = \mathbb{R}^{\mathrm{d}}$ with diffusivity parameter η_t , the explicit representations for ρ_t^s and $\mathbf{w}_t^s$ are given by

$$\rho_t^s := \mathcal{T}_{\mathfrak{G}_t^s}[\rho_t^\varepsilon] \quad \text{and} \quad \mathbf{w}_t^s := \vec{\mathcal{T}}_{\mathfrak{G}_t^s}[\mathbf{w}_t^\varepsilon] - s\partial_t\eta_t\nabla\mathcal{T}_{\mathfrak{G}_t^s}[\rho_t^\varepsilon].$$

The better a priori bounds are then given by:

Theorem 6.3.2 (Better A Priori Bounds)**.** *Let $\varepsilon \in (0,\varepsilon_*)$. Then, there exists a non-negative constant C, which depends only on T and the initial datum ρ^0 such that the WED-approximation $(\rho_t^\varepsilon, \mathbf{w}_t^\varepsilon)$ with respect to $\mathcal{E}$ emanating from $\rho^0 \in \mathcal{D}(\mathcal{I})$ satisfies*

$$\int_0^T (1-e^{-t/2})^2 \int_\Omega \|\mathrm{D}^2(\sqrt{\rho_t^\varepsilon})\|^2 \,\mathrm{d}x\,\mathrm{d}t \leq C. \tag{6.3.3}$$

Proof. Fix $T > 0$. Given the WED-approximation $(\rho_t^\varepsilon, \mathbf{w}_t^\varepsilon)$ with respect to $\mathcal{I}$, let $(\rho_t^s, \mathbf{w}_t^s)$ be the perturbation defined by (6.2.14) with η_t given by (6.2.12). Due to the minimality property of $(\rho_t^\varepsilon, \mathbf{w}_t^\varepsilon)$ and due to $(\rho_t^s, \mathbf{w}_t^s) \in \mathcal{C}(\rho^0)$, we have

$$0 \leq \int_0^\infty \frac{e^{-t/\varepsilon}}{2}\frac{1}{s}\left[\mathcal{K}(\rho_t^s, \mathbf{w}_t^s) - \mathcal{K}(\rho_t^\varepsilon, \mathbf{w}_t^\varepsilon)\right]\mathrm{d}t + \int_0^\infty \frac{e^{-t/\varepsilon}}{\varepsilon}\frac{1}{s}\left[\mathcal{I}(\rho_t^s) - \mathcal{I}(\rho_t^\varepsilon)\right]\mathrm{d}t. \tag{6.3.4}$$

As before, by passing in (6.3.4) to the limit $s \searrow 0$, we obtain the better a priori bound (6.3.3). For fixed $t > 0$ the first variation of the Fisher information $\mathcal{I}$ along the heat flow is given by, cf. [47, Theorem 5.1] or [68, Lemma 4.4.]:

$$\limsup_{s\searrow 0} \frac{1}{s}\left(\mathcal{I}(\rho_t^s) - \mathcal{I}(\rho_t^\varepsilon)\right) \leq -C\eta_t \int_\Omega \|\mathrm{D}^2(\sqrt{\rho_t^\varepsilon})\|^2 \,\mathrm{d}x \tag{6.3.5}$$

where C is a given non negative constant. Note, (6.3.5) incorporates the fact that the Fisher information is a Lyapunov function for the heat flow and therefore one has $\mathcal{I}(\rho_t^s) \leq \mathcal{I}(\rho_t^\varepsilon)$. Applying Fatou's lemma yields

$$\begin{aligned}\limsup_{s\searrow 0} \int_0^\infty \frac{e^{-t/\varepsilon}}{\varepsilon}\frac{1}{s}\left[\mathcal{I}(\rho_t^s) - \mathcal{I}(\rho_t^\varepsilon)\right]\mathrm{d}t &\leq - \int_0^\infty \frac{e^{-t/\varepsilon}}{\varepsilon} C\eta_t \int_\Omega \|\mathrm{D}^2(\sqrt{\rho_t^\varepsilon})\|^2 \,\mathrm{d}x\,\mathrm{d}t \\ &\leq - C\int_0^T (1-e^{-t/2})^2 \int_\Omega \|\mathrm{D}^2(\sqrt{\rho_t^\varepsilon})\|^2 \,\mathrm{d}x\,\mathrm{d}t.\end{aligned}$$

To estimate the difference quotient of the kinetic energy $\mathcal{K}$ we proceed similar to the proof of lemma 6.2.2. I.e., we use (6.2.17) given by:

$$\frac{1}{s}\left[\mathcal{K}(\rho_t^s, \mathbf{w}_t^s) - \mathcal{K}(\rho_t^\varepsilon, \mathbf{w}_t^\varepsilon)\right] \leq 4\sqrt{\mathcal{K}(\rho_t^\varepsilon, \mathbf{w}_t^\varepsilon)}\,|\partial_t\eta_t|\,\sqrt{\mathcal{I}(\rho_t^s)} + 4s\,|\partial_t\eta_t|^2\,\mathcal{I}(\rho_t^s).$$

Applying Young's inequality on the first part of the right-hand side yields

$$\begin{aligned}\frac{1}{s}\left[\mathcal{K}(\rho_t^s,\mathbf{w}_t^s)-\mathcal{K}(\rho_t^\varepsilon,\mathbf{w}_t^\varepsilon)\right] &\leq 2\frac{\eta_t}{\varepsilon}\mathcal{K}(\rho_t^\varepsilon,\mathbf{w}_t^\varepsilon)+4(1+s)\varepsilon\frac{|\partial_t\eta_t|^2}{\eta_t}\mathcal{I}(\rho_t^s)\\ &\leq\big(2\mathcal{K}(\rho_t^\varepsilon,\mathbf{w}_t^\varepsilon)+4(1+s)\mathcal{I}(\rho_t^s)\big)\mathbf{1}_{\{t\leq 4T\}}.\end{aligned}$$

where we used the properties (6.2.13) from η_t. Using the monotonicity of the Fisher information, an integration in time yields

$$\int_0^\infty \frac{e^{-t/\varepsilon}}{2}\frac{1}{s}\left[\mathcal{K}(\rho_t^s,\mathbf{w}_t^s)-\mathcal{K}(\rho_t^\varepsilon,\mathbf{w}_t^\varepsilon)\right]\mathrm{d}t\leq\int_0^{4T}\mathcal{K}(\rho_t^\varepsilon,\mathbf{w}_t^\varepsilon)+2(1+s)\mathcal{I}(\rho_t^\varepsilon)\,\mathrm{d}t.$$

Hence, the limit $s\to 0$ is given by

$$\limsup_{s\searrow 0}\int_0^\infty \frac{e^{-t/\varepsilon}}{2}\frac{1}{s}\left[\mathcal{K}(\rho_t^s,\mathbf{w}_t^s)-\mathcal{K}(\rho_t^\varepsilon,\mathbf{w}_t^\varepsilon)\right]\mathrm{d}t\leq\int_0^{4T}\mathcal{K}(\rho_t^\varepsilon,\mathbf{w}_t^\varepsilon)\,\mathrm{d}t+2\int_0^{4T}\mathcal{I}(\rho_t^\varepsilon)\,\mathrm{d}t.$$

So we can pass to the limit $s\to 0$ in (6.3.4) and conclude

$$0\leq\int_0^{4T}\mathcal{K}(\rho_t^\varepsilon,\mathbf{w}_t^\varepsilon)\,\mathrm{d}t+2\int_0^{4T}\mathcal{I}(\rho_t^\varepsilon)\,\mathrm{d}t-C\int_0^T(1-e^{-t/2})^2\int_\Omega\left\|\mathrm{D}^2(\sqrt{\rho_t^\varepsilon})\right\|^2\,\mathrm{d}x\,\mathrm{d}t.$$

Rearranging terms and estimating the first two integral with (6.1.10) yields the desired result. □

6.3.3 Convergence

With the better a priori bounds of theorem 6.3.2 at hand we can prove convergence. The cornerstone of the proof is a version of Aubin-Lions theorem for Banach spaces, cf. [85].

Theorem 6.3.3 (Convergence). *Given a vanishing sequence $(\varepsilon_n)_{n\in\mathbb{N}}$ with $\varepsilon_n\in(0,\varepsilon_*)$. Let $(\rho_t^{\varepsilon_n})_{n\in\mathbb{N}}$ be the WED-approximation the with respect to $\mathcal{E}$ emanating from $\rho^0\in\mathcal{D}(\mathcal{E})$. Then there exits a (non-relabeled) subsequence ε_n and a limit function $\rho_t^*:[0,\infty)\times\Omega\to\mathbb{R}$ with $\rho_t^*\in\mathcal{P}_2^{ac}(\Omega)$ or each $t>0$ such that for all $p\in[1,\infty)$ and for all $T>0$:*

$$\begin{aligned}\rho_t^{\varepsilon_n}&\to\rho_t^* &&\text{strongly in } \mathrm{L}^p(0,T;\mathrm{L}^1(\Omega)),\\ \sqrt{\rho_t^{\varepsilon_n}}&\to\sqrt{\rho_t^*} &&\text{strongly in } \mathrm{L}^2(0,T;\mathrm{H}^1(\Omega)),\\ \sqrt{\rho_t^{\varepsilon_n}}&\rightharpoonup\sqrt{\rho_t^*} &&\text{weakly in } \mathrm{L}^2(0,T;\mathrm{H}^2(\Omega)).\end{aligned}$$

This limit function ρ_t^* is a solution of the Derrida-Lebowitz-Speer-Spohn equation (6.3.1) *in the weak sense of* (6.3.6).

Proof. Fix $T>0$. To derive in this case the convergence result we proceed similar as before, applying Theorem 2.5.4 once. The other convergence results follow by the Banach-Alaoglu theorem and an interpolation argument.

We seek to apply the Aubin-Lions theorem 2.5.4 to the sequence $(\rho_t^{\varepsilon_n})_{n\in\mathbb{N}}$ with the underlying Banach space $\mathbf{X} = \mathrm{L}^1(\Omega)$. We consider as normal coercive functional $\mathcal{A} : \mathrm{L}^1(\Omega) \to [0,\infty]$ and as pseudo-distance g on $\mathrm{L}^1(\Omega)$:

$$\mathcal{A}(u) := \begin{cases} \left\|\sqrt{\rho}\right\|_{\mathrm{H}^1(\Omega)} + \boldsymbol{M}_2(\rho) & \text{if } \sqrt{\rho} \in \mathrm{H}^1(\Omega) \text{ and } \rho \in \mathcal{P}_2^{\mathrm{ac}}(\Omega), \\ \infty & \text{else,} \end{cases}$$

$$g(\rho,\eta) := \begin{cases} \mathbf{W}_2(\rho,\eta) & \text{if } \rho,\eta \in \mathcal{P}_2^{\mathrm{ac}}(\Omega), \\ +\infty & \text{else.} \end{cases}$$

The functional $\mathcal{A}_t$ is measurable, lower semi-continuous and has compact sublevels with respect to the $\mathrm{L}^1(\Omega)$-topology and g is a lower semi-continuous pseudo-distance, cf. lemma 2.5.3.

Due to the ε-independent bounds(6.1.10), the sequence $(\rho_t^{\varepsilon_n})_{n\in\mathbb{N}}$ is tight with respect to $\mathcal{A}_t$. To be more precise, we use

$$\boldsymbol{M}_2(\rho_t^{\varepsilon_n}) \le 2\boldsymbol{M}_2(\rho_0) + 2\mathbf{W}_2^2(\rho_0,\rho_t^{\varepsilon_n}) \le 2\boldsymbol{M}_2(\rho_0) + 2t\int_0^t \mathcal{K}(\rho_r^{\varepsilon_n},\mathbf{w}_r^{\varepsilon_n})\,\mathrm{d}r$$

since $(\rho_{st}^{\varepsilon_n}, t\mathbf{w}_{st}^{\varepsilon_n})_{s\in[0,1]}$ is a feasible competitor in the Benamou-Brennier formula for ρ_0 and $\rho_t^{\varepsilon_n}$. Hence, integrating this with respect to t yields the uniform integrablity and we have the tightness of the sequence $(\rho_t^{\varepsilon_n})_{n\in\mathbb{N}}$. The weak integral equi-continuity condition for $(\rho_t^{\varepsilon_n})_{n\in\mathbb{N}}$ follows by the same calculations as in the proof of theorem 6.2.6. We can conclude by theorem 2.5.4 that (on a subsequence) $(\rho_t^{\varepsilon_n})_{n\in\mathbb{N}}$ converges to some ρ_t^* in $\mathcal{M}(0,T;\mathrm{L}^1(\Omega))$. Clearly, we have also uniform $\mathrm{L}^\infty(0,T;\mathrm{L}^1(\Omega))$ bounds. So we can use Remark 2.1.1 to conclude the strong convergence result of $(\rho_t^{\varepsilon_n})_{n\in\mathbb{N}}$ to some ρ_t^* in $\mathrm{L}^p(0,T;\mathrm{L}^1(\Omega))$ for every $p \ge 1$.

The refined a priori estimates (6.3.3) yield also by the Banach-Alaoglu theorem the convergence of $(\sqrt{\rho_t^{\varepsilon_n}})_{n\in\mathbb{N}}$ in the weak $\mathrm{L}^2(0,T;\mathrm{H}^2(\Omega))$-topology. An additional interpolation argument of this weak convergence result in $\mathrm{L}^2(0,T;\mathrm{H}^2(\Omega))$ and the strong convergence in $\mathrm{L}^2(0,T;\mathrm{L}^2(\Omega))$ yields the desired strong convergence of $(\sqrt{\rho_t^{\varepsilon_n}})_{n\in\mathbb{N}}$ in the $\mathrm{L}^2(0,T;\mathrm{H}^1(\Omega))$-topology. An additional diagonal argument in $T \to \infty$ yields the desired convergence result for all $T > 0$.

It remains to prove, that ρ_t^* is a solution of the DLSS-equation (6.3.1). Therefore, for a given test functions $\varphi_t \in \mathcal{C}_c^\infty([0,\infty)\times\overline{\Omega})$ with $\nabla\varphi_t \cdot \boldsymbol{n} = 0$ on $\partial\Omega$ we pass in the Euler-Lagrange equations (6.3.2) to the limit as $\varepsilon_n \searrow 0$. First, define $T > 0$ such that $\varphi_t = 0$ for all $t \ge T$ and split up the different parts of the Euler-Lagrange equation

(6.3.2):

$$\begin{aligned}
0 =& \int_0^\infty \int_\Omega \langle \nabla \varphi_t, \mathbf{w}_t^{\varepsilon_n} \rangle \,\mathrm{d}x \,\mathrm{d}t \\
&- \int_0^\infty (1 - e^{-t/\varepsilon_n}) \int_\Omega \frac{1}{2} \langle \nabla(\Delta \varphi_t), \nabla \rho_t^{\varepsilon_n} \rangle + 2 \langle \operatorname{Hess} \varphi_t \nabla \sqrt{\rho_t^{\varepsilon_n}}, \nabla \sqrt{\rho_t^{\varepsilon_n}} \rangle \,\mathrm{d}x \,\mathrm{d}t \\
&+ \varepsilon_n \int_0^\infty (1 - e^{-t/\varepsilon_n}) \int_\Omega \langle \partial_t \nabla \varphi_t, \mathbf{w}_t^{\varepsilon_n} \rangle \,\mathrm{d}x \,\mathrm{d}t \\
&+ \varepsilon_n \int_0^\infty (1 - e^{-t/\varepsilon_n}) \int_\Omega \frac{\langle \mathbf{w}_t^{\varepsilon_n}, \operatorname{Hess} \varphi_t \,\mathbf{w}_t^{\varepsilon_n} \rangle}{\rho_t^{\varepsilon_n}} \,\mathrm{d}x \,\mathrm{d}t \\
=:& I_1 - I_2 + \varepsilon_n (I_3 + I_4).
\end{aligned}$$

As before, we simplify I_1 by using the fact that each density-flux pair $(\rho_t^\varepsilon, \mathbf{w}_t^\varepsilon)$ solves the continuity equation. Then we obtain in the limit $\varepsilon_n \searrow 0$

$$\lim_{n \to \infty} I_1 = - \int_0^\infty \int_\Omega \rho_t^* \partial_t \varphi_t \,\mathrm{d}x \,\mathrm{d}t - \int_\Omega \varphi_0 \rho^0 \,\mathrm{d}x$$

thanks to the strong $\mathrm{L}^1(0, T; \mathrm{L}^1(\Omega))$ convergence of $(\rho_t^{\varepsilon_n})_{n \in \mathbb{N}}$ to ρ_t^*. The limit of I_2 is given by

$$\begin{aligned}
\lim_{n \to \infty} I_2 &= \lim_{n \to \infty} \int_0^\infty (1 - e^{-t/\varepsilon_n}) \int_\Omega \frac{1}{2} \langle \nabla(\Delta \varphi_t), \nabla \rho_t^{\varepsilon_n} \rangle + 2 \langle \operatorname{Hess} \varphi_t \nabla \sqrt{\rho_t^{\varepsilon_n}}, \nabla \sqrt{\rho_t^{\varepsilon_n}} \rangle \,\mathrm{d}x \,\mathrm{d}t \\
&= \int_0^\infty \int_\Omega \frac{1}{2} \langle \nabla(\Delta \varphi_t), \nabla \rho_t^* \rangle + 2 \langle \operatorname{Hess} \varphi_t \nabla \sqrt{\rho_t^*}, \nabla \sqrt{\rho_t^*} \rangle \,\mathrm{d}x \,\mathrm{d}t
\end{aligned}$$

due to the strong convergence of $(\sqrt{\rho_t^{\varepsilon_n}})_{n \in \mathbb{N}}$ in the $\mathrm{L}^2(0, T; \mathrm{H}^1(\Omega))$-topology and the pointwise convergence of $t \mapsto (1 - e^{-t/\varepsilon_n})$. The integrals I_3 and I_4 are estimated as in the previous proof such that we have $\lim_{n \to \infty} \varepsilon_n (I_3 + I_4) = 0$. In conclusion, the limit function ρ_t^* solves

$$\begin{aligned}
&- \int_0^\infty \int_\Omega \frac{1}{2} \langle \nabla(\Delta \varphi_t), \nabla \rho_t^* \rangle + 2 \langle \operatorname{Hess} \varphi_t \nabla \sqrt{\rho_t^*}, \nabla \sqrt{\rho_t^*} \rangle \,\mathrm{d}x \,\mathrm{d}t \\
&= \int_0^\infty \int_\Omega \rho_t^* \partial_t \varphi_t \,\mathrm{d}x \,\mathrm{d}t + \int_\Omega \varphi_0 \rho^0 \,\mathrm{d}x
\end{aligned} \tag{6.3.6}$$

which proves that ρ_t^* is a solution of the DLSS equation (6.3.1) in the weak sense. □

Bibliography

[1] Georgios Akrivis and Michel Crouzeix, *Linearly implicit methods for nonlinear parabolic equations*, Mathematics of Computation, 73 (2004), pp. 613–635.

[2] Georgios Akrivis and Christian Lubich, *Fully implicit, linearly implicit and implicit–explicit backward difference formulae for quasi-linear parabolic equations*, Numerische Mathematik, 131 (2015), pp. 713–735.

[3] Samuel M Allen and John W Cahn, *A microscopic theory for antiphase boundary motion and its application to antiphase domain coarsening*, Acta Metallurgica, 27 (1979), pp. 1085–1095.

[4] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré, *Gradient flows in metric spaces and in the space of probability measures*, Lectures in Mathematics ETH Zürich, Birkhäuser Verlag, Basel, second ed., 2008.

[5] Claudio Baiocchi and Michel Crouzeix, *On the equivalence of A-stability and G-stability*, Appl. Numer. Math., 5 (1989), pp. 19–22. Recent theoretical results in numerical ordinary differential equations.

[6] Martino Bardi and Italo Capuzzo-Dolcetta, *Optimal control and viscosity solutions of Hamilton-Jacobi-Bellman equations*, Springer Science & Business Media, 2008.

[7] Jean-David Benamou and Yann Brenier, *A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem*, Numerische Mathematik, 84 (2000), pp. 375–393.

[8] Jean-David Benamou, Guillaume Carlier, Quentin Mérigot, and Edouard Oudet, *Discretization of functionals involving the Monge–Ampère operator*, Numerische Mathematik, 134 (2016), pp. 611–636.

[9] Adrien Blanchet, Vincent Calvez, and José A Carrillo, *Convergence of the mass-transport steepest descent scheme for the subcritical Patlak–Keller–Segel model*, SIAM Journal on Numerical Analysis, 46 (2008), pp. 691–721.

[10] A. Blanchet and P. Laurençot, *The parabolic-parabolic Keller-Segel system with critical diffusion as a gradient flow in* $\mathbb{R}^d$, $d \geq 3$, Comm. Partial Differential Equations, 38 (2013), pp. 658–686.

[11] Folkmar Bornemann and Peter Deuflhard, *Scientific Computing with Ordinary Differential Equations* -, Springer Science & Business Media, Berlin Heidelberg, 2012.

[12] Andrea Braides, *Γ-convergence for beginners*, vol. 22 of Oxford Lecture Series in Mathematics and its Applications, Oxford University Press, Oxford, 2002.

[13] James H Bramble, Joseph E Pasciak, Peter H Sammon, and Vidar Thomée, *Incomplete iterations in multistep backward difference methods for parabolic problems with smooth and nonsmooth data*, Mathematics of Computation, 52 (1989), pp. 339–367.

[14] Haïm Brezis, *On a characterization of flow-invariant sets*, Communications on Pure and Applied Mathematics, 23 (1970), pp. 261–263.

[15] John W Cahn and John E Hilliard, *Free energy of a nonuniform system. i. interfacial free energy*, The Journal of chemical physics, 28 (1958), pp. 258–267.

[16] Vincent Calvez and Thomas O Gallouët, *Blow-up phenomena for gradient flows of discrete homogeneous functionals*, Applied Mathematics & Optimization, (2017), pp. 1–29.

[17] JA Carrillo, FS Patacchini, P Sternberg, and Gershon Wolansky, *Convergence of a particle method for diffusive gradient flows in one dimension*, SIAM Journal on Mathematical Analysis, 48 (2016), pp. 3708–3741.

[18] Jose A Carrillo, Katy Craig, Li Wang, and Chaozhen Wei, *Primal dual methods for Wasserstein gradient flows*, arXiv preprint arXiv:1901.08081, (2019).

[19] J. A. Carrillo, M. DiFrancesco, A. Figalli, T. Laurent, and D. Slepčev, *Global-in-time weak measure solutions and finite-time aggregation for nonlocal interaction equations*, Duke Math. J., 156 (2011), pp. 229–271.

[20] Jose Antonio Carrillo, Maria Pia Gualdani, Ansgar Jüngel, et al., *Convergence of an entropic semi-discretization for nonlinear Fokker-Planck equations in* $\mathbb{R}^d$, Publicacions Matemàtiques, 52 (2008), pp. 413–433.

[21] José A Carrillo, Robert J McCann, Cédric Villani, et al., *Kinetic equilibration rates for granular media and related equations: entropy dissipation and mass transportation estimates*, Revista Matematica Iberoamericana, 19 (2003), pp. 971–1018.

[22] José A Carrillo and J Salvador Moll, *Numerical simulation of diffusive and aggregation phenomena in nonlinear continuity equations by evolving diffeomorphisms*, SIAM Journal on Scientific Computing, 31 (2009), pp. 4305–4329.

[23] José A Carrillo, Helene Ranetbauer, and Marie-Therese Wolfram, *Numerical simulation of nonlinear continuity equations by evolving diffeomorphisms*, Journal of Computational Physics, 327 (2016), pp. 186–202.

[24] Wikimedia Commons, *File:transport-plan.svg — wikimedia commons, the free media repository*, 2018. [Online; accessed 5-February-2019].

[25] Peter E Crouch and R Grossman, *Numerical integration of ordinary differential equations on manifolds*, Journal of Nonlinear Science, 3 (1993), pp. 1–33.

[26] Germund Dahlquist, *Convergence and stability in the numerical integration of ordinary differential equations*, Mathematica Scandinavica, (1956), pp. 33–53.

[27] Germund G Dahlquist, *A special stability problem for linear multistep methods*, BIT Numerical Mathematics, 3 (1963), pp. 27–43.

[28] Sara Daneri and Giuseppe Savaré, *Eulerian calculus for the displacement convexity in the Wasserstein distance*, SIAM J. Math. Anal., 40 (2008), pp. 1104–1122.

[29] Ennio De Giorgi, *New problems on Minimizing Movements*, Ennio de Giorgi: Selected Papers, (1993), pp. 699–713.

[30] Ennio De Giorgi, Antonio Marino, and Mario 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.

[31] M Di Francesco, A Esposito, and S Fagioli, *Nonlinear degenerate cross-diffusion systems with nonlocal interaction*, Nonlinear Analysis, 169 (2018), pp. 94–117.

[32] Bertram Düring, Daniel Matthes, and Josipa Pina Milišic, *A gradient flow scheme for nonlinear fourth order equations*, Discrete Contin. Dyn. Syst. Ser. B, 14 (2010), pp. 935–959.

[33] Charles M Elliott, *The Cahn-Hilliard model for the kinetics of phase separation*, in Mathematical models for phase change problems, Springer, 1989, pp. 35–73.

[34] Charles M Elliott and Harald Garcke, *On the Cahn–Hilliard equation with degenerate mobility*, Siam journal on mathematical analysis, 27 (1996), pp. 404–423.

[35] Etienne Emmrich, *Stability and error of the variable two-step BDF for semilinear parabolic problems*, Journal of Applied Mathematics and Computing, 19 (2005), pp. 33–55.

[36] ——, *Convergence of the variable two-step BDF time discretisation of nonlinear evolution problems governed by a monotone potential operator*, BIT Numerical Mathematics, 49 (2009), pp. 297–323.

[37] ——, *Two-step BDF time discretisation of nonlinear evolution problems governed by monotone operators with strongly continuous perturbations*, Computational Methods in Applied Mathematics Comput. Methods Appl. Math., 9 (2009), pp. 37–62.

[38] Matthias Erbar et al., *The heat equation on manifolds as a gradient flow in the Wasserstein space*, in Annales de l'Institut Henri Poincaré, Probabilités et Statistiques, vol. 46, Institut Henri Poincaré, 2010, pp. 1–23.

[39] Lucas CF Ferreira and Julio C Valencia-Guevara, *Gradient flows of time-dependent functionals in metric spaces and applications to PDEs*, Monatshefte für Mathematik, 185 (2018), pp. 231–268.

[40] Till Daniel Frank, *Nonlinear Fokker-Planck equations: fundamentals and applications*, Springer Science & Business Media, 2005.

[41] Thomas Gallouët and Quentin Mérigot, *A Lagrangian scheme for the incompressible Euler equation using Optimal Transport*, arXiv preprint arXiv:1605.00568, (2016).

[42] C. Gardiner, *Stochastic Methods: A Handbook for the Natural and Social Sciences*, Springer Series in Synergetics, Springer Berlin Heidelberg, 2009.

[43] C. William Gear, *Numerical Initial Value Problems in Ordinary Differential Equations*, Prentice Hall PTR, Upper Saddle River, NJ, USA, 1971.

[44] C. W. Gear and L. R. Petzold, *ODE methods for the solution of differential/algebraic systems*, SIAM J. Numer. Anal., 21 (1984), pp. 716–728.

[45] Lorenzo Giacomelli and Felix Otto, *Variatonal formulation for the lubrication approximation of the Hele-Shaw flow*, Calculus of Variations and Partial Differential Equations, 13 (2001), pp. 377–403.

[46] ——, *Variatonal formulation for the lubrication approximation of the Hele-Shaw flow*, Calculus of Variations and Partial Differential Equations, 13 (2001), pp. 377–403.

[47] U. Gianazza, G. Savaré, and G. Toscani, *The Wasserstein gradient flow of the Fisher information and the quantum drift-diffusion equation*, Arch. Ration. Mech. Anal., 194 (2009), pp. 133–220.

[48] Enrico Giusti, *Minimal surfaces and functions of bounded variation*, vol. 80, Birkhauser Verlag, 1984.

[49] Annegret Glitzky and Alexander Mielke, *A gradient structure for systems coupling reaction–diffusion effects in bulk and interfaces*, Zeitschrift für angewandte Mathematik und Physik, 64 (2013), pp. 29–52.

[50] Ernst Hairer and Christian Lubich, *Energy-diminishing integration of gradient systems*, IMA J. Numer. Anal., 34 (2014), pp. 452–461.

[51] Ernst Hairer and Gerhard Wanner, *Solving Ordinary Differential Equations II - Stiff and Differential - Algebraic Problems*, Springer Science & Business Media, Berlin Heidelberg, 2013.

[52] Eskil Hansen, *Convergence of multistep time discretizations of nonlinear dissipative evolution equations*, SIAM Journal on Numerical Analysis, 44 (2006), pp. 55–65.

[53] W. Horsthemke and R. Lefever, *Noise-Induced Transitions: Theory and Applications in Physics, Chemistry, and Biology*, Springer Series in Synergetics, Springer Berlin Heidelberg, 2006.

[54] Richard Jordan, David Kinderlehrer, and Felix Otto, *The variational formulation of the Fokker-Planck equation*, SIAM J. Math. Anal., 29 (1998), pp. 1–17.

[55] Oliver Junge, Daniel Matthes, and Horst Osberger, *A fully discrete variational scheme for solving nonlinear Fokker-Planck equations in higher space dimensions*, arXiv preprint arXiv:1509.07721, (2015).

[56] Ansgar Jüngel, *Quasi-hydrodynamic semiconductor equations*, vol. 41, Birkhäuser, 2011.

[57] Young-Heon Kim and Robert J McCann, *Towards the smoothness of optimal maps on Riemannian submersions and Riemannian products (of round spheres in particular)*, Journal für die reine und angewandte Mathematik (Crelles Journal), 2012 (2012), pp. 1–27.

[58] David Kinderlehrer, Léonard Monsaingeon, and Xiang Xu, *A Wasserstein gradient flow approach to Poisson-Nernst-Planck equations*, ESAIM: Control, Optimisation and Calculus of Variations, (2016).

[59] Horst Kreth, *Time-discretisations for nonlinear evolution equations*, in Numerical treatment of differential equations in applications, Springer, 1978, pp. 57–63.

[60] Philippe Laurençot and Bogdan-Vasile Matioc, *A gradient flow approach to a thin film approximation of the Muskat problem*, Calc. Var. Partial Differential Equations, 47 (2013), pp. 319–341.

[61] ——, *A gradient flow approach to a thin film approximation of the muskat problem*, Calculus of Variations and Partial Differential Equations, 47 (2013), pp. 319–341.

[62] Marie-Noëlle Le Roux, *Semidiscretization in time for parabolic problems*, Mathematics of Computation, 33 (1979), pp. 919–931.

[63] L Gary Leal, *Advanced transport phenomena: fluid mechanics and convective transport processes*, vol. 7, Cambridge University Press, 2007.

[64] Stefano Lisini, Daniel Matthes, and Giuseppe Savaré, *Cahn-Hilliard and thin film equations with nonlinear mobility as gradient flows in weighted-Wasserstein metrics*, J. Differential Equations, 253 (2012), pp. 814–850.

[65] Stefano Lisini, Daniel Matthes, Giuseppe Savaré, and Simon Plazotta, *Weighted Energy-Dissipation principle for* L^2*-Wasserstein gradient flows*, in preparation, (2019).

[66] Daniel Loibl, Daniel Matthes, and Jonathan Zinsl, *Existence of weak solutions to a class of fourth order partial differential equations with Wasserstein gradient structure*, Potential Analysis, 45 (2016), pp. 755–776.

[67] Peter A. Markowich, Christian A. Ringhofer, and Christian Schmeiser, *Semiconductor Equations*, Springer Vienna, 1990.

[68] Daniel Matthes, Robert J. McCann, and Giuseppe Savaré, *A family of nonlinear fourth order equations of gradient flow type*, Comm. Partial Differential Equations, 34 (2009), pp. 1352–1397.

[69] Daniel Matthes and Horst Osberger, *Convergence of a variational Lagrangian scheme for a nonlinear drift diffusion equation*, ESAIM: Mathematical Modelling and Numerical Analysis, 48 (2014), pp. 697–726.

[70] Daniel Matthes and Simon Plazotta, *A variational formulation of the BDF2 method for metric gradient flows*, ESAIM: M2AN, to appear, (2019).

[71] Daniel Matthes and Benjamin Söllner, *Convergent Lagrangian discretization for drift-diffusion with nonlocal aggregation*, in Innovative Algorithms and Analysis, Springer, 2017, pp. 313–351.

[72] Daniel Matthes and Jonathan Zinsl, *Existence of solutions for a class of fourth order cross-diffusion systems of gradient flow type*, Nonlinear Analysis, 159 (2017), pp. 316–338.

[73] Robert J. McCann, *A convexity principle for interacting gases*, Adv. Math., 128 (1997), pp. 153–179.

[74] Alexander Mielke, *A gradient structure for reaction-diffusion systems and for energy-drift-diffusion systems*, Nonlinearity, 24 (2011), p. 1329.

[75] Alexander Mielke and Michael Ortiz, *A class of minimum principles for characterizing the trajectories and the relaxation of dissipative systems*, ESAIM: Control, Optimisation and Calculus of Variations, 14 (2008), pp. 494–516.

[76] Alexander Mielke and Ulisse Stefanelli, *Weighted Energy-Dissipation functionals for gradient flows*, ESAIM: Control, Optimisation and Calculus of Variations, 17 (2011), pp. 52–85.

[77] Mitio Nagumo, *Über die lage der integralkurven gewöhnlicher differentialgleichungen*, Proceedings of the Physico-Mathematical Society of Japan. 3rd Series, 24 (1942), pp. 551–559.

[78] Felix Otto, *Lubrication approximation with prescribed nonzero contact angle*, Communications in partial differential equations, 23 (1998), pp. 2077–2164.

[79] ——, *The geometry of dissipative evolution equations: the porous medium equation*, Comm. Partial Differential Equations, 26 (2001), pp. 101–174.

[80] Gabriel Peyré, *Entropic approximation of Wasserstein gradient flows*, SIAM Journal on Imaging Sciences, 8 (2015), pp. 2323–2351.

[81] Simon Plazotta, *A BDF2-approach for the non-linear Fokker-Planck equation*, Discrete & Continuous Dynamical Systems - A, (2019).

[82] Simon Plazotta and Jonathan Zinsl, *High-frequency limit of non-autonomous gradient flows of functionals with time-periodic forcing*, Journal of Differential Equations, 261 (2016), pp. 6806–6855.

[83] Hannes Risken, *Fokker-planck equation*, in The Fokker-Planck Equation, Springer, 1996, pp. 63–95.

[84] Riccarda Rossi, Alexander Mielke, and Giuseppe Savaré, *A metric approach to a class of doubly nonlinear evolution equations and applications*, Annali della Scuola Normale Superiore di Pisa - Classe di Scienze, Ser. 5, 7 (2008), pp. 97–169.

[85] Riccarda Rossi and Giuseppe Savaré, *Tightness, integral equicontinuity and compactness for evolution problems in banach spaces*, Annali della Scuola Normale Superiore di Pisa-Classe di Scienze-Serie V, 2 (2003), p. 395.

[86] Riccarda Rossi, Giuseppe Savaré, Antonio Segatti, and Ulisse Stefanelli, *Weighted Energy-Dissipation principle for gradient flows in metric spaces*, arXiv preprint arXiv:1801.04988, (2018).

[87] Filippo Santambrogio, *Optimal Transport for Applied Mathematicians - Calculus of Variations, PDEs, and Modeling*, Birkhauser, Basel, 1st ed. 2015 ed., 2015.

[88] ANTONIO Segatti, *A variational approach to gradient flows in metric spaces*, Boll. Unione Mat. Ital.(9), 6 (2013), pp. 765–780.

[89] Karl-Theodor Sturm, *Convex functionals of probability measures and nonlinear diffusions on manifolds*, Journal de mathématiques pures et appliquées, 84 (2005), pp. 149–168.

[90] Vidar Thomée, *Galerkin Finite Element Methods for Parabolic Problems*, vol. 25, Springer Berlin Heidelberg, 2 ed., 2006.

[91] Chad M Topaz, Andrea L Bertozzi, and Mark A Lewis, *A nonlocal continuum model for biological aggregation*, Bulletin of mathematical biology, 68 (2006), pp. 1601–1623.

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

[93] ——, *Optimal Transport: Old and New*, vol. 338, Springer Science & Business Media, 2008.

[94] Michael Westdickenberg and Jon Wilkening, *Variational particle schemes for the porous medium equation and for the system of isentropic Euler equations*, ESAIM: Mathematical Modelling and Numerical Analysis, 44 (2010), pp. 133–166.

[95] Shixin Xu, Ping Sheng, and Chun Liu, *An energetic variational approach for ion transport*, arXiv preprint arXiv:1408.4114, (2014).

[96] Jonathan Zinsl and Daniel Matthes, *Exponential convergence to equilibrium in a coupled gradient flow system modeling chemotaxis*, Analysis & PDE, 8 (2015), pp. 425–466.

[97] ——, *Transport distances and geodesic convexity for systems of degenerate diffusion equations*, Calculus of Variations and Partial Differential Equations, 54 (2015), pp. 3397–3438.