Operator-splitting schemes for degenerate, non-local, conservative-dissipative systems

Duong, Manh Hong; Adams, Daniel; dos Reis, Goncalo

DOI:
10.3934/dcds.2022109

Document Version
Peer reviewed version

Citation for published version (Harvard):
Duong, MH, Adams, D & dos Reis, G 2022, 'Operator-splitting schemes for degenerate, non-local, conservative-dissipative systems', *Discrete and Continuous Dynamical Systems - Series A*, vol. 42, no. 11, pp. 5453-5486. https://doi.org/10.3934/dcds.2022109

Operator-splitting schemes for degenerate, non-local, conservative-dissipative systems

Daniel Adams[a,*] d.t.s.adams@sms.ed.ac.uk
Manh Hong Duong[b†] h.duong@bham.ac.uk
Gonçalo dos Reis[c,d,‡] G.dosReis@ed.ac.uk

[a] Maxwell Institute for Mathematical Sciences School of Mathematics University of Edinburgh Edinburgh UK EH9 3FD
[b] School of Mathematics, University of Birmingham, Birmingham B15 2TT, UK
[c] School of Mathematics, University of Edinburgh, The King's Buildings, Edinburgh, UK
[d] Centro de Matemática e Aplicações (CMA), FCT, UNL, Portugal

Abstract

In this paper, we develop a natural operator-splitting variational scheme for a general class of non-local, degenerate conservative-dissipative evolutionary equations. The splitting-scheme consists of two phases: a conservative (transport) phase and a dissipative (diffusion) phase. The first phase is solved exactly using the method of characteristic and DiPerna-Lions theory while the second phase is solved approximately using a JKO-type variational scheme that minimizes an energy functional with respect to a certain Kantorovich optimal transport cost functional. In addition, we also introduce an entropic-regularisation of the scheme. We prove the convergence of both schemes to a weak solution of the evolutionary equation. We illustrate the generality of our work by providing a number of examples, including the kinetic Fokker-Planck equation and the (regularized) Vlasov-Poisson-Fokker-Planck equation.

Keywords: Wasserstein gradient flows; degenerate diffusions; variational principle; operator-splitting methods; non-local partial differential equations; optimal transport; entropic regularisation.

2020 AMS subject classifications:
Primary: 35K15, 35K55, Secondary: 65K05, 90C25

Contents

[*]D.A was supported by The Maxwell Institute Graduate School in Analysis and its Applications, a Centre for Doctoral Training funded by the UK Engineering and Physical Sciences Research Council (grant EP/L016508/01), the Scottish Funding Council, Heriot-Watt University and the University of Edinburgh.

[†]The research of MHD was supported by EPSRC Grants EP/W008041/1 and EP/V038516/1.

[‡]G.d.R. acknowledges support from the *Fundação para a Ciência e a Tecnologia* (Portuguese Foundation for Science and Technology) through the project UIDB/00297/2020 (Centro de Matemática e Aplicações CMA/FCT/UNL).

1 Introduction

1.1 Dissipative systems and Wasserstein gradient flows

In their seminal work [JKO98] Jordan, Otto and Kinderlehrer show that the linear Fokker-Planck Equation (FPE)

$$\partial_t \rho = \operatorname{div}(\rho \nabla f) + \Delta \rho, \tag{1.1}$$

which is the forward Kolmogorov equation of the overdamped Langevin dynamics

$$dX(t) = -\nabla f(X(t))\, dt + \sqrt{2} dW(t), \tag{1.2}$$

can be viewed as a gradient flow of the free energy functional with respect to the Wasserstein metric. Thereby the solution of the FPE can be iteratively approximated by the following minimising movement (steepest descent) scheme: given a time-step $h > 0$ and defining $\rho_h^0 := \rho_0$, then the solution ρ_h^n at the n-th step, $n = 1, ..., \lfloor \frac{T}{h} \rfloor$, is determined as the unique minimiser of the following minimisation problem

$$\min_{\rho} \Big\{ \frac{1}{2h} W_2^2(\rho_h^{n-1}, \rho) + \mathcal{F}_{\text{fpe}}(\rho) \Big\} \quad \text{with} \quad \mathcal{F}_{\text{fpe}}(\rho) := \int f\rho + \rho \log \rho, \tag{1.3}$$

over the space of the probability measures with finite second moment. Since then, the theory of Wasserstein gradient flows on the space of probability measures has made enormous progress, spanning research activity in various branches of mathematics including partial differential equations, probability theory, optimal transportation as well as geometric analysis. It provides a unified framework and tools for studying well-posedness, regularity, stability and quantitative convergence to equilibrium of dissipative systems. Over the last twenty years, many evolutionary PDEs for models in biology, chemistry, mechanics, and physics have been analysed via this framework. Examples include the Fokker-Planck equation, porous medium equations, thin-film equations, nonlinear aggregation-diffusion equations, interface evolutions, as well as pattern formation and evolution [Vil21, AGS08, San15]. More recently, the theory has been extended to other settings such as discrete spaces [Maa11, EPSS21] and quantum evolutions [CM17] and made intimate connections to the large deviation theory of stochastic processes [ADPZ11, DPZ13, MPR14].

1.2 Conservative-dissipative systems

Many important evolutionary equations arising from biology and physics are not gradient flows, but contain both conservative and dissipative dynamics. A prototypical example is the (generalized[1]) Kramers' (or kinetic Fokker-Planck) equation [Kra40, Ris89],

$$\partial_t \rho = \underbrace{\Big(- \operatorname{div}_q(\rho p) + \operatorname{div}_p(\rho \nabla_q V) \Big)}_{\text{conservative part}} + \underbrace{\Big(\operatorname{div}_p(\rho \nabla_p F) + \Delta_p \rho \Big)}_{\text{dissipative part}}, \tag{1.4}$$

[1] In the classical Kramers equation, $F(p) = \frac{p^2}{2}$.

for a density ρ depending on $t \in \mathbb{R}_+, q, p \in \mathbb{R}^d$. In the above equation, we use the notation div_q and similar to indicate that the differential operator acts only on one variable. The Kramers equation is the forward Kolmogorov equation of the underdamped Langevin dynamics

$$d\begin{pmatrix} Q \\ P \end{pmatrix} = \underbrace{\begin{pmatrix} P \\ -\nabla V(Q) \end{pmatrix} dt}_{\text{conservative dynamics}} + \underbrace{\begin{pmatrix} 0 \\ -\nabla F(P)\, dt + \sqrt{2}\, dW_t \end{pmatrix}}_{\text{dissipative dynamics}} \tag{1.5}$$

The Langevin dynamics (1.5) describes the movement of a particle (with unity mass) at position Q and with momentum P under the influence of three forces: an external force field ($-\nabla V(Q)$), a (possibly non-linear) friction ($-\nabla F(P)$) and a stochastic noise ($\sqrt{2}dW_t$). The Kramers equation (1.4) characterizes the time evolution of the probability of finding the particle at time t at position q and with momentum p. Unlike the Fokker-Planck equation (1.1), which is purely dissipative, the Kramers equation (1.4) is a mixture of both conservative and dissipative dynamics. The first part in (1.5) is a deterministic Hamiltonian system with Hamiltonian energy $H(q,p) = p^2/2 + V(q)$. The evolution of this part is reversible and conserves the Hamiltonian. Correspondingly, the first part of (1.4) is also reversible and conserves the expectation of H,

$$\mathcal{H}(\rho) := \int_{\mathbb{R}^{2d}} \rho(q,p) H(q,p)\, dqdp.$$

On the other hand, the second part of (1.5) is an overdamped Langevin dynamics (cf. (1.2)), but only in the p-variable. The corresponding part in (1.4) is precisely a Fokker-Planck equation in p-variable (cf. (1.1)), which is a Wasserstein gradient flow in the p-variable. Because of the mixture of both conservative and dissipative effects, the full Kramers equation (1.4) is not a gradient flow, and the theory of Wasserstein gradient flows, in particular the JKO-minimizing movement scheme (1.3), is not directly applicable.

It is desirable to develop operator-splitting methods for conservative-dissipative systems that reflect the same division between conservative and dissipative effects. Developing structure-preserving schemes is currently of great interest both theoretically and computationally, according to [Ött18] "an important challenge for the future is how the structure of thermodynamically admissible evolution equations can be preserved under time-discretization, which is a key to successful numerical calculations". For the Kramers equation, such an operator-splitting scheme is introduced in [DPZ14]. However, the scheme developed in that work uses a complicated optimal transport cost functional for the dissipative part which does not capture the fact that it is simply a Wasserstein gradient flow in the momentum variable. More recently in [CL17] the authors introduce an operator-splitting scheme for a non-degenerate conservative-dissipative non-local-nonlinear diffusion equation

$$\partial_t \rho + \mathrm{div}\big(\rho b[\rho]\big) = \Delta P(\rho) + \mathrm{div}\big(\rho \nabla f\big),$$

where $b[\rho]$ is a divergence-free vector field for each ρ, and P is the non-linear pressure function. The above equation does not cover the Kramers equation since the latter is a degenerate diffusion, in which the Laplacian only acts on the momentum variables. The degeneracy of the Kramers equation can also be seen in the Langevin dynamics (1.5) where the noise is present only in the momentum but not on the position variables. A natural question arises

Can we develop operator-splitting schemes for non-local, degenerate conservative-dissipative systems?

In this paper, we address this question by providing a simple operator-splitting scheme for a general class of non-local, degenerate, conservative-dissipative systems.

1.3 Our contribution

We consider a general class of degenerate, non-local, conservative-dissipative evolutionary equations of the form

$$\partial_t \rho + \mathrm{div}\big(\rho b[\rho]\big) = \mathrm{div}\big(A\big(\nabla \rho + \rho \nabla f\big)\big), \quad \rho(0,\cdot) = \rho_0(\cdot), \tag{1.6}$$

where the unknown ρ is a time dependent probability distribution on $[0,T] \times \mathbb{R}^d$, $A \in \mathbb{R}^{d\times d}$ is a semi-positive definite (symmetric) matrix (possibly degenerate), $f : \mathbb{R}^d \to \mathbb{R}$ a given energy potential, $b : \mathcal{P}(\mathbb{R}^d) \times \mathbb{R}^d \to \mathbb{R}^d$ a divergence free non-local vector field, and the probability density $\rho_0 \in \mathcal{P}_2^r(\mathbb{R}^d)$ is the initial condition. Here, and throughout we denote $\mathcal{P}(\mathbb{R}^d)$ the space of Borel probability measures on $\mathbb{R}^d$, $\mathcal{P}_2(\mathbb{R}^d)$ those with finite second moment, and $\mathcal{P}_2^r(\mathbb{R}^d)$ as those which are additionally absolutely continuous (with respect to the Lebesgue measure). Equation (1.6) can be viewed as the forward Kolmogorov equation describing the time evolution of the distribution ρ associated to the stochastic process X satisfying the following SDE of McKean type

$$dX(t) = b[\rho_t](X(t))dt - A\nabla f(X(t))dt + \sqrt{2\sigma} dW(t), \qquad \rho_t = \mathrm{Law}(X_t), \tag{1.7}$$

for a constant diffusion matrix σ, with $\sigma\sigma^T = A$. This serves as a general model for the dynamic limit of interacting particles, evolving under the influence of (weakly, self-stabilising) interaction force $b[\rho]$ depending on the own law of the process, and a potential drift ∇f, whilst being perturbed by noise $W(t)$. Like the Kramers equation, (1.6) contains both conservative and dissipative effects. The conservative part is represented via the divergence-free vector field (the transport part in the left-hand side of (1.6)), in particular implying that the entropy will be preserved under this part. On the other hand, the dissipative part is given by the right hand side of (1.6), which resembles a A-Wasserstein gradient flow [Lis09] (but note that A can be degenerate). The aim of this paper is to develop operator-splitting schemes, which capture the conservative-dissipative splitting and take into account the degeneracy of the diffusion matrix, for solving (1.6).

Our operator-splitting scheme can be summarised as follows (details follow in Section 2).

The Operator-splitting scheme. We split the dynamics described in (1.6) by two phases:

1. *Conservative (transport) phase*: for a given ρ, we solve the conservative part, which is simply a transport equation, using the method of characteristics
$$\partial_t \rho + \operatorname{div}\big(\rho b[\rho]\big) = 0$$
The existence of a solution to the above equation under a transport/push-forward map is guaranteed by DiPerna Lions theory [DL89].

2. *Dissipative (diffusion) phase*: we solve the dissipative (diffusion) part using a JKO-miminimizing movement scheme
$$\partial_t \rho = \operatorname{div}\big(A\big(\nabla \rho + \rho \nabla f\big)\big). \tag{1.8}$$
We emphasize again that we allow the diffusion matrix A to be degenerate. Because of the degeneracy of A, the JKO-scheme (1.3) using the A-weighted Wasserstein matrix developed in [Lis09] is not applicable. To overcome this difficulty, we use a simple idea, that is to use a small perturbation of A to get a symmetric positive definite matrix. The key novelty here is that we perturb A by $A_h := A + hI$ where h is the time-step in the discretisation scheme. Note that this perturbation is equivalent to adding a small amount of noise into each component of the stochastic dynamics (1.7). Thereby, we solve the dissipative (diffusion) equation iteratively using the minimizing movement scheme: ρ_h^{n+1} is determined as the unique minimizer of the minimization problem
$$\min_{\rho} \Big\{ \frac{1}{2h} W_{c_h}(\rho, \rho_h^n) + \int f\rho + \rho \log \rho \Big\},$$
where (see Section 2; $\Pi(\mu,\nu)$ denotes the set of transport plans between the measures μ, ν)
$$W_{c_h}(\mu,\nu) = \inf_{\gamma \in \Pi(\mu,\nu)} \int_{\mathbb{R}^{2d}} \big\langle (A+hI)^{-1}(x-y), (x-y) \big\rangle \, d\gamma(x,y).$$

Our main result, Theorem 2.4, establishes the convergence of the above splitting-scheme to a weak solution of (1.6) as the time step h tends to zero. Our operator-splitting scheme is simple and natural capturing the conservative-dissipative splitting of the dynamics, in particular the expectation that the dissipative part is a A-weighted Wasserstein gradient flow. Furthermore, motivated by the efficiency of entropic regularisation methods in computational performances, in Theorem 4.2 we also provide an entropic regularisation of the above scheme. We expect that the entropic regularisation scheme will be useful when one performs numerical simulations although we do not pursue it here. Our result offers a unified approach to establish existence results for a wide class of degenerate, non-local, conservative-dissipative systems. In fact, the class of (1.6) is rich and includes many cases of interest: the linear and kinetic Fokker-Planck [Ris89], the (regularised) Vlasov Poisson Fokker-Planck [HJ13], the linear Wigner Fokker-Planck [AGG$^+$12], and higher-order degenerate diffusions approximating the generalised Langevin and generalised Vlasov equations [OP11, Duo15]. We will discuss in details these concrete applications in Section 5.

Comparison to existing literature. There is a vast literature on operator-splitting methods for solving PDEs, see e.g. [GO16]. We now compare our work with the most relevant literature where the dissipative dynamics involves a Wassertein-type gradient flow. The closest paper to ours are [CDPS17, Ber18] where the authors consider equations of the form (1.6) and introduce similar operator-splitting schemes. However, these papers are limited to non-degenerate diffusion matrices A ($A = I$ in these papers). In fact, [Ber18] does not deal with mixed dynamics, the splitting is carried out at the level of the gradient flow. In [YB13], the authors implement a numerical method that splits an aggregation-diffusion equation, where they exploit its transport structure using a Lagrangian method for the aggregation part, and employ an implicit finite-difference scheme for the diffusion part. Our splitting method is of a different nature, in that we would treat [YB13][Equation (1.1)] as a dissipative equation with no conservative dynamics. Other works that also develop operator-splitting schemes for degenerate PDEs are [CG04, MS20a], however these works only deal with a linear, local conservative dynamics and use a rather involved splitting mechanism. Several papers including [Hua00, DPZ14, DT18, ADdR21] also develop JKO-type minimizing movement schemes for degenerate diffusion equations; however these papers use one-step

schemes where the cost functions are often non-homogeneous, time-step dependent and do not induce a metric. We also mention recent works in which operator-splitting methods have been investigated for partial differential equations containing a Wasserstein gradient flow part and a non-Wasserstein part. The papers [BA15, DL19] construct operator-splitting schemes for fractional Fokker-Planck equations, in which the transport phase is solved by a JKO-type minimizing movement scheme while the fractional diffusion is solved exactly by convolution with the fractional heat kernel. More recently, [LWW21] builds operator-splitting scheme for reaction-diffusion systems with detailed balanced based on an energetic variational formulation of the systems.

Future work. From a modelling perspective the non-local term b captures the interactions between a large ensemble of particles. In this case, it takes the form of a convolution between the density distribution and a certain kernel, and our assumptions require the kernel to be uniformly bounded and Lipschitz. However, many fundamental models of interacting particle systems compose of singular interaction kernels [JW18, Ser20]. This leads to the natural and challenging question: can our method be generalized to deal with singular interaction kernels? In this paper, we demonstrate via the regularised Vlasov-Poisson-Fokker-Planck equation that our method is applicable when one regularises the Coulomb interaction (see Section 5.3) Another interesting question is whether we can use the variational structure developed in this paper to study exponential convergence to the equilibrium of degenerate PDEs of the form (1.6). This is related to the hypocoercivity theory introduced by Villani [Vil09], and a method using variational structures would provide further physical insight to the theory. These themes are left for future work.

1.4 Organisation of the paper

In Section 2 we set up the notations and present the operator-splitting scheme, assumptions, and the main result of this paper. The proof of the main result is given in Section 3. In Section 4 we show how the scheme can be regularised. Section 5 provides several explicit examples to which our work can be applied to. Finally, the Appendix contains detailed computations and proofs of technical results.

2 The operator-splitting scheme, assumptions and our main result

In this section, we first introduce notations that will be used throughout the paper, then we present the operator-splitting scheme and assumptions and finally, we state the main result of this paper, Theorem 2.4.

2.1 Notation

Throughout $d \in \mathbb{N}$ will be the dimension of the space. A fixed $T > 0$ denotes the length of the time interval we consider. Throughout, C denotes a constant whose value may change without indication and depends on the problem's involved constants, but, critically, it is independent of key parameters of this work, namely the time step $h > 0$ and number of iterates $N \in \mathbb{N}$ of the scheme introduced in Section 1. The Euclidean inner product between two vectors $x, y \in \mathbb{R}^d$ will be written as $x \cdot y$ or sometimes $\langle x, y \rangle$. We write $\|\cdot\|$ as the Euclidean norm on $\mathbb{R}^d$, and $|\cdot|$ when $d = 1$. The symbol $\|\cdot\|$ is also used as the 2-norm on $\mathbb{R}^{d\times d}$. For a matrix A let A^T be its transpose.

Let $\Omega \subseteq \mathbb{R}^d$, we write $|\Omega|$ as its $d-$dimensional Lebesgue measure. The space of Lebesgue $m-$integrable functions on Ω is denoted by $L^m(\Omega)$. The Sobolev space of functions in $L^1(\Omega)$ with first weak derivatives also in $L^1(\Omega)$ is denoted $W^{1,1}(\Omega)$. We say that $f \in L^1_{\text{loc}}(\mathbb{R}^d)$ if $f \in L^1(\Omega)$ for any compact $\Omega \subset \mathbb{R}^d$. We define the space $f \in W^{1,1}_{\text{loc}}(\mathbb{R}^d)$ similarly. The supremum norm $\|\cdot\|_{\infty,\Omega}$ of a vector field $\phi : \Omega \to \mathbb{R}^d$, or a function $\phi : \Omega \to \mathbb{R}$, is used to denote $\sup_{x\in\Omega} \|\phi(x)\|$, $\sup_{x\in\Omega} |\phi(x)|$ respectively, when $\Omega = \mathbb{R}^d$ we just write $\|\cdot\|_\infty$. We use the Landau "big-O" notation $\phi(h) = O(\varphi(h))$, for functions $\phi, \varphi : \mathbb{R}_+ \to \mathbb{R}$ to mean that there exists $C, h_0 > 0$ such that $|\phi(h)| \le C\varphi(h)$ for all $h < h_0$ and we say a matrix $B \in \mathbb{R}^{d\times d}$ is $O(h)$ if $\max_{i,j} |B_{i,j}| \le Ch$.

Let $A, B \subseteq \mathbb{R}^d$, define $C^k(A; B)$ as the k-times continuously differentiable functions from A to B with continuous k^{th} derivative. Define $C^\infty_c(A; B)$ as the set of infinitely differentiable functions from A to B with compact support. We specifically write $C^\infty_c(\mathbb{R}^d)$ to denote infinitely differentiable functions from $\mathbb{R}^d$ to $\mathbb{R}$ with compact support. Let $C_b(\mathbb{R}^d)$ be the set of continuous bounded functions from $\mathbb{R}^d$ to $\mathbb{R}$. Let $\nabla\phi$, $\Delta\phi$, and $\nabla^2\phi$ be the gradient, Laplacian, and Hessian respectively, of a sufficiently smooth function $\phi : \mathbb{R}^d \to \mathbb{R}$. For a sufficiently smooth vector field $\eta : \mathbb{R}^d \to \mathbb{R}^d$ let $\text{div}(\eta)$, and $D\eta$ be its divergence and Jacobian respectively. We call 'id' the identity map on any space.

Denote the space of Borel probability measures on $\mathbb{R}^d$ as $\mathcal{P}(\mathbb{R}^d)$. The second moment M of a measure $\rho \in \mathcal{P}(\mathbb{R}^d)$ is defined as

$$\mathcal{P}(\mathbb{R}^d) \ni \rho \mapsto M(\rho) := \int_{\mathbb{R}^d} \|x\|^2 \rho(dx). \tag{2.1}$$

The set of probability measures with finite second moments is denoted $\mathcal{P}_2(\mathbb{R}^d)$,

$$\mathcal{P}_2(\mathbb{R}^d) := \{\rho \in \mathcal{P}(\mathbb{R}^d) \ : \ M(\rho) < \infty\}. \tag{2.2}$$

Define $\mathcal{P}_2^r(\mathbb{R}^d)$ as those $\rho \in \mathcal{P}_2(\mathbb{R}^d)$ which are absolutely continuous. Throughout, when a measure is said to be 'absolutely continuous' we implicitly mean with respect to the Lebesgue measure . We will use the same symbol ρ to denote a measure $\rho \in \mathcal{P}_2^r(\mathbb{R}^d)$ as well as its associated density. Define H to be the negative of Boltzmann entropy,

$$\mathcal{P}(\mathbb{R}^d) \ni \rho \mapsto H(\rho) := \begin{cases} \int_{\mathbb{R}^d} \rho \log \rho, & \text{if } \rho \in \mathcal{P}_2^r(\mathbb{R}^d) \\ +\infty, & \text{otherwise,} \end{cases} \tag{2.3}$$

which throughout we will just refer to as the entropy. Also define the positive part of the entropy as

$$\mathcal{P}(\mathbb{R}^d) \ni \rho \mapsto H_+(\rho) := \begin{cases} \int_{\mathbb{R}^d} \max\{\rho \log \rho, 0\}, & \text{if } \rho \in \mathcal{P}_2^r(\mathbb{R}^d) \\ +\infty, & \text{otherwise,} \end{cases} \tag{2.4}$$

and the negative part of the entropy as

$$\mathcal{P}(\mathbb{R}^d) \ni \rho \mapsto H_-(\rho) := \begin{cases} \int_{\mathbb{R}^d} |\min\{\rho \log \rho, 0\}|, & \text{if } \rho \in \mathcal{P}_2^r(\mathbb{R}^d) \\ +\infty, & \text{otherwise.} \end{cases} \tag{2.5}$$

The set of transport plans between given measures $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ is denoted by $\Pi(\mu, \nu) \subset \mathcal{P}_2(\mathbb{R}^{2d})$. That is, for $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$, $\gamma \in \Pi(\mu, \nu)$ if $\gamma(\mathcal{B} \times \mathbb{R}^d) = \mu(\mathcal{B})$ and $\gamma(\mathbb{R}^d \times \mathcal{B}) = \nu(\mathcal{B})$ for all Borel sets $\mathcal{B} \subset \mathbb{R}^d$. Let $\Pi^r(\mu, \nu)$ be those $\gamma \in \Pi(\mu, \nu)$ which are absolutely continuous. We denote a sequence of probability measures indexed by $k \in \mathbb{N}$ as $\{\mu_k\}_{k \in \mathbb{N}}$ which we relax to $\{\mu_k\}$. We use the symbol $\rightharpoonup$ to mean the weak convergence of measures, that is $\rho_k \rightharpoonup \rho$ (weakly) if

$$\lim_{k \to \infty} \int f d\rho_k = \int f d\rho, \ \ \forall f \in C_b(\mathbb{R}^d). \tag{2.6}$$

We also recall that if it is known that $\rho \in \mathcal{P}(\mathbb{R}^d)$ then the weak convergence (2.6) is equivalent to narrow convergence, that is convergence against $C_c^\infty(\mathbb{R}^d)$ functions. For any two subsets $P, Q \subset \mathcal{P}_2(\mathbb{R}^d)$ we denote $\Pi(P, Q)$ as the set of transport plans whose marginals lie in P and Q respectively. For a vector field $\eta : \mathbb{R}^d \to \mathbb{R}^d$ and measure $\mu \in \mathcal{P}(\mathbb{R}^d)$ we write $(\eta)_\# \mu$ as the push-forward of μ by η. For any probability measure γ and function c on $\mathbb{R}^{2d}$ we write

$$(c, \gamma) := \int_{\mathbb{R}^{2d}} c(x, y) d\gamma(x, y).$$

We use the symbol $*$ to denote the convolution, that is for a vector field $K : \mathbb{R}^{d_1} \to \mathbb{R}^{d_1}$ and a measure $\rho \in \mathcal{P}(\mathbb{R}^{d_2})$, $K * \rho : \mathbb{R}^{d_1} \to \mathbb{R}^{d_1}$ is defined as

$$K * \rho(x) := \int_{\mathbb{R}^{d_2}} K(x - x') \rho(x', z) dx' dz, \tag{2.7}$$

where $x, x' \in \mathbb{R}^{d_1}$ and $z \in \mathbb{R}^{d_2 - d_1}$. Lastly, the 2-Wasserstein distance on $\mathcal{P}_2(\mathbb{R}^d)$ is denoted by W_2.

2.2 The operator-splitting scheme

We now present our operator-splitting scheme for solving (1.6). Denote the free energy $\mathcal{F} : \ \mathcal{P}_2^r(\mathbb{R}^d) \to \mathbb{R}$ as the sum of the potential energy and the entropy

$$\mathcal{F}(\rho) := F(\rho) + H(\rho),$$

where

$$F(\rho) := \int_{\mathbb{R}^d} \rho f dx, \quad \text{and} \quad H(\rho) := \int_{\mathbb{R}^d} \rho \log(\rho) dx.$$

Operator-splitting scheme: Let $T > 0$ denotes the terminal time and $\rho_0 \in \mathcal{P}_2^r(\mathbb{R}^d)$ be given, with $\mathcal{F}(\rho_0) < \infty$. Let $h > 0$, $N \in \mathbb{N}$ be such that $hN = T$, and let $n \in \{0, \dots, N-1\}$. Set $\rho_h^0 = \tilde{\rho}_h^0 = \rho_0$. Given ρ_h^n, our operator-splitting to determine ρ_h^{n+1} consists of two phases

1. *Conservative (transport) phase*: first we introduce the push forward by the Hamiltonian flow as

$$\tilde{\rho}_h^{n+1} = X_h^n(h, \cdot)_\# \rho_h^n, \tag{2.8}$$

where $X_h^n : \mathbb{R}^+ \times \mathbb{R}^d \to \mathbb{R}^d$ solves the ODE

$$\begin{cases} \partial_t X_h^n = b[\rho_h^n] \circ X_h^n, \\ X_h^n(0, \cdot) = \text{id}. \end{cases} \tag{2.9}$$

2. *Dissipative (diffusion) phase*: next, define ρ_h^{n+1} as the minimizer of the following JKO-type optimal transport minimization problem

$$\rho_h^{n+1} = \operatorname*{argmin}_{\rho\in\mathcal{P}_2^r(\mathbb{R}^d)} \Big\{ \frac{1}{2h} W_{c_h}(\tilde{\rho}_h^{n+1}, \rho) + \mathcal{F}(\rho) \Big\}, \tag{2.10}$$

where W_{c_h} is a Kantorovich optimal transport cost functional, defined for $h > 0$ as

$$W_{c_h}(\mu,\nu) := \inf_{\gamma\in\Pi(\mu,\nu)} \int c_h(x,y) d\gamma(x,y), \tag{2.11}$$

with the cost function $c_h : \mathbb{R}^{2d} \to \mathbb{R}$ given by

$$c_h(x,y) := \big\langle A_h^{-1}(x-y), (x-y) \big\rangle, \tag{2.12}$$

for the matrix $A_h \in \mathbb{R}^{d\times d}$ defined as

$$A_h := A + hI. \tag{2.13}$$

Note that since A is symmetric positive semi-definite, the addition of hI to A guarantees that A_h is symmetric positive definite (see Lemma 3.3). Hence, c_h is well defined for all $h > 0$ and $\sqrt{c_h}$ defines a metric on $\mathbb{R}^d$, which in-turn means $W_{c_h}^{1/2}$ defines a metric on $\mathcal{P}_2(\mathbb{R}^d)$. This is precisely a A_h-weighted Wasserstein distance [Lis09]. The above perturbation can be also effectively achieved by adding a small noise $\sqrt{2h}dW(t)$ to the SDE (1.7). We mention that if the matrix A is invertible then there is no need to perform the perturbation. Instead we can adopt the scheme with $c_h(x,y) = c(x,y) := \big\langle A^{-1}(x-y), (x-y) \big\rangle$ and all results would remain true. This is the case for the Linear Wigner Fokker-Planck, see Section 5.1.

For each $n \in \{0,\dots,N\}$ we denote $\tilde{\gamma}_h^{n,c}, \tilde{\gamma}_h^n \in \Pi(\tilde{\rho}_h^n, \rho_h^n)$, as the following optimal couplings (respectively)

$$W_{c_h}(\tilde{\rho}_h^n, \rho_h^n) = \int_{\mathbb{R}^{2d}} c_h(x,y) d\tilde{\gamma}_h^{n,c}(x,y), \qquad W_2^2(\tilde{\rho}_h^n, \rho_h^n) = \int_{\mathbb{R}^{2d}} \|x-y\|^2 d\tilde{\gamma}_h^n(x,y), \tag{2.14}$$

and for $n \in \{0,\dots,N-1\}$ we define $\gamma_h^n \in \Pi(\rho_h^n, \tilde{\rho}_h^{n+1})$ as the optimal coupling

$$W_2^2(\rho_h^n, \tilde{\rho}_h^{n+1}) = \int_{\mathbb{R}^{2d}} \|x-y\|^2 d\gamma_h^n(x,y). \tag{2.15}$$

The optimal couplings in (2.14) and (2.15) are all well defined, see Lemma A.3. Throughout this work we will adopt the notation that $t_n = nh$ for $n \in \{0,\dots,N\}$. Consider the following piece-wise constant in time interpolations of $\{\rho_h^n\}_{n=0}^N$

$$\rho_h(t,\cdot) := \rho_h^{n+1} \text{ for } t \in [t_n, t_{n+1}), \tag{2.16}$$

and of $\{\tilde{\rho}_h^n\}_{n=0}^N$

$$\tilde{\rho}_h(t,\cdot) := \tilde{\rho}_h^{n+1} \text{ for } t \in [t_n, t_{n+1}), \tag{2.17}$$

and consider the interpolation of $\{\tilde{\rho}^n\}_{n=0}^N$, which continuously follows the conservative dynamics

$$\rho_h^\dagger(t,\cdot) := \big(X_h^n(t-t_n,\cdot)\big)_\# \rho_h^n \text{ for } t \in [t_n, t_{n+1}), \tag{2.18}$$

so that for $t \in [t_n, t_{n+1})$, $\rho_h^\dagger(t) = \mu(t-t_n)$ where μ is the solution of the continuity equation (see Lemma 3.1)

$$\begin{cases} \partial_t \mu(t,\cdot) + \operatorname{div}\big(\mu(t,\cdot) b[\rho_h^n]\big) = 0 \\ \mu(t,\cdot)|_{t=0} = \rho_h^n. \end{cases} \tag{2.19}$$

We now introduce assumptions on the potential f, the non-local vector field b, and the diffusion matrix A. Under these assumptions we will prove the well-posedness of the splitting scheme and the convergence of the interpolations (2.16)-(2.18) to a weak solution of (1.6).

Assumption 2.1. The potential energy $f \in C^1(\mathbb{R}^d)$ is assumed to be non-negative $f(x) \geq 0$, and Lipschitz, that is there exists a constant $C > 0$ such that for any $x, y \in \mathbb{R}^d$

$$|f(x) - f(y)| \leq C\|x-y\|. \tag{2.20}$$

For the non-local drift $b : \mathcal{P}(\mathbb{R}^d) \times \mathbb{R}^d \to \mathbb{R}^d$, we assume that there exists $C > 0$ such that for any $\mu \in \mathcal{P}_2(\mathbb{R}^d)$

$$\|b[\mu](x)\| \leq C\big(1 + \|x\|\big),\ \forall x \in \mathbb{R}^d, \quad b[\mu] \in W^{1,1}_{\text{loc}}(\mathbb{R}^d), \qquad \operatorname{div}(b[\mu]) = 0. \tag{2.21}$$

Moreover, we assume there exists $C > 0$ for all $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$

$$\int_{\mathbb{R}^d} \|b[\nu](x) - b[\mu](x)\|^2 d\nu(x) \leq C W_2^2(\nu,\mu). \tag{2.22}$$

Lastly assume A is a semi-positive definite (symmetric), constant matrix $A \in \mathbb{R}^{d\times d}$.

Remark 2.2 (Commenting on the assumptions)**.** The Lipschitz assumption on f is standard when working on the space of probability measures with finite second moments, particularly ensuring that the free energy functional is well-defined. In terms of the assumptions on the non-local vector field b, (2.21) implies well-posedness of the transport problem via DiPerna-Lions theory [DL89]. Moreover, imposing the regularity in the measure component (2.22) allows us to obtain upper-bounds for some error terms when proving the convergence of the scheme to a weak solution of (1.6). Note that when b takes the form of a convolution with an interaction kernel, (2.22) is satisfied when the kernel is uniformly bounded, Lipschitz and differentiable, which are the cases for the examples in Section 5. Note that the above assumptions have been also made in [CL17].

We now make the definition of a weak solution to (1.6) precise.

Definition 2.3 (Weak solution)**.** The curve $\rho : [0,T] \to \mathcal{P}_2^r(\mathbb{R}^d)$, $t \mapsto \rho(t,\cdot)$, is called a weak solution to the general evolution equation (1.6) if for all $\varphi \in C_c^\infty([0,T]\times\mathbb{R}^d)$ we have

$$\int_0^T \int_{\mathbb{R}^d} \rho(t,x)\Big(\partial_t\varphi(t,x) + \big(b[\rho(t,\cdot)](x) - A\nabla f(x)\big)\cdot\nabla\varphi(t,x) + \operatorname{div}\big(A\nabla\varphi(t,x)\big)\Big)dxdt + \int_{\mathbb{R}^d}\rho^0(x)\varphi(0,x)dx = 0 \tag{2.23}$$

The main (abstract) result of this work is the following theorem which gives the existence of weak solutions of the evolution equation (1.6). We do not deal with uniqueness here, but in principle, it can be obtained via displacement convexity arguments and an exponential in time contraction on the W_2 distance between two solutions started from different initial data, cf. [Lab17].

Theorem 2.4. Let $\rho_0 \in \mathcal{P}_2^r(\mathbb{R}^d)$ satisfy $\mathcal{F}(\rho_0) < \infty$. Let $h > 0$, $N \in \mathbb{N}$ with $hN = T$, and let $\{\rho_h^n\}_{n=0}^N, \{\tilde\rho_h^n\}_{n=0}^N$ be the solution of the scheme (2.8)-(2.10). Define the piecewise constant interpolations $\rho_h, \tilde\rho_h$ by (2.16)-(2.17) and the interpolation $\rho_h^\dagger$ by (2.18). Suppose that Assumption 2.1 holds. Then

(i) for each $t \in [0,T]$ as $h \to 0$ ($N \to \infty$ abiding by $hN = T$) we have

$$\rho_h(t,\cdot), \tilde\rho_h(t,\cdot), \rho_h^\dagger(t,\cdot) \underset{h\to 0}{\longrightarrow} \rho(t) \quad \text{in} \quad L^1\big(\mathbb{R}^d\big). \tag{2.24}$$

(ii) Moreover, there exists a map $[0,T] \ni t \mapsto \rho(t,\cdot)$ in $\mathcal{P}_2^r(\mathbb{R}^d)$ such that

$$\lim_{h\to 0}\sup_{t\in[0,T]} \max\Big\{W_2(\rho_h(t,\cdot),\rho(t,\cdot)), W_2(\tilde\rho_h(t,\cdot),\rho(t,\cdot)), W_2(\rho_h^\dagger(t,\cdot),\rho(t,\cdot))\Big\} = 0. \tag{2.25}$$

The ρ maps appearing in the above limits are weak solutions of the evolution equation (1.6) in the sense of Definition 2.3.

Note that the convergence (2.24) is stronger than weak $L^1((0,T)\times\mathbb{R}^d)$ convergence. The proof of the above theorem is given in Section 3.

Remark 2.5. If one were to instead consider the evolution equation, for a non-linear function P,

$$\partial_t\rho + \operatorname{div}(\rho b[\rho]) = \operatorname{div}\Big(A\big(\nabla P(\rho) + \rho\nabla f\big)\Big),$$

then following the strategy in [CL17], to deal with the non-linear term, we expect one could construct a similar scheme to the one detailed above by adjusting the free energy functional $\mathcal{F}$. We leave this for now to not over complicate the presentation.

3 Proof of the Main Result

The objective of this section is to prove the main result, Theorem 2.4. Once a suitable optimal transport cost functional has been identified, the proof of the convergence of the discrete variational approximation scheme to a weak solution of the evolutionary equation is now a well-established procedure following the celebrated strategy of [JKO98]: firstly we prove the well-posedness of the scheme, then we derive discrete Euler-Lagrange equations for the minimisers of (2.10) and necessary a priori estimates, and finally we prove the convergence of the scheme to a weak solution of (1.6). An additional step in our proof for the constructed operator-splitting scheme is to combine the two (conservative and dissipative) phases together. Since the outcome of the conservative phase $\tilde\rho_h^{n+1}$ becomes an input of the dissipative phase, we need to show that the second moments, the free-energy functionals and the distances involved, with respect to this

density are controllable. This is where we make use of the divergence-free property and assumptions of the non-local vector field b. To this end, our proof follows the methods in [DPZ14, CL17] and we will omit similar computations.

Recall from Section 2.2 the definitions of the sequences $\rho_h^n, \tilde{\rho}_h^n$, interpolations $\rho_h, \tilde{\rho}_h, \rho_h^\dagger$, and optimal couplings $\tilde{\gamma}_h^{n,c}$, $\tilde{\gamma}_h^n$, γ_h^n. Also recall that the constant $C \geq 0$ that appears will be independent of h and $n \in \{0, \dots N\}$, but may depend on the final time T. The following results hold under the assumptions of Theorem 2.4, and for all $0 < h < 1$, note that we are ultimately interested in the case where $h \to 0$.

3.1 Preliminary results and well-posedness

The main result here is that the scheme proposed in Section 2.2 is well-posed, this is shown using the direct method of calculus of variations with respect to the weak topology.

We also make some preliminary observations on the matrix A_h, and on solutions to the continuity equation which will be useful later on.

3.1.1 The transport equation

By our assumptions on b, we can use DiPerna-Lions theory [DL89] to conclude that there exists a solution to the ODE (2.9), which when pushing forward the initial density solves the continuity equation (2.19). Moreover, we note that the conservative dynamics preserves the entropy.

Lemma 3.1. Let $\rho_h^n \in \mathcal{P}_2^r(\mathbb{R}^d)$. Then the following results hold for any $n \in \{0, \dots, N-1\}$.

(i) There exists a unique $X_h^n : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}^d$, such that $X_h^n(0, \cdot) = \mathrm{id}$, and for a.e. $x \in \mathbb{R}^d$ the map $t \mapsto X_h^n(t, x)$ solves (2.9),

$$X_h^n(t, x) = x + \int_0^t b[\rho_h^n] \circ X_h^n(s, x) ds.$$

Moreover, $\mathbb{R}^d \ni x \mapsto X(\cdot, x) \in L^1_{\mathrm{loc}}(\mathbb{R}^d; C(\mathbb{R}))$, and for a.e. $x \in \mathbb{R}^d$ the map $\mathbb{R}^+ \ni t \mapsto X_h^n(t, x) \in C^1(\mathbb{R})$.

(ii) For $t \in [t_n, t_{n+1})$, $\rho_h^\dagger(t, \cdot)$ solves the continuity equation (2.19) over the interval $[0, h)$.

(iii) We have the following entropy preservation identities

$$H\big(\rho_h^\dagger(t, \cdot)\big) = H(\rho_h^n) \; \forall t \in [t_n, t_{n+1}), \qquad H(\tilde{\rho}_h^{n+1}) = H(\rho_h^n). \tag{3.1}$$

Proof. Since $b[\rho_h^n]$ satisfies Assumption 2.1, (i) and (ii) follow by [DL89, Theorem III.1]. In regard to (iii), note that for all $t \geq 0$ the map $X_h^n(t, \cdot)$ preserves the Lebesgue measure since b is a divergence free vector field. The result is thus immediate. □

The following lemma bounds the change of the distribution under the Hamiltonian dynamics over the interval $(0, h)$ by its 2nd moments.

Lemma 3.2. The following result holds for any $n \in \{0, \dots, N-1\}$. Let $\rho_h^n \in \mathcal{P}_2^r(\mathbb{R}^d)$. Let μ be the solution of (2.19) over the interval $[0, h]$ and let $0 \leq s_1 \leq s_2 \leq h$. Then

$$W_2^2\big(\mu(s_1, \cdot), \mu(s_2, \cdot)\big) \leq Ch \int_{s_1}^{s_2} \big(1 + M(\mu(s, \cdot))\big) ds. \tag{3.2}$$

Moreover, for any $t \in [t_n, t_{n+1})$, $M(\rho_h^\dagger(t, \cdot)), M(\tilde{\rho}_h(t, \cdot)) < C\big(M(\rho_h^n) + 1\big)$.

Proof. Let μ solve (2.19). For any $0 \leq s_1 \leq s_2 \leq h$, from Benamou-Brenier formula [AGS08, Chapter 8] and (2.21), we have

$$\begin{aligned} W_2^2\big(\mu(s_1, \cdot), \mu(s_2, \cdot)\big) \leq & (s_2 - s_1) \int_{s_1}^{s_2} \int_{\mathbb{R}^d} \|b[\rho_h^n](x)\|^2 \mu(s, x) dx ds \\ \leq & (s_2 - s_1) C \int_{s_1}^{s_2} \int_{\mathbb{R}^d} \big(1 + \|x\|^2\big) \mu(s, x) dx ds \leq hC \int_{s_1}^{s_2} \big(1 + M(\mu(s, \cdot))\big) ds, \end{aligned}$$

which is (3.2). Now consider

$$\begin{aligned}\partial_t M(\mu(t,\cdot)) = \partial_t \int_{\mathbb{R}^d} \|X_h^n(t,x)\|^2 \rho_h^n(x)dx =& 2\int_{\mathbb{R}^d} X_h^n(t,x)\cdot \partial_t X_h^n(t,x)\rho_h^n(x)dx\\ =& 2\int_{\mathbb{R}^d} X_h^n(t,x)\cdot b[\rho_h^n]\circ X_h^n(t,x)\rho_h^n(x)dx\\ \leq& C\int_{\mathbb{R}^d}\big(1+\|X_h^n(t,x)\|^2\big)\rho_h^n(x)dx\\ =& C\int_{\mathbb{R}^d}\big(1+\|x\|^2\big)\big(X_h^n(t,\cdot)\big)_{\#}\rho_h^n(x)dx = C\big(1+M(\mu(t,\cdot))\big).\end{aligned}$$

Employing Grönwall's inequality, we have for any $t\in[0,h]$ (recalling that throughout this article $h\leq T$)

$$M(\mu(t,\cdot)) \leq C\big(M(\mu(0,\cdot))+1\big) = C\big(M(\rho_h^n)+1\big). \tag{3.3}$$

For $t\in[t_n,t_{n+1})$, recall $\rho_h^\dagger(t,\cdot)$ is equal to the solution of (2.19) over the interval $[0,h)$. Hence for any $t\in[t_n,t_{n+1})$, by (3.3),

$$M(\rho_h^\dagger(t,\cdot)) \leq C(M(\rho_h^n)+1).$$

Moreover, for all $t\in[t_n,t_{n+1})$ we have $\tilde\rho_h(t,\cdot)=\tilde\rho_h^{n+1}=\mu(h,\cdot)$, where again μ solves (2.19), and hence by (3.3)

$$M(\tilde\rho_h(t,\cdot)) = M(\tilde\rho_h^{n+1}) = M(\mu(h,\cdot)) \leq C(M(\rho_h^n)+1),$$

for any $t\in[t_n,t_{n+1})$. This completes the proof. □

3.1.2 The optimal transportation problem

In this section we discuss the well-posedness of the minimization problem (2.10). It is natural to achieve well-posedness of the scheme through finiteness, lower semi-continuity, and convexity of the functionals which appear in it. First observe that A_h is indeed positive.

Lemma 3.3 (The cost function)**.** The matrix A_h defined in (2.13) is positive definite (i.e., invertible) which implies,

$$\|x-y\|^2 \leq C c_h(x,y),\quad \forall x,y\in\mathbb{R}^d. \tag{3.4}$$

Proof. This is well-known [ADdR21, Appendix B.1] .

□

The next result addresses the existence of a unique minimiser to (2.10). This type of result is classical, for completeness the details of the proof can be found in Appendix A.1.

Proposition 3.4. Let $\mu\in\mathcal{P}_2^r(\mathbb{R}^d)$ with $\mathcal{F}(\mu)<\infty$. Then, there exists a unique $\nu^*\in\mathcal{P}_2^r(\mathbb{R}^d)$ such that

$$\nu^* = \operatorname*{argmin}_{\nu\in\mathcal{P}_2^r(\mathbb{R}^d)}\Big\{\frac{1}{2h}W_{c_h}(\mu,\nu)+\mathcal{F}(\nu)\Big\}. \tag{3.5}$$

3.2 Discrete Euler-Lagrange Equations

The following results are by now classical [JKO98, Proposition 4.1], so we state them without proof.

Lemma 3.5. Let $\eta\in C_c^\infty(\mathbb{R}^d,\mathbb{R}^d)$, and let Φ be the solution of the following ODE:

$$\partial_s\Phi_s = \eta(\Phi_s),\ \Phi_0 = \mathrm{id}. \tag{3.6}$$

Then for any $\nu\in\mathcal{P}_2^r(\mathbb{R}^d)$ we have

$$\delta\mathcal{F}(\nu,\eta) := \frac{d}{ds}\mathcal{F}\big((\Phi_s)_{\#}\nu\big)\Big|_{s=0} = \int_{\mathbb{R}^d}\nu(y)\eta(y)\cdot\nabla f(y)dy - \int_{\mathbb{R}^d}\nu(y)\mathrm{div}(\eta(y))dy. \tag{3.7}$$

Lemma 3.6. Let $\mu\in\mathcal{P}_2^r(\mathbb{R}^d)$. Let ν be the optimal solution in (3.5), and let γ be the corresponding optimal plan in $W_{c_h}(\mu,\nu)$. Then, for any $\eta\in C_c^\infty(\mathbb{R}^d,\mathbb{R}^d)$ we have

$$0 = \frac{1}{2h}\int_{\mathbb{R}^{2d}}\Big\langle\eta(y),\nabla_y c_h(x,y)\Big\rangle d\gamma(x,y) + \delta\mathcal{F}(\nu,\eta). \tag{3.8}$$

In particular, for any $\varphi\in C_c^\infty(\mathbb{R}^d)$, by choosing $\eta(x)=A_h\nabla\varphi(x)$, and $\tilde\gamma_h^{n+1,c}$ defined in (2.14), we have

$$0 = \frac{1}{h}\int_{\mathbb{R}^{2d}}\Big\langle y-x,\nabla\varphi(x)\Big\rangle d\tilde\gamma_h^{n+1,c}(x,y) + \delta\mathcal{F}(\rho_h^{n+1},A_h\nabla\varphi). \tag{3.9}$$

3.3 A priori estimates

In this section we establish a priori estimates which will allow us to prove the convergence of the scheme to a weak solution of (1.6) in Section 3.4. More precisely, we will show the uniformly boundedness of the second moments and of free energies of the minimization iterations (2.10). These uniform bounds are preserved under the conservative dynamics, this is explained in the next lemma.

Lemma 3.7. Let $n \in \{0, 1, \dots, N-1\}$. If there exists a constant $C_1 > 0$, independent of h and n, such that $M(\rho_h^n), \mathcal{F}(\rho_h^n) < C_1$, then $\tilde{\rho}_h^{n+1}$ obtained from (2.8) satisfies

$$M(\tilde{\rho}_h^{n+1}), \mathcal{F}(\tilde{\rho}_h^{n+1}) < C.$$

As usual, the constant C appearing is also independent of h and n, but will depend on C_1.

Proof. The bound for the moments clearly hold by Lemma 3.2. For the free energy functional, we have $\mathcal{F}(\tilde{\rho}_h^{n+1}) = F(\tilde{\rho}_h^{n+1}) + H(\rho_h^n)$ by the conservation of entropy in Lemma 3.1. Therefore, since f is Lipschitz

$$\begin{aligned}\mathcal{F}(\tilde{\rho}_h^{n+1}) = \int f(x)\tilde{\rho}_h^{n+1}(x)dx + H(\rho_h^n) &\leq C\int \big(\|x\| + 1\big)\tilde{\rho}_h^{n+1}(x)dx + H(\rho_h^n) \\ &\leq C(M(\tilde{\rho}_h^{n+1}) + 1) + H(\rho_h^n) \leq C(M(\tilde{\rho}_h^{n+1}) + 1) + \mathcal{F}(\rho_h^n) \leq C.\end{aligned}$$

□

The following lemma controls the sum of the optimal transport costs of the JKO steps, by using $\tilde{\rho}_h^{n+1}$ as a competitor to ρ_h^{n+1} in (2.10). This estimate is of a similar type to [JKO98][Equation (46)], however, because of the splitting nature of our scheme, we don't use ρ_h^n as a competitor in (2.10)- making the estimate more involved.

Lemma 3.8. For any $n \in \{1, \dots, N-1\}$ it holds that

$$\sum_{i=0}^{n-1} W_{c_h}(\tilde{\rho}_h^{i+1}, \rho_h^{i+1}) \leq Ch\Big(1 + \mathcal{F}(\rho^0) + \big(M(\rho_h^n) + 1\big)^\alpha\Big). \tag{3.10}$$

Proof. Let $n \in \{0, 1, \dots, N-1\}$. Since ρ_h^{n+1} attains the infimum in (2.10) we can compare it against $\tilde{\rho}_h^{n+1}$. This gives

$$\frac{1}{2h} W_{c_h}(\tilde{\rho}_h^{n+1}, \rho_h^{n+1}) \leq \mathcal{F}(\tilde{\rho}_h^{n+1}) - \mathcal{F}(\rho_h^{n+1}).$$

Using Lemma 3.1 for the entropy, the above is equivalent to

$$\frac{1}{2h} W_{c_h}(\tilde{\rho}_h^{n+1}, \rho_h^{n+1}) \leq F(\tilde{\rho}_h^{n+1}) - F(\rho_h^{n+1}) + H(\rho_h^n) - H(\rho_h^{n+1}). \tag{3.11}$$

Recall now that $(c_h, \tilde{\gamma}_h^{n+1,c}) = W_{c_h}(\tilde{\rho}_h^{n+1}, \rho_h^{n+1})$. Using that f is Lipschitz and Young's inequality with $\sqrt{\sigma}$ for some $\sigma > 0$, we can see

$$\begin{aligned}F(\tilde{\rho}_h^{n+1}) - F(\rho_h^{n+1}) = \int_{\mathbb{R}^{2d}} (f(x) - f(y))d\tilde{\gamma}_h^{n+1,c}(x,y) &\leq C\int \|x-y\| d\tilde{\gamma}_h^{n+1,c}(x,y) \\ &\leq \frac{C}{2\sigma}\int \|x-y\|^2 d\tilde{\gamma}_h^{n+1,c}(x,y) + \frac{C\sigma}{2} \\ &\leq \frac{C}{2\sigma}\int c_h(x,y) d\tilde{\gamma}_h^{n+1,c}(x,y) + \frac{C\sigma}{2},\end{aligned}$$

where in the last step we used (3.4). Substituting this into (3.11) yields

$$\frac{1}{2h} W_{c_h}(\tilde{\rho}_h^{n+1}, \rho_h^{n+1}) \leq \frac{C}{2\sigma}\int c_h(x,y) d\tilde{\gamma}_h^{n+1,c}(x,y) + \frac{C\sigma}{2} + H(\rho_h^n) - H(\rho_h^{n+1}).$$

Choosing $\sigma = 2Ch$ leads to

$$\frac{1}{2h} W_{c_h}(\tilde{\rho}_h^{n+1}, \rho_h^{n+1}) \leq \frac{1}{4h} W_{c_h}(\tilde{\rho}_h^{n+1}, \rho_h^{n+1}) + Ch + H(\rho_h^n) - H(\rho_h^{n+1}).$$

After rearranging we finally conclude

$$W_{c_h}(\tilde{\rho}_h^{n+1}, \rho_h^{n+1}) \leq C\Big(h^2 + h\big(H(\rho_h^n) - H(\rho_h^{n+1})\big)\Big). \tag{3.12}$$

The sum of (3.12) over $i \in \{0, \dots, n-1\}$ contains a telescopic component which allows for the simplified expression

$$\sum_{i=0}^{n-1} W_{c_h}(\tilde{\rho}_h^{i+1}, \rho_h^{i+1}) \leq Ch\Big(1 + H(\rho^0) - H(\rho_h^n)\Big),$$

where we have used that $Nh = T$. To deal with the terms in the right hand side of the above expression, we employ (A.3) to deal with $H(\rho_h^i)$, while for $H(\rho^0)$ we just use the positivity of f. This leads to

$$\sum_{i=0}^{n-1} W_{c_h}(\tilde{\rho}_h^{i+1}, \rho_h^{i+1}) \leq Ch\Big(1 + \mathcal{F}(\rho^0) + \big(M(\rho_h^n) + 1\big)^\alpha\Big).$$

□

The next Lemma provides uniform bounds in n and h for the second moments, free energy functionals, and positive part of the entropy functionals of the solutions from the scheme (2.8) - (2.10). The proof is inspired by the procedure found in [ADdR21, DPZ14, Hua00], first obtaining bounds locally and then extending them over the full time interval.

Lemma 3.9 (Boundedness of the energy functionals, second moments and the positive part of the entropy functionals)**.** For all $n \in \{0, 1, \dots, N\}$, we have

$$M(\rho_h^n), \mathcal{F}(\rho_h^n), H_+(\rho_h^n) \leq C \qquad \text{and} \qquad M(\tilde{\rho}_h^n), \mathcal{F}(\tilde{\rho}_h^n), H_+(\tilde{\rho}_h^n) \leq C.$$

Proof. Throughout this proof the constant $\bar{C}$ will change from line to line, and importantly it is independent of ρ^0.

For any $n \in \{1, \dots, N\}$ we have that

$$\begin{aligned} M(\rho^n) \leq 2\Big(M(\rho^0) + W_2^2(\rho^0, \rho^n)\Big) &\leq 2\Big(M(\rho^0) + n\sum_{i=0}^{n-1} W_2^2(\rho^i, \rho^{i+1})\Big) \\ &\leq 4\Big(M(\rho^0) + n\sum_{i=0}^{n-1} W_2^2(\rho^i, \tilde{\rho}^{i+1}) + W_2^2(\tilde{\rho}^{i+1}, \rho^{i+1})\Big) \\ &\leq 4\Big(M(\rho^0) + n\sum_{i=0}^{n-1} W_2^2(\rho^i, \tilde{\rho}^{i+1}) + \bar{C}W_{c_h}(\tilde{\rho}^{i+1}, \rho^{i+1})\Big). \end{aligned} \tag{3.13}$$

From Lemma 3.2 we have $4TW_2^2(\rho^i, \tilde{\rho}^{i+1}) \leq \bar{C}h^2(1 + M(\rho^i))$ for a constant $\bar{C}$ (independent of the initial condition), substituting this, and the bound (3.10) into (3.13), we have, whilst noting $hN = T$,

$$\begin{aligned} M(\rho^n) &\leq 4\Big(M(\rho^0) + \bar{C}\big(1 + \mathcal{F}(\rho^0)\big)\Big) + \bar{C}\Big((1 + M(\rho^n))^\alpha + h\sum_{i=0}^{n-1}(1 + M(\rho^i))\Big) \\ &\leq C + \bar{C}\Big((1 + M(\rho^n))^\alpha + h\sum_{i=0}^{n-1}(1 + M(\rho^i))\Big) \\ &\leq C + \bar{C}\Big((1 + M(\rho^n))^\alpha + h\sum_{i=0}^{n-1} M(\rho^i)\Big), \end{aligned} \tag{3.14}$$

for a constant C depending only on $\mathcal{F}(\rho^0)$ and $M(\rho^0)$, and constant $\bar{C}$ independent of ρ^0. Since the $\bar{C}$ appearing in (3.14) is fixed and independent of the initial condition, we can find $h_0 > 0, N_0 \in \mathbb{N}$ (independent of the initial condition) such that for all $h \leq h_0$ we have $hN_0\bar{C} \leq \frac{1}{2}$. Set $M_{N_0} := \max_{n=1,\dots,N_0} M(\rho^n)$. Then (3.14) implies

$$\begin{aligned} M_{N_0} &\leq C + \bar{C}\Big((1 + M_{N_0})^\alpha + hN_0M_{N_0}\Big) \\ &\leq C + \bar{C}(1 + M_{N_0})^\alpha + \frac{1}{2}M_{N_0}, \end{aligned}$$

which implies

$$M_{N_0} \leq 2\big(C + \bar{C}(1 + M_{N_0})^\alpha\big), \tag{3.15}$$

from which we can conclude $M(\rho^n) \leq C$, for all $n = 1, \dots, N_0$, and all $h \leq h_0$. For the free energy, note that by definition of ρ^{i+1}, we have that

$$\mathcal{F}(\rho^{i+1}) - F(\tilde{\rho}^{i+1}) - H(\tilde{\rho}^{i+1}) \leq 0,$$

adding and subtracting $F(\rho^i)$, and recalling that $H(\rho^i) = H(\tilde{\rho}^{i+1})$, implies

$$\mathcal{F}(\rho^{i+1}) - \mathcal{F}(\rho^i) \leq |F(\rho^i) - F(\tilde{\rho}^{i+1})|. \tag{3.16}$$

Summing (3.16) from $i = 0, \dots, n-1$, using that f is Lipschitz, and applying Young's inequality for some $\sigma > 0$, we have

$$\mathcal{F}(\rho^n) - \mathcal{F}(\rho^0) \leq \sum_{i=0}^{n-1} |F(\rho^i) - F(\tilde{\rho}^{i+1})| \leq C \sum_{i=0}^{n-1} \int_{\mathbb{R}^{2d}} \|x - y\| d\gamma^i(x, y) \leq C \sum_{i=0}^{n-1} \Big(\frac{1}{\sigma} W_2^2(\rho^i, \tilde{\rho}^{i+1}) + \sigma \Big). \tag{3.17}$$

Now let N_0, h_0 be chosen as before, and let $n = 1, \dots, N_0$. We know, by A.1 and the bounded moments just proved, that $W_2^2(\rho^i, \tilde{\rho}^{i+1}) \leq Ch^2(1 + M(\rho^i)) \leq Ch^2$ for $i \leq n$. Therefore, choosing $\sigma = h$ in (3.17) implies the uniform bounded energies $\mathcal{F}(\rho^n) \leq C$. Note that $\mathcal{F}(\rho^n) \leq C$ implies $H(\rho^n) \leq C$, moreover, (A.3) and the uniform bounds on $M(\rho_h^n)$ imply that $H_-(\rho^n) \leq C$, therefore we have that $H_+(\rho^n) \leq C$. So far we have established the uniform bounds

$$M(\rho^n), \mathcal{F}(\rho^n), H_+(\rho^n) \leq C, \quad \forall n = 1, \dots, N_0,\ h \leq h_0. \tag{3.18}$$

Since the N_0 and h_0 we have chosen are independent of the initial data we can extend the bound (3.18) to all $n \in \{1, \dots, N\}$ similarly as has been done in [Hua00, Lemma 5.3], see also [DPZ14]. The uniform bounds $M(\rho_h^n), \mathcal{F}(\rho_h^n), H_+(\rho_h^n) \leq C$, Lemma 3.7, and another application of (A.3) establishes $M(\tilde{\rho}_h^n), \mathcal{F}(\tilde{\rho}_h^n), H_+(\tilde{\rho}_h^n) \leq C$, completing the proof.

□

Lemma 3.9 states the uniform bounds for the discrete elements of our schemes. The following Lemma induces those bounds for the interpolations (2.16), (2.17) and (2.18).

Lemma 3.10 (A priori estimates for the interpolations)**.** For all $n \in \{0, 1, \dots, N\}$, the moments, free-energies and the positive part of the entropies are uniformly bounded (in n, h, t), namely,

$$M(\rho_h(t,\cdot)), M(\tilde{\rho}_h(t,\cdot)), M(\rho_h^\dagger(t,\cdot)), \mathcal{F}(\rho_h(t,\cdot)), \mathcal{F}(\tilde{\rho}_h(t,\cdot)), \mathcal{F}(\rho_h^\dagger(t,\cdot)), H_+(\rho_h(t,\cdot)), H_+(\tilde{\rho}_h(t,\cdot)), H_+(\rho_h^\dagger(t,\cdot)) \leq C.$$

Proof. These results for the interpolations follow easily from Lemma 3.9. Indeed, it is immediate from their definitions how this is inferred for the interpolations $\rho_h(t,\cdot), \tilde{\rho}(t,\cdot)$. For $\rho_h^\dagger(t,\cdot)$, just notice from Lemma 3.2 that we have $M(\rho_h^\dagger(t,\cdot)) \leq C$. This uniform moment bound gives us the other two bounds for $\rho_h^\dagger(t,\cdot)$: for the free energy one follows the argument in Lemma 3.7 (using the bounded energy of ρ_h^n), and for the positive entropy one uses again (A.3). □

The uniform bounds established in Lemma 3.9 allow us to control the transport cost (w.r.t to both cost functions c_h and $\|\cdot\|^2$) of the JKO step.

Lemma 3.11 (Estimates of the sum of optimal transport costs)**.** We have

$$\sum_{n=0}^{N-1} W_{c_h}(\tilde{\rho}_h^{n+1}, \rho_h^{n+1}) \leq Ch, \quad \text{and} \quad \sum_{n=0}^{N-1} W_2^2(\tilde{\rho}_h^{n+1}, \rho_h^{n+1}) \leq Ch. \tag{3.19}$$

Proof. The estimate (3.10), together with the uniform bounds of Lemma 3.9, gives the first result of (3.19). The second result is immediate from the first and (3.4), since

$$\sum_{n=0}^{N-1} W_2^2(\tilde{\rho}_h^{n+1}, \rho_h^{n+1}) \leq C \sum_{n=0}^{N-1} W_{c_h}(\tilde{\rho}_h^{n+1}, \rho_h^{n+1}) \leq Ch.$$

□

The uniform moment bounds (in conjunction with the preliminary observation of Lemma 3.2) allow us to control the Wasserstein cost of the conservative phase.

Lemma 3.12 (Estimates of the sum of optimal transport costs for the conservative dynamics)**.** We have

$$\sum_{n=0}^{N-1} W_2^2(\rho_h^n, \tilde{\rho}_h^{n+1}) \leq Ch. \tag{3.20}$$

Proof. We recall (3.2), implying that for any $n = 0, \dots, N-1$, $W_2^2(\rho_h^n, \tilde{\rho}_h^{n+1}) \leq Ch^2(1 + M(\rho_h^n))$, the uniform bounded moment estimates then give the result. □

3.4 Convergence of the operator-splitting scheme

Having obtained a priori estimates, in this section we prove the main theorem, Theorem 2.4, that is the convergence of the time-interpolations of the discrete solutions constructed from the operator-splitting scheme in Section 2.2 to a weak solution of the main evolutionary equation (1.6). The following Lemma shows that these interpolations converge to limits which are equal almost everywhere to some curve $[0,T] \ni t \mapsto \rho(t,\cdot) \in \mathcal{P}_2^r(\mathbb{R}^d)$, and moreover, the sequences $\rho_h(t,\cdot), \tilde{\rho}_h(t,\cdot), \rho_h^\dagger(t,\cdot)$ converge in W_2 to ρ, uniformly in time.

Lemma 3.13. [Convergence of the time-interpolations] There exists a curve $[0,T] \ni t \mapsto \rho(t,\cdot) \in \mathcal{P}_2^r(\mathbb{R}^d)$, such that

$$\lim_{h\to 0} \sup_{t\in[0,T]} \max\Big\{W_2(\rho_h(t,\cdot),\rho(t,\cdot)), W_2(\tilde{\rho}_h(t,\cdot),\rho(t,\cdot)), W_2(\rho_h^\dagger(t,\cdot),\rho(t,\cdot))\Big\} = 0. \tag{3.21}$$

Proof. The proof follows an adapted version of [AGS08, Theorem 11.1.6], since we will require uniform in time convergence we utilise a refined version of the Arzela-Ascoli theorem [AGS08, Proposition 3.3.1] in a similar manner to [CL17][Section 5.1]. We provide the argument for ρ_h only, the approach for $\tilde{\rho}_h, \rho_h^\dagger$ is similar. To obtain the uniform convergence we set up an Arzela-Ascoli argument. Since the paths ρ_h are not continuous we introduce the continuous concatenation of $\{\rho_h^n\}_n$ by geodesics. Let $n \in \{1,\dots,N-1\}$. Fix any $s,t \in [0,T]$, define the path $\nu_h : [0,T] \to \mathcal{P}_2(\mathbb{R}^d)$ by concatenating ρ_h^n and ρ_h^{n+1} on $[t_{n-1},t_n]$ by a constant speed geodesic. Then for $t \in [t_{n-1},t_n)$

$$\begin{aligned} W_2(\rho_h(t),\nu_h(t)) = W_2(\rho_h^n,\nu_h(t)) =& W_2(\nu_h(t_{n-1}),\nu_h(t)) \\ \leq& W_2(\rho_h^n,\rho_h^{n+1})(t-t_{n-1}) \leq W_2(\rho_h^n,\rho_h^{n+1})h \leq Ch, \end{aligned}$$

by the bounded moments of Lemma 3.10. Let $t<s$, $t \in [ih,(i+1)h)$, $s \in [jh,(j+1)h)$, for some $i,j \in \{0,\dots,N-1\}$. We then have

$$\begin{aligned} W_2(\nu_h(t),\nu_h(s)) \leq& W_2(\nu_h(t),\rho_h^{i+1}) + W_2(\rho_h^{i+1},\rho_h^{j+1}) + W_2(\rho_h^{j+1},\nu_h(s)) \\ \leq& hW_2(\rho_h^{i+2},\rho_h^{i+1}) + W_2(\rho_h^{i+1},\rho_h^{j+1}) + hW_2(\rho_h^{j+1},\rho_h^{j+2}) \\ \leq& Ch + W_2(\rho_h^{i+1},\rho_h^{j+1}). \end{aligned}$$

Using the triangle property and then the Cauchy Schwartz inequality we have

$$\begin{aligned} W_2(\rho_h^{i+1},\rho_h^{j+1}) \leq \sum_{n=i+1}^{j} W_2(\rho_h^n,\rho_h^{n+1}) \leq \sqrt{(j-i)h}\sqrt{\frac{1}{h}\sum_{n=i+1}^{j} W_2^2(\rho_h^n,\rho_h^{n+1})} \leq \sqrt{(j-i)h}\sqrt{\frac{1}{h}\sum_{n=0}^{N-1} W_2^2(\rho_h^n,\rho_h^{n+1})} \\ \leq \sqrt{(j-i)h}\sqrt{\frac{2}{h}\sum_{n=0}^{N-1} W_2^2(\rho_h^n,\tilde{\rho}_h^{n+1}) + W_2^2(\tilde{\rho}^{n+1},\rho^{n+1})} \leq C\sqrt{(j-i)h} \leq C\sqrt{s-t}. \end{aligned}$$

Therefore the family ν_h is uniformly equicontinuous, [AGS08, Proposition 3.3.1]. This implies that for all $t \in [0,T]$ $\{\nu_h(t,\cdot)\}_{h>0} \subset (\mathcal{P}_2(\mathbb{R}^d), W_2)$ is (point-wise) precompact, and hence we can use Arzela-Ascoli [BS18, Theorem 1.1.11] to obtain uniform convergence (taking subsequences if necessary) in W_2 (over $[0,T]$ as $h \to 0$) of the paths ν_h to a limit, which we call ρ. We are then able to deduce the uniform convergence of ρ_h to ρ from that of ν_h, namely

$$\begin{aligned} \lim_{h\to 0} \sup_{t\in[0,T]} W_2(\rho_h(t),\rho(t)) \leq& \lim_{h\to 0} \sup_{t\in[0,T]} \Big(W_2\big(\rho_h(t),\nu_h(t)\big) + W_2\big(\nu_h(t),\rho(t)\big)\Big) \\ \leq& \lim_{h\to 0} \Big(Ch + \sup_{t\in[0,T]} W_2\big(\nu_h(t),\rho(t)\big)\Big) = 0. \end{aligned}$$

Since W_2 induces a stronger topology than that of weak convergence we have for all $t \in [0,T]$ $\rho_h(t,\cdot) \rightharpoonup \rho(t,\cdot) \in \mathcal{P}_2(\mathbb{R}^d)$. Moreover, by the uniform moment and entropy bounds (see Lemma 3.10) we have by (A.4) that the limit $\rho(t,\cdot) \in \mathcal{P}_2^r(\mathbb{R}^d)$. By an almost identical procedure (this time concatenating geodesics between $\{\tilde{\rho}_h^n\}_n$, and using (3.2) for $\rho_h^\dagger$) we get the same convergence of $\tilde{\rho}_h, \rho_h^\dagger$ to some limit $\tilde{\rho} \in \mathcal{P}_2^r(\mathbb{R}^d)$. It remains only to show that $\rho = \tilde{\rho}$ a.e., note we have, for instance using the Dominated Convergence theorem, letting $\varphi \in C_c^\infty([0,T),\mathbb{R}^d)$

$$\begin{aligned} \int_0^T \int_{\mathbb{R}^d} \big(\tilde{\rho}(t,x) - \rho(t,x)\big)\varphi(t,x)dxdt =& \lim_{h\to 0} \int_0^T \int_{\mathbb{R}^d} \big(\tilde{\rho}_h(t,x) - \rho_h(t,x)\big)\varphi(t,x)dxdt \\ =& \lim_{h\to 0} \sum_{n=0}^{N-1} \int_{t_n}^{t_{n+1}} \int_{\mathbb{R}^d} \big(\tilde{\rho}_h^{n+1}(x) - \rho_h^{n+1}(x)\big)\varphi(t,x)dxdt \\ =& \lim_{h\to 0} \sum_{n=0}^{N-1} \int_{t_n}^{t_{n+1}} \int_{\mathbb{R}^{2d}} \big(\varphi(t,x) - \varphi(t,y)\big)\tilde{\gamma}^{n+1}(dx,dy)dt, \end{aligned}$$

where we recall $\tilde{\gamma}^{n+1}$ is the optimal coupling between ρ^{n+1} and $\tilde{\rho}^{n+1}$ in W_2. By Taylor's theorem, Jensen inequality and then Cauchy Schwartz, we have

$$\begin{aligned}\int_0^T \int_{\mathbb{R}^d} \big(\tilde{\rho}(t,x) - \rho(t,x)\big)\varphi(t,x)dxdt &\leq \lim_{h\to 0} h \sup \|\nabla\varphi\| \sum_{n=0}^{N-1} \int_{\mathbb{R}^{2d}} \|x-y\| \gamma^{n+1}(dx,dy) \\ &\leq \lim_{h\to 0} h \sup \|\nabla\varphi\| \sum_{n=0}^{N-1} W_2(\tilde{\rho}_h^{n+1}, \rho_h^{n+1}) \\ &\leq \lim_{h\to 0} h\sqrt{N} \sup \|\nabla\varphi\| \sqrt{\sum_{n=0}^{N-1} W_2^2(\tilde{\rho}_h^{n+1}, \rho_h^{n+1})} \\ &\leq \lim_{h\to 0} Ch\sqrt{T} \sup \|\nabla\varphi\| = 0,\end{aligned}$$

where in the last line we used Lemma 3.11. We are then able to conclude that $\tilde{\rho}$ and ρ are equal a.e.

□

We can also ascertain the L^1 convergence (2.24), i.e. fix $t \in [0,T]$, we show that we have weak $L^1(\mathbb{R}^d)$ convergence of $\rho_h(t,\cdot), \tilde{\rho}_h(t,\cdot)$, and $\rho_h^\dagger(t,\cdot)$ to $\rho(t,\cdot)$ (the same limit as found in the previous Lemma 3.13), that is convergence against $L^\infty(\mathbb{R}^d)$ functions not just those in $C_b(\mathbb{R}^d)$. Indeed, since $x \mapsto \max\{x\log x, 0\}$ is a superlinear function, the uniform bounds on the positive entropy (Lemma 3.10) implies the families $\{\rho_h(t,\cdot)\}_h, \{\tilde{\rho}_h(t,\cdot)\}_h, \{\rho_h^\dagger(t,\cdot)\}_h$ are equi-integrable, and hence, by the weak convergence of the previous lemma, [San15, Box 8.2 (p301)] implies the weak $L^1(\mathbb{R}^d)$ convergence. Recall this implies weak $L^1((0,T)\times\mathbb{R}^d)$ convergence.

The following lemma is a key step in our analysis linking the conservative and the disspative phases together.

Lemma 3.14. For any $\varphi \in C_c^\infty([0,T)\times\mathbb{R}^d)$ we have that

$$\begin{aligned}\sum_{n=0}^{N-1} \int_{\mathbb{R}^d} \Big(\tilde{\rho}_h^{n+1}(x) - \rho_h^{n+1}(x)\Big)\varphi(t_{n+1},x)dx = \int_0^T \int_{\mathbb{R}^d} \rho_h^\dagger(t,x)\big(\partial_t\varphi(t,x) + b[\rho_h(t-h)](x)\cdot\nabla\varphi(t,x)\big)dxdt \\ + \int_{\mathbb{R}^d} \rho^0(x)\varphi(0,x)dx. \qquad (3.22)\end{aligned}$$

Proof. Let $n \in \{0,\dots N-1\}$. First notice that for $t \in [t_n, t_{n+1}]$ by (2.9) and the chain rule, we have

$$\partial_t\big(\varphi(t, X_h^n(t-t_n,x))\big) = \Big(\partial_t\varphi + b[\rho_n^h]\cdot\nabla\varphi\Big)(t, X_h^n(t-t_n,x)). \qquad (3.23)$$

Now consider

$$\begin{aligned}&\sum_{n=0}^{N-1} \int_{\mathbb{R}^d} \Big(\tilde{\rho}_h^{n+1}(x) - \rho_h^{n+1}(x)\Big)\varphi(t_{n+1},x)dx - \int_{\mathbb{R}^d} \rho^0(x)\varphi(0,x)dx \\ &= \sum_{n=0}^{N-1} \int_{\mathbb{R}^d} \Big(\tilde{\rho}_h^{n+1}(x)\varphi(t_{n+1},x) - \rho_h^n(x)\varphi(t_n,x)\Big)dx \qquad (3.24) \\ &= \sum_{n=0}^{N-1} \int_{\mathbb{R}^d} \rho_h^n(x)\Big(\varphi(t_{n+1}, X_h^n(h,x)) - \varphi(t_n,x)\Big)dx \\ &= \sum_{n=0}^{N-1} \int_{t_n}^{t_{n+1}} \int_{\mathbb{R}^d} \rho_h^n(x)\partial_t\big(\varphi(t, X_h^n(t-t_n,x))\big)dxdt \\ &= \sum_{n=0}^{N-1} \int_{t_n}^{t_{n+1}} \int_{\mathbb{R}^d} \rho_h^n(x)\Big(\partial_t\varphi + b[\rho_h^n]\cdot\nabla\varphi\Big)(t, X_h^n(t-t_n,x))dxdt \qquad (3.25) \\ &= \int_0^T \int_{\mathbb{R}^d} \rho_h^\dagger(t,x)\Big(\partial_t\varphi + b[\rho_h(t-h,\cdot)]\cdot\nabla\varphi\Big)(t,x)dxdt, \qquad (3.26)\end{aligned}$$

where in (3.24) follows since φ has compact support, in (3.25) we have applied (3.23), and in (3.26) we have used the definitions of the interpolations $\rho_h, \rho_h^\dagger$. □

Now following the classical procedure we can interpolate across the discrete Euler-Lagrange equations (3.9).

Lemma 3.15. For any $\varphi \in C_c^\infty([0,T)\times\mathbb{R}^d)$ we have

$$\int_0^T \int_{\mathbb{R}^d} \rho_h^\dagger(t,x)\big(\partial_t\varphi(t,x) + b[\rho(t-h,\cdot)](x)\cdot\nabla\varphi(t,x)\big)dxdt$$
$$= -\int_{\mathbb{R}^d}\rho^0(x)\varphi(0,x)dx - h\sum_{n=0}^{N-1}\delta\mathcal{F}(\rho_h^{n+1}, A_h\nabla\varphi(t_{n+1},\cdot)) + O(h). \tag{3.27}$$

Proof. Let $n\in\{0,\dots,N-1\}$. By Taylor's Theorem we have

$$\int_{\mathbb{R}^d}\Big(\rho_h^{n+1}(x)-\tilde\rho_h^{n+1}(x)\Big)\varphi(t_{n+1},x)dx = \int_{\mathbb{R}^{2d}}\Big(\varphi(t_{n+1},y)-\varphi(t_{n+1},x)\Big)d\tilde\gamma_h^{n+1,c}(x,y)$$
$$= \int_{\mathbb{R}^{2d}}\Big\langle y-x,\nabla\varphi(t_{n+1},y)\Big\rangle d\tilde\gamma_h^{n+1,c}(x,y)+\kappa_n(t_{n+1}). \tag{3.28}$$

By Lemma 3.3 we can bound the remainder term κ_n, namely,

$$|\kappa_n(t)| \le C\sup_{t\in[0,T),x\in\mathbb{R}^d}\|\nabla^2\varphi\|\int_{\mathbb{R}^{2d}}\|x-y\|^2 d\tilde\gamma_h^{n+1,c}(x,y) \le C\sup_{t\in[0,T),x\in\mathbb{R}^d}\|\nabla^2\varphi\|\int_{\mathbb{R}^{2d}}c_h(x,y)d\tilde\gamma_h^{n+1,c}(x,y). \tag{3.29}$$

Using (3.28) in combination with the Euler-Lagrange equation (3.9) yields the identity

$$\int_{\mathbb{R}^d}\Big(\rho_h^{n+1}(x)-\tilde\rho_h^{n+1}(x)\Big)\varphi(t_{n+1},x)dx = \kappa_n(t_{n+1}) - h\delta\mathcal{F}(\rho_h^{n+1}, A_h\nabla\varphi(t_{n+1},\cdot)).$$

Summing the previous expression over $n=0,\dots,N-1$ gives

$$\sum_{n=0}^{N-1}\int_{\mathbb{R}^d}\Big(\rho_h^{n+1}(x)-\tilde\rho_h^{n+1}(x)\Big)\varphi(t_{n+1},x)dx = O(h) - h\sum_{n=0}^{N-1}\delta\mathcal{F}(\rho_h^{n+1}, A_h\nabla\varphi(t_{n+1},\cdot)), \tag{3.30}$$

where we have combined (3.29) with Lemma 3.11 to conclude $|\sum_{n=0}^{N-1}\kappa_n(t_{n+1})|\le Ch$. Finally, using (3.22) on the left hand side of (3.30), multiplying through by -1, delivers the sought result. □

We are now ready to prove the main theorem, Theorem 2.4.

Proof of Theorem 2.4. Recall the convergence result of Lemma 3.13. To prove Theorem 2.4 we need only to argue that the limit $h\to 0$, $N\to\infty$ in (3.27) can be taken. Clearly the error term $O(h)$ in (3.27) goes to zero (as $h\to 0$), and for any $\varphi\in C_c^\infty([0,T),\mathbb{R}^d)$ we have

$$\lim_{h\to 0}\int_0^T\int_{\mathbb{R}^d}\rho_h^\dagger(t,x)\partial_t\varphi(t,x)dxdt = \int_0^T\int_{\mathbb{R}^d}\rho(t,x)\partial_t\varphi(t,x)dxdt.$$

We now address the remaining terms of (3.27): the free energy and the divergence free part. We start with the free energy term $\delta\mathcal{F}$. Note that we can write

$$h\sum_{n=0}^{N-1}\delta\mathcal{F}\big(\rho_h^{n+1}, A_h\nabla\varphi(t_{n+1},\cdot)\big)$$
$$= \sum_{n=0}^{N-1}\int_{t_n}^{t_{n+1}}\Big(\int_{\mathbb{R}^d}\rho_h^{n+1}(x)\big(A_h\nabla\varphi(t_{n+1},x)\cdot\nabla f(x)\big)dx - \int_{\mathbb{R}^d}\rho_h^{n+1}(x)\mathrm{div}\big(A_h\nabla\varphi(t_{n+1},x)\big)dx\Big)dt$$
$$= \sum_{n=0}^{N-1}\int_{t_n}^{t_{n+1}}\Big(\int_{\mathbb{R}^d}\rho_h(t,x)\big(A_h\nabla\varphi(t_{n+1},x)\cdot\nabla f(x)\big)dx - \int_{\mathbb{R}^d}\rho_h(t,x)\mathrm{div}\big(A_h\nabla\varphi(t_{n+1},x)\big)dx\Big)dt. \tag{3.31}$$

Consider the first term on the right hand side of (3.31) (the second term can be dealt with in a similar manner). Adding and subtracting $A_h\nabla\varphi(t,x)$ and $A\nabla\varphi(t,x)$, we get

$$\sum_{n=0}^{N-1}\int_{t_n}^{t_{n+1}}\int_{\mathbb{R}^d}\rho_h^{n+1}\Big(A_h\nabla\varphi(t_{n+1},x)\cdot\nabla f(x)\Big)dxdt$$
$$= \sum_{n=0}^{N-1}\int_{t_n}^{t_{n+1}}\int_{\mathbb{R}^d}\Big(\rho_h\Big(A\nabla\varphi\cdot\nabla f\Big)(t,x) + \rho_h\Big(\big(A_h-A\big)\nabla\varphi\cdot\nabla f\Big)(t,x)$$
$$+\rho_h(t)\Big(A_h\big(\nabla\varphi(t_{n+1})-\nabla\varphi(t)\big)\cdot\nabla f\Big)(x)\Big)dxdt. \tag{3.32}$$

Then, as $h \to 0$, the first term tends to $\int_0^T \int_{\mathbb{R}^d} \rho\big(A\nabla\varphi \cdot \nabla f\big)(t,x)dxdt$ by the weak $L^1([0,T]\times\mathbb{R}^d)$ convergence, the second term tends to zero by Cauchy Schwartz and the fact that $\lim_{h\to 0}\|A_h - A\| = 0$ and again the weak $L^1([0,T]\times\mathbb{R}^d)$ convergence. The third term in (3.32) also tends to zero, since

$$\lim_{h\to 0}\sum_{n=0}^{N-1}\int_{t_n}^{t_{n+1}}\int_{\mathbb{R}^d}\rho_h(t)\Big(A_h\big(\nabla\varphi(t_{n+1}) - \nabla\varphi(t)\big)\cdot\nabla f\Big)(x)dxdt$$

$$\leq \lim_{h\to 0} C\|A_h\| \sup_{x\in\mathbb{R}^d}\|\nabla f(x)\| \sup_{[u_h,r_h]\subset[0,T),|u_h-r_h|\leq h}\ \sup_{s\in[u_h,r_h],x\in\mathbb{R}^d}\|\nabla\varphi(r_h,x) - \nabla\varphi(s,x)\|\int_0^T\int_{\mathbb{R}^d}\rho_h(t,x)dxdt = 0,$$

where we have used that $\|A_h\|, \sup\|\nabla f\| \leq C$, that ρ_h is a probability density, and that $\nabla\varphi$ is uniformly continuous.

Lastly, we address the divergence free term in (3.27). Adding and subtracting $\rho_h^\dagger b[\rho_h^\dagger]\cdot\nabla\varphi$ gives

$$\int_0^T\int_{\mathbb{R}^d}\rho_h^\dagger(t,x)b[\rho_h(t-h,\cdot)](x)\cdot\nabla\varphi(t,x)dxdt$$
$$= \int_0^T\int_{\mathbb{R}^d}\rho_h^\dagger(t,x)\Big(b[\rho_h(t-h,\cdot)](x) - b[\rho_h^\dagger(t,\cdot)]\Big)\cdot\nabla\varphi(t,x)dxdt + \int_0^T\int_{\mathbb{R}^d}\rho_h^\dagger(t,x)b[\rho_h^\dagger(t,\cdot)]\cdot\nabla\varphi(t,x)dxdt. \quad (3.33)$$

The first term in (3.33) converges to zero as $h\to 0$ since

$$\Big|\int_0^T\int_{\mathbb{R}^d}\rho_h^\dagger(t,x)\Big(b[\rho_h(t-h,\cdot)](x) - b[\rho_h^\dagger(t,\cdot)](x)\Big)\cdot\nabla\varphi(t,x)dxdt\Big|$$
$$\leq C\int_0^T\Big(\int_{\mathbb{R}^d}\rho_h^\dagger(t,x)\|b[\rho_h(t-h,\cdot)](x) - b[\rho_h^\dagger(t,\cdot)](x)\|^2dx\Big)^{\frac{1}{2}}dt \quad (3.34)$$
$$\leq C\int_0^T W_2(\rho_h(t-h,\cdot),\rho_h^\dagger(t,\cdot))dt \quad (3.35)$$
$$= C\sum_{n=0}^{N-1}\int_{t_n}^{t_{n+1}} W_2(\rho_h^n,\rho_h^\dagger(t,\cdot))dt$$
$$\leq CTh, \quad (3.36)$$

where in (3.34) we have used the Cauchy Schwartz and Jensen's inequality, in (3.35) we have used Assumption (2.22), and in (3.36) we have used (3.2) and the bounded moments result of Lemma 3.10. The second term on the right hand side of (3.33) has already the desired convergence, indeed consider

$$\Big|\int_0^T\int_{\mathbb{R}^d}\Big(\rho_h^\dagger(t,x)b[\rho_h^\dagger(t,\cdot)](x) - \rho(t,x)b[\rho(t,\cdot)](x)\Big)\cdot\nabla\varphi(t,x)dxdt\Big|$$
$$\leq C\int_0^T\Big(\int_{\mathbb{R}^d}\rho_h^\dagger(t,x)\|b[\rho_h^\dagger(t,\cdot)](x) - b[\rho(t,\cdot)](x)\|^2dx\Big)^{\frac{1}{2}}dt \quad (3.37)$$
$$+\Big|\int_0^T\int_{\mathbb{R}^d}\big(\rho(t,x) - \rho_h^\dagger(t,x)\big)b[\rho(t,\cdot)](x)\cdot\nabla\varphi(t,x)dxdt\Big|$$
$$\leq CT\sup_{t\in[0,T]}W_2(\rho_h^\dagger(t,\cdot),\rho(t,\cdot)) + \Big|\int_0^T\int_{\mathbb{R}^d}\big(\rho(t,x) - \rho_h^\dagger(t,x)\big)b[\rho(t,\cdot)](x)\cdot\nabla\varphi(t,x)dxdt\Big|, \quad (3.38)$$

where in (3.37) we have added and subtracted $\rho_h^\dagger b[\rho]$, used Cauchy Schwartz and Jensen's inequality, and in (3.38) we used again Assumption (2.22). The two terms in (3.38) go to zero from the convergence of $\rho_h^\dagger$ in Lemma 3.13 and that $b[\rho(t,\cdot)]\cdot\nabla\varphi \in L^\infty((0,T)\times\mathbb{R}^d)$.

Having the above estimates, by passing to the limit $h\to 0$ in (3.27) we obtain precisely the weak formulation (2.3) of the evolutionary equation (1.6). This completes the proof of Theorem 2.4. □

4 Entropy Regularised Scheme

From a computational point of view, implementing a JKO-type scheme (1.3) directly is expensive since at each iteration it requires the resolution of an optimal transportation problem. The entropic regularisation technique developed in [Cut13] overcomes this difficulty by transforming the transport problem into a strictly convex problem that can be solved more efficiently with matrix scaling algorithms such as the Sinkhorn's algorithm [KS67]. In recent years, this regularisation technique has found applications in a variety of domains such as machine learning, image processing,

graphics and biology. In particular, several works have developed entropic regularisation schemes for solving evolutionary equations, such as nonlinear diffusion equations [CDPS17, Pey15], flux-limited gradient flows [MS20b] and a tumour growth model of Hele-Shaw type [DMC20]. We refer the reader to the recent monograph [PC19] for a great detailed account of the entropic regularisation technique.

In this section, we provide an entropy regularised version of the scheme introduced in Section 1. The regularised scheme, presented below, differs only in that we have penalised the weighted Wasserstein distance by an entropy term. The convergence of this new scheme is stated in Theorem 4.2, the proof of which is sketched since it only differs slightly to that of Theorem 2.4. The results and techniques of this section are similar to those appearing in [CDPS17, ADdR21]. The following assumption introduces a theoretical constraint on the scaling of the time step and strength of entropic regularisation. It ensures that the error made by the regularisation goes to zero sufficiently fast.

Assumption 4.1 (The regularisation's scaling parameters)**.** Take three sequences $\{N_k\}_{k\in\mathbb{N}} \subset \mathbb{N}$, $\{\varepsilon_k\}_{k\in\mathbb{N}} \subset \mathbb{R}_+$, and $\{h_k\}_{k\in\mathbb{N}} \subset \mathbb{R}_+$, which, for any $k \in \mathbb{N}$, abide by the following scaling

$$h_k N_k = T, \quad \text{and} \quad 0 < \varepsilon_k \leq \varepsilon_k |\log \varepsilon_k| \leq C h_k^2, \tag{4.1}$$

and are such that $h_k, \varepsilon_k \to 0$ and $N_k \to \infty$ as $k \to \infty$.

An entropic regularisation of the operator-splitting scheme: Let the sequences $\{h_k\}_{k\in\mathbb{N}}, \{\varepsilon_k\}_{k\in\mathbb{N}}, \{N_k\}_{k\in\mathbb{N}}$, satisfy Assumption 4.1. Throughout the section, for the sake of notational clarity, we have mostly suppressed the dependence of ε, h and N on k. Let $\mathcal{F}(\rho_0) < \infty$, and set $\rho_k^0 = \tilde{\rho}_k^0 = \rho^0$. Let $n \in \{0, \dots, N_k - 1\}$. Given ρ_k^n we find ρ_k^{n+1} as follows, first introduce the push forward of ρ_k^n by the Hamiltonian flow as

$$\tilde{\rho}_k^{n+1} = X_k^n(h, \cdot)_\# \rho_k^n, \tag{4.2}$$

where X_k^n solves

$$\begin{cases} \partial_t X_k^n = b[\rho_k^n] \circ X_k^n, \\ X_k^n(0, \cdot) = \mathrm{id}. \end{cases} \tag{4.3}$$

Next define ρ_k^{n+1} as the minimiser of the regularised JKO type descent step

$$\rho_k^{n+1} = \mathrm{argmin}_{\rho \in \mathcal{P}_2^r(\mathbb{R}^d)} \big\{ \frac{1}{2h} W_{c_h,\varepsilon}(\tilde{\rho}_k^{n+1}, \rho) + \mathcal{F}(\rho) \big\}, \tag{4.4}$$

where $W_{c_h,\varepsilon}$ is the regularised weighted Wasserstein

$$W_{c_h,\varepsilon}(\mu, \nu) := \inf_{\gamma \in \Pi(\mu,\nu)} \Big\{ \int_{\mathbb{R}^{2d}} c_h(x, y) d\gamma(x, y) + \varepsilon H(\gamma) \Big\}, \tag{4.5}$$

for the same cost function defined in (2.12). Let $\tilde{\gamma}_k^{n,c}$ be the optimal plan associated to $W_{c_h,\varepsilon}(\tilde{\rho}_k^n, \rho_k^n)$, and define the interpolations $\rho_k, \tilde{\rho}_k, \rho_k^\dagger$ analogously to the un-regularised case but now with respect to the new sequences $\{\rho_k^n\}_{n=0}^N$ and $\{\tilde{\rho}_k^n\}_{n=0}^N$.

The convergence of the above entropic regularised scheme is established in the next result.

Theorem 4.2. Assume that f, b and A satisfy Assumption 2.1, and let the sequences $\{h_k\}_{k\in\mathbb{N}}, \{\varepsilon_k\}_{k\in\mathbb{N}}, \{N_k\}_{k\in\mathbb{N}}$ satisfy Assumption 4.1. Let $\rho_0 \in \mathcal{P}_2^r(\mathbb{R}^d)$ satisfy $\mathcal{F}(\rho_0) < \infty$. Let $\{\rho_k^n\}_{n=0}^{N_k}, \{\tilde{\rho}_k^n\}_{n=0}^{N_k}$ to be the solution of the regularised scheme (4.2)-(4.4), with interpolations $\rho_k, \tilde{\rho}_k, \rho_k^\dagger$ as defined above.

Then

(i)
$$\rho_k, \tilde{\rho}_k, \rho_k^\dagger \underset{k\to\infty}{\longrightarrow} \rho \quad \text{in} \quad L^1((0, T) \times \mathbb{R}^d). \tag{4.6}$$

(ii) Moreover, there exists a map $[0, T] \ni t \mapsto \rho(t, \cdot)$ in $\mathcal{P}_2^r(\mathbb{R}^d)$ such that

$$\sup_{t\in[0,T]} \max \Big\{ W_2\big(\rho_k(t, \cdot), \rho(t, \cdot)\big),\ W_2\big(\tilde{\rho}_k(t, \cdot), \rho(t, \cdot)\big),\ W_2\big(\rho_k^\dagger(t, \cdot), \rho(t, \cdot)\big) \Big\} \underset{k\to\infty}{\longrightarrow} 0, \tag{4.7}$$

where the limits ρ appearing above are weak solutions of the evolution equation (1.6) in the sense of Definition 2.3.

The proof does not change much from that of Theorem 2.4, so we provide only a sketch, highlighting the parts that are different.

Proof of Theorem 4.2. Let $n \in \{0, \dots, N-1\}$.

The well-posedness. The well-posedness of the regularised scheme relies on the well-posedness of the minimisation problem (4.5), the proof of which can be found in [ADdR21, Section 3].

A priori estimates. In the proof of Theorem 2.4 we compare the quantity $\frac{1}{2h}W_{c_h}(\tilde{\rho}_h^{n+1}, \rho_h^{n+1}) + \mathcal{F}(\rho_h^{n+1})$ against $\frac{1}{2h}W_{c_h}(\tilde{\rho}_h^{n+1}, \tilde{\rho}_h^{n+1}) + \mathcal{F}(\tilde{\rho}_h^{n+1})$. The term $W_{c_h}(\tilde{\rho}_h^{n+1}, \tilde{\rho}_h^{n+1})$ is zero, and hence we end up with a control of $W_{c_h}(\tilde{\rho}_h^{n+1}, \rho_h^{n+1})$ in terms of the free energy. However, since $W_{c_h,\varepsilon}(\tilde{\rho}_k^{n+1}, \tilde{\rho}_k^{n+1}) \neq 0$, we need to select a new distribution to compare the performance of ρ_k^{n+1} against. We judiciously choose a distribution ρ_ϵ (with optimal plan γ_ϵ) as to make the cost of transporting mass zero, i.e as $\epsilon \to 0$ we aim to have $(c_h, \gamma_\epsilon) \to 0$. We construct such a candidate distribution ρ_ε in the following way, let $G \in C_c^\infty(\mathbb{R}^d)$ be a probability density, such that $M(G) = 1$ and $H(G) < \infty$. Define $G_\varepsilon(\cdot) := \varepsilon^{-2d} G(\frac{\cdot}{\varepsilon^2})$, and

$$\gamma_\varepsilon(x, y) := \tilde{\rho}_k^{n+1}(x) G_\varepsilon\big(y - x\big),$$

as the joint distribution with first marginal $\tilde{\rho}_k^{n+1}$, and second marginal $\rho_\varepsilon(y) := \int \gamma_\varepsilon(x, y)dx$. One can then calculate/express $H(\gamma_\varepsilon), \mathcal{F}(\gamma_\varepsilon), (c_h, \gamma_\varepsilon)$ in terms of $\tilde{\rho}_k^{n+1}$ (see [ADdR21, Lemma 4.3]). Comparing the performance of ρ_k^{n+1} against ρ_ε in (4.4) we get, making use of the scaling (4.1) and that $H(\tilde{\gamma}_k^{n+1,c}) \geq H(\rho_k^{n+1}) + H(\tilde{\rho}_k^{n+1})$, the following inequality

$$(c_h, \tilde{\gamma}_k^{n+1,c}) \leq Ch^2\Big(M(\tilde{\rho}_k^{n+1}) + 1\Big) - \varepsilon H(\rho_k^{n+1}) + 2h\Big(\mathcal{F}(\tilde{\rho}_k^{n+1}) - \mathcal{F}(\rho_k^{n+1})\Big). \tag{4.8}$$

We are able to obtain bounded 2nd moments, energy, and entropy estimates in an almost identical fashion as to Lemma 3.9, specifically in (3.13) we use $(c_h, \tilde{\gamma}_k^{i+1,c})$ in place of $W_{c_h}(\tilde{\rho}^{i+1}, \rho^{i+1})$, and apply (4.8). Moreover, summing (4.8) and using such estimates yields the bound

$$\sum_{n=0}^{N-1} (c_h, \tilde{\gamma}_k^{n+1,c}) \leq Ch. \tag{4.9}$$

It is easy to conclude that we also have

$$\sum_{n=0}^{N-1} W_2^2(\tilde{\rho}_k^{n+1}, \rho_k^{n+1}) \leq Ch \quad \text{and} \quad \sum_{n=0}^{N-1} W_2^2(\rho_k^{n}, \tilde{\rho}_k^{n+1}) \leq Ch. \tag{4.10}$$

The Discrete Euler-Lagrange Equation and concluding the convergence. Since ρ_k^{n+1} solves the minimisation problem (4.4), the associated discrete Euler-Lagrange equation reads, for any $\varphi \in C_c^\infty(\mathbb{R}^d)$,

$$0 = \frac{1}{h}\int_{\mathbb{R}^{2d}} \Big\langle x - y, \nabla\varphi(y)\Big\rangle d\tilde{\gamma}^{n+1,c}(x, y) + \delta\mathcal{F}(\rho_k^{n+1}, A_h\nabla\varphi) - \frac{\varepsilon}{2h}\int_{\mathbb{R}^d} \rho_k^{n+1}(y)\text{div}\big(A_h\nabla\varphi(y)\big)dy. \tag{4.11}$$

Therefore, the analogous result to Lemma 3.15 is

$$\begin{aligned}
&\int_0^T \int_{\mathbb{R}^d} \rho_k^\dagger(t, x)\big(\partial_t\varphi(t, x) + b[\rho(t, \cdot)](x) \cdot \nabla\varphi(t, x)\big)dxdt \\
&= -\int_{\mathbb{R}^d} \rho^0(x)\varphi(0, x)dx - \sum_{n=0}^{N-1}\Big(h\delta\mathcal{F}(\rho_k^{n+1}, A_h\nabla\varphi(t_n, \cdot)) + \frac{\varepsilon}{2h}\int_{t_n}^{t_{n+1}}\int_{\mathbb{R}^d} \rho_k^{n+1}(y)\text{div}\big(A_h\nabla\varphi(t_n, y)\big)dy\Big) + O(h).
\end{aligned} \tag{4.12}$$

The convergence claimed in (4.6) and (4.7) follows by a priori estimates identical to those of Lemma 3.13. Hence to complete the proof of Theorem 4.2 we need only to deal with the term appearing from the regularisation

$$\sum_{n=0}^{N-1} \frac{\varepsilon}{2h}\int_{t_n}^{t_{n+1}}\int_{\mathbb{R}^d} \rho_k^{n+1}(y)\text{div}\big(A_h\nabla\varphi(t_n, y)\big)dy,$$

and show it goes to zero as $\varepsilon, h \to 0$. This is clear by a similar argument to that in the end of the proof of Section 3.4: using the convergence (4.6) of ρ_k and the scaling (4.1) which implies that $\frac{\varepsilon}{h} \to 0$.

□

5 Examples

In this section, we present four concrete examples of evolutionary equations that can all be written in the general form (1.6): the Vlasov-Fokker-Planck equation, a degenerate non-linear diffusion equation of Kolmogorov-type, the linear Wigner FPE, the regularised Vlasov-Poisson FPE, and a generalised Vlasov-Langevin equation.

Applicability of Theorems 2.4 and 4.2 to the examples. In all these examples, we will show explicitly the (non-local) vector field b and the diffusion matrix A. Assuming the drift vector fields and the diffusion matrix are such that Assumption 2.1 is satisfied, then Theorem 2.4 and/or of Theorem 4.2 provides novel operator-splitting variational schemes for solving these evolutionary equations. It will be clear that from the explicit formulas that A is symmetric positive semi-definite and $b[\rho]$ is divergence-free. It remains to consider the first and second conditions in (2.21), which are assumptions on the growth and regularity of the vector field, and (2.22). In all examples, the vector field $b[\rho](x)$ consists of a local part and a non-local part, where the non-local part is a convolution of ρ with an interaction kernel K, namely of the form

$$b_{\text{non-local}}[\rho](x) := (K * \rho)(x) = \int K(x - x')d\rho(x').$$

Assumption 2.1 is satisfied for instance when the local part is Lipschitz, and the kernel K is uniformly bounded, Lipschitz, and differentiable. Under this assumption, (2.21) and (2.22) are straightforward for the local part. Now we show that the non-local part also fulfills these assumptions if it is uniformly bounded and Lipschitz.

Lemma 5.1. Suppose that K is uniformly bounded and Lipschitz. Then for all $\rho, \mu \in \mathcal{P}_2(\mathbb{R}^d),\ z \in \mathbb{R}^d$, we have

$$\int_{\mathbb{R}^d} \|K * \rho(z) - K * \mu(z)\|^2 \, d\rho(z) \leq CW_2^2(\rho, \mu), \tag{5.1}$$

$$\|K * \mu(z)\| \leq C(1 + \|z\|), \tag{5.2}$$

$$K * \mu \in W^{1,1}_{\text{loc}}(\mathbb{R}^d). \tag{5.3}$$

Proof. We first prove (5.1). We will use the following equivalent formulation of the Wasserstein distance [Vil21]

$$W_2^2(\rho, \mu) = \inf \Big[\mathbb{E}(\|X - Y\|^2)\Big], \tag{5.4}$$

where the infimum is taken over all couples of random variables X and Y with $Y \sim \rho$ and $X \sim \mu$.

Now let $\mu, \rho \in \mathcal{P}_2(\mathbb{R}^d)$ and take random variables X and Y with $Y \sim \rho$ and $X \sim \mu$. We have

$$\begin{aligned}
\int_{\mathbb{R}^d} \|K * \rho(z) - K * \mu(z)\|^2 \, d\rho(z) &= \int_{\mathbb{R}^d} \| \int_{\mathbb{R}^d} K(z - z')\big(d\rho(z') - d\mu(z')\big)\|^2 \, d\rho(z) \\
&= \int_{\mathbb{R}^d} \|\mathbb{E}[K(z - Y) - K(z - X)]\|^2 \, d\rho(z) \\
&\leq \int_{\mathbb{R}^d} \mathbb{E}[\|K(z - Y) - K(z - X)\|^2] \, d\rho(z) \\
&\leq C \int_{\mathbb{R}^d} \mathbb{E}[\|Y - X\|^2] \, d\rho(z) = C\mathbb{E}[\|Y - X\|^2].
\end{aligned}$$

Taking infimum over all X and Y and using (5.4) yields (5.1). Verifying (5.2) is straightforward by the uniform bound on K. Finally, we check (5.3). Let Ω be an arbitrary compact set in $\mathbb{R}^d$. Firstly it is clear that $K * \mu(z) \in L^1_{\text{loc}}(\mathbb{R}^d)$ since K is uniformly bounded. Let $i, j \in 1, \ldots, d$. It remains to show $\partial_{z_j} K_i * \mu(z) \in L^1_{\text{loc}}(\mathbb{R}^d)$. In fact, since $\|\nabla K_i\| \leq C$, we have

$$\int_\Omega \| \int \partial_{z_j} K_i(z - z')d\mu(z')\|dz \leq \int_\Omega \| \int \nabla K_i(z - z')d\mu(z')\|dz \leq C|\Omega|.$$

This completes the proof of this lemma. □

We now discuss concrete applications of our work.

5.1 Linear Wigner FPE

The Wigner Fokker-Planck equation is the quantum mechanical analogue to the classical VFPE discussed in Section 5.2. It has been used in the modelling of semiconductor devices, see [MRS90] and references therein. In particular, the linearized Wigner Fokker-Planck equation is

$$\partial_t \rho = -v \cdot \nabla_x \rho + \beta \text{div}_v\big(v\rho\big) + \sigma\Delta_v\rho + \alpha\Delta_x\rho + 2\lambda\text{div}_v(\nabla_x\rho), \tag{5.5}$$

where the variables $x, v \in \mathbb{R}^d$, and the constant $\beta > 0$ is the friction parameter, and $\sigma, \alpha, \lambda > 0$ form the diffusion matrix of the system. The above equation is an instance of (1.6) with

$$b(x, v) = \begin{pmatrix} (\frac{\beta\lambda}{\sigma} + 1)v \\ 0 \end{pmatrix}, \qquad A = \begin{pmatrix} \alpha & \lambda \\ \lambda & \sigma \end{pmatrix}, \qquad f(x, v) = f(v) = \frac{\beta}{2\sigma}\|v\|^2,$$

Note that each entry of A stands for a $d \times d$ identity matrix times that entry. Note that the above equation is local and non-degenerate. In the operator-splitting scheme, one needs not to perturb the diffusion matrix A (see discussion at the end of the introduction of the scheme in Section 2.2).

It should be mentioned that the full Wigner Fokker-Planck equation includes a pseudo-differential operator acting on a non-local term coupled to the Poisson equation. At the moment, it is not clear to us whether we can extend our results to the full equation.

5.2 Vlasov-Fokker-Planck Equation (VFPE)

The Vlasov-Fokker-Planck equation, which describes the probability of finding a particle at time t with position $x \in \mathbb{R}^d$ and velocity $v \in \mathbb{R}^d$ moving under the influence of an external potential ∇g, an interaction force K, a frictional force ∇f and a stochastic noise, is given by

$$\partial_t \rho = -v \cdot \nabla_x \rho + \nabla g \cdot \nabla_v \rho + \mathrm{div}_v\big(\rho K * \rho\big) + \mathrm{div}_v\big(\rho \nabla f\big) + \Delta_v \rho.$$

It is the forward Kolmogorov equation of the follow stochastic differential equation

$$\begin{cases} dX_t = V_t dt \\ dV_t = -\big(K * \rho_t(X_t) + \nabla g(X_t)\big)dt - \nabla f(V_t)dt + \sqrt{2} dW_t \\ \rho_t = \mathrm{Law}(X_t, V_t). \end{cases}$$

The VFPE is a special case of (1.6) with

$$b[\rho](x,v) = \begin{pmatrix} v \\ -\big(\nabla g(x) + K * \rho(x)\big) \end{pmatrix}, \qquad A = \begin{pmatrix} 0 & 0 \\ 0 & I \end{pmatrix}, \quad f(x,v) = f(v), \quad (x,v) \in \mathbb{R}^{2d}. \tag{5.6}$$

When there is no interaction (i.e., $K = 0$) the VFPE reduces to the Kramers equation. As mentioned in the introduction, various variational schemes have been developed for the Kramers equation [DPZ14, Hua00, CG04, MS20a], see [ADdR21] for extensions of these work to non-linear models. This paper not only provides a novel scheme but also incorporates the interaction force.

5.3 Regularized Vlasov-Poisson-Fokker-Planck equation

The Vlasov-Poisson-Fokker-Planck equation is given by

$$\partial_t \rho = -v \cdot \nabla_x \rho + \nabla\big(g(x) + \phi[\rho](x)\big) \cdot \nabla_v \rho - \beta \mathrm{div}_v\big(\rho v\big) + \sigma \Delta_v \rho. \tag{5.7}$$

for positive constants σ, β and variables $x, v \in \mathbb{R}^d$, where ϕ solves the Poisson equation

$$\Delta \phi(x) = -\int_{\mathbb{R}^d} \rho(x,v) dv,$$

the solution of which is

$$\phi(x) = \int_{\mathbb{R}^{2d}} \Gamma(x-y)\rho(y,v) dy dv, \tag{5.8}$$

for Γ defined as

$$\Gamma(r) := \begin{cases} \frac{\omega_d}{\|r\|^{\frac{d-2}{2}}} & \text{for } d > 2, \\ \omega_2 \log \|r\| & \text{for } d = 2, \end{cases}$$

where ω_d is the surface area of the unit ball in $\mathbb{R}^d$. This equation is of great importance in plasma physics, as it models a cloud of charged particles influencing each other through a Coulomb interaction, whilst subject to deterministic and random forcing, and friction. Since Γ is singular our methods can not be directly applied where it is easy to check that (2.21) fails to hold. However, if we consider ϕ^ϵ defined analogously to (5.8) but with Γ replaced by

$$\Gamma^\epsilon(r) = \begin{cases} \frac{\omega_d}{\big(\|r\|^2 + \epsilon\big)^{\frac{d-2}{2}}} & \text{for } d > 2 \\ \frac{\omega_d}{2} \log\big(\|r\|^2 + \epsilon\big) & \text{for } d = 2 \end{cases},$$

then we arrive at the regularised Vlasov-Poisson Fokker-Planck equation

$$\partial_t \rho^\epsilon = -v \cdot \nabla_x \rho^\epsilon + \nabla\big(g(x) + \phi^\epsilon[\rho^\epsilon](x)\big) \cdot \nabla_v \rho^\epsilon - \beta \mathrm{div}_v\big(\rho^\epsilon v\big) + \sigma \Delta_v \rho^\epsilon. \tag{5.9}$$

Here we have regularised the Kernel appearing in the convolution (this is different from the regularisation discussed in Section 4). For any $\epsilon > 0$, Γ^ϵ is no longer singular, and $\|\nabla\Gamma^\epsilon\|$ is uniformly bounded. Moreover, $\nabla\Gamma^\epsilon$ is Lipschitz, indeed, the Hessian is uniformly bounded which can be seen from the following explicit computations, for $d \geq 2$ we have

$$\partial_{x_i}\partial_{x_j}\Gamma^\epsilon(x) = C_d \begin{cases} \frac{1}{(\|x\|^2+\epsilon)^{\frac{d}{2}}} - \frac{dx_i^2}{(\|x\|^2+\epsilon)^{\frac{d+2}{2}}} & \text{if } i = j, \\ -\frac{dx_i x_j}{(\|x\|^2+\epsilon)^{\frac{d+2}{2}}} & \text{if } i \neq j, \end{cases}$$

for some constant C_d depending only on the dimension. Hence by Lemma 5.1 assumptions (2.21) and (2.22) are satisfied.

The solutions ρ^ϵ to (5.9) have been shown to converge (as $\epsilon \to 0$) to the solution of the original system (5.7) [CS95]. One-step variational schemes (in the space of probability measures) have already been proposed for (5.9), see [HJ00]. However, the cost function used in [HJ00] is not a metric, the free energy depends on the time step *and* contains a mix of conservative and dissipative terms. Our approach in the present paper naturally splits the conservative and dissipative dynamics.

5.4 A generalised Vlasov-Langevin equation

Next we consider the following generalised Vlasov-Langevin equation [OP11, Duo15]

$$\partial_t\rho = -p\cdot\nabla_q\rho + \big(\mathcal{A}(q) + K*\rho(q) - \sum_{j=1}^m \Lambda^j z^j\big)\cdot\nabla_p\rho + \sum_{j=1}^m \mathrm{div}_{z^j}[(\Lambda^j p + \alpha^j z^j)\rho] + \Delta_z\rho_t. \tag{5.10}$$

Note that in the above equation, the coordinates are $(q,p,z) \in \mathbb{R}^{2d+md}$, with $q,p \in \mathbb{R}^d$ and $z \in \mathbb{R}^{md}$ for some $m \in \mathbb{N}$. Equation (5.10) is the forward Kolmogorov equation of the SDE system

$$\begin{cases} dQ_t = P_t\,dt, \\ dP_t = -\mathcal{A}(Q_t)\,dt - K*\rho_t(Q_t)\,dt + \sum_{j=1}^m \Lambda^j Z_t^j\,dt, \\ dZ_t^j = -\Lambda^j P_t\,dt - \alpha^j Z_t^j\,dt + \sqrt{2}\,dW_t^j, \quad j = 1,\dots,m. \\ \rho_t = \mathrm{Law}(Q_t, P_t, Z_t^1, \dots, Z_t^m), \end{cases} \tag{5.11}$$

where W_t^j are independent $d-$dimensional Brownian motions, $\mathcal{A}: \mathbb{R}^d \to \mathbb{R}^d$ is an external potential, $K: \mathbb{R}^d \to \mathbb{R}^d$ an interaction kernel, $\Lambda^j, \alpha^j \in \mathbb{R}^{d\times d}$ constant diagonal matrices $\forall j \in \{1,\dots,m\}$. When $\mathcal{A}$ is the gradient of a potential and no Kernel is present, $K = 0$, then (5.11) can be viewed as the coupling of a deterministic Hamiltonian system (Q_t, P_t) to a heat bath Z_t, the literature on this subject is vast. In this setup, for large m the Markovian system (5.11) approximates the Generalised Langevin equation (GLE). The GLE serves as a standard model in non-Markovian non-equilibrium statistical mechanics, where the Hamiltonian system is in contact with one or more heat baths. The heat baths are modeled by the linear wave equation and are initialised according to Gibbs distribution, see [Kup04, OP11, RB06] and references therein. When $K \neq 0$, the mean field term $K*\rho$ models the particle interactions in the underlying deterministic system (via the positions Q_t). In this case, (5.11) is the McKean-Vlasov limit of a system of weakly interacting particles [Duo15].

Again the generalised Vlasov-Langevin equation is another example of (1.6) where free vector field, diffusion matrix, and potential energy are given by

$$b[\rho](q,p,z) = \begin{pmatrix} p \\ -\mathcal{A}(q) - K*\rho(q) + \sum_{j=1}^m \Lambda^j z^j \\ -\Lambda^1 p \\ \vdots \\ -\Lambda^m p \end{pmatrix}, \quad A = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & I \end{pmatrix}, \quad f(q,p,z) = f(z) = \sum_{j=1}^m \frac{1}{2}\|\alpha^j z^j\|^2,$$

where I is the $md \times md$ identity matrix.

5.5 A degenerate diffusion equation of Kolmogorov-type

The final example that we consider is the following non-linear degenerate equation of Kolmogorov type

$$\partial_t\rho = -\sum_{i=2}^n x_i\cdot\nabla_{x_{i-1}}\rho + \mathrm{div}_{x_n}(\nabla f(x_n)\rho) + \Delta_{x_n}\rho. \tag{5.12}$$

In the above equation, the coordinates are $\mathbf{x} = \big(x_1, x_2, \ldots, x_{n-1}, x_n\big)^T$, where $x_i \in \mathbb{R}^d$ for each $i \in \{1, \ldots, n\}$. Equation (5.12) is the forward Kolmogorov equation of the associated stochastic differential equations

$$\begin{cases} dX_1 = X_2\, dt \\ dX_2 = X_3\, dt \\ \quad \vdots \\ dX_{n-1} = X_n\, dt \\ dX_n = -\nabla f(X_n)\, dt + \sqrt{2}\, dW_t, \end{cases} \tag{5.13}$$

where W_t is a d-dimensional Wiener process. System (5.13) describes the motion of n coupled oscillators connected to their nearest neighbours with the last oscillator additionally forced by a random noise which propagates through the system. The simplest cases of $n = 1, n = 2$ correspond to the heat equation and Kramers' equation (with no background potential) respectively. When $n > 2$ this type of equations arise as models of simplified finite Markovian approximations of generalised Langevin dynamics [OP11], or harmonic oscillator chains [BL08, DM10]. Recent works [DT17, ADdR21] have constructed a one-step scheme for (5.12), however, the cost function used there (the mean squared derivative cost function [DT18][(11)]), although explicit, does not take a simple form.

Equation (5.12) is yet another special case of (1.6) with the following divergence free vector field, diffusion matrix, and potential energy

$$b(\mathbf{x}) = (x_2, x_3, \ldots, x_n, 0)^T, \qquad A = \begin{pmatrix} 0 & 0 \\ 0 & I \end{pmatrix}, \qquad f(\mathbf{x}) = f(x_n), \tag{5.14}$$

where, in the matrix A, I is the $d \times d$-dimensional identity matrix, and remaining elements are all 0.

A Appendix

In this section, we provide detailed computations and proof for some technical results used in previous sections for the completeness.

Lemma A.1. For any $h > 0$, and any μ and ν in $\mathcal{P}_2(\mathbb{R}^d)$, it is true that

$$M(\nu) \le 2\big(W_2^2(\mu, \nu) + M(\mu)\big) \tag{A.1}$$

and

$$M(\nu) \le C\big(W_{c_h}(\mu, \nu) + M(\mu)\big). \tag{A.2}$$

Proof. The result (A.1) for W_2 is obvious. For (A.2) just use (A.1) in conjunction with (3.4). □

Lemma A.2. [JKO98, Proposition 4.1] There exists a $C > 0$ and $0 < \alpha < 1$ such that

$$H(\mu) \ge -C(M(\mu) + 1)^\alpha, \ \ \forall \mu \in \mathcal{P}_2^r(\mathbb{R}^d) \quad \text{and} \quad H_-(\mu) \le C(M(\mu) + 1)^\alpha, \ \ \forall \mu \in \mathcal{P}_2^r(\mathbb{R}^d). \tag{A.3}$$

Moreover, H is weakly lower semi-continuous under bounded moments, i.e., if $\{\mu_k\}_{k\in\mathbb{N}} \subset \mathcal{P}_2(\mathbb{R}^d)$, $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ with $\mu_k \rightharpoonup \mu$, and there exists $C > 0$ such that $M(\mu_k), M(\mu) < C$ for all $k \in \mathbb{N}$, then

$$H(\mu) \le \liminf_{k\to\infty} H(\mu_k). \tag{A.4}$$

Lemma A.3. Let $h > 0$. Given $\mu, \nu \in \mathcal{P}_2^r(\mathbb{R}^d)$ there exists $\gamma \in \Pi(\mu, \nu)$ such that

$$W_{c_h}(\mu, \nu) = (c_h, \gamma).$$

Moreover, the map $\gamma \mapsto (c_h, \gamma)$ is weakly lower semi-continuous.

Proof. See [Vil08, Theorem 4.1]. □

Lemma A.4 (Lower Semi-Continuity of the functionals)**.** Let $\{\nu_k\}_{k\in\mathbb{N}} \subset \mathcal{P}_2^r(\mathbb{R}^d)$, $\mu, \nu \in \mathcal{P}_2^r(\mathbb{R}^d)$, with $\nu_k \rightharpoonup \nu$ as $k \to \infty$. Assume that for all $k \in \mathbb{N}$ the probability measures ν_k, μ, ν have uniformly bounded entropy and second moments. Then

$$\mathcal{F}(\nu) \le \liminf_{k\to\infty} \mathcal{F}(\nu_k) \quad \text{and} \quad W_{c_h}(\mu, \nu) \le \liminf_{k\to\infty} W_{c_h}(\mu, \nu_k). \tag{A.5}$$

Proof. Let $\{\nu_k\}, \mu, \nu$ be as assumed above, and $\{\gamma_k\}$ be the associated optimal plans in $W_{c_h,\varepsilon}(\mu,\nu_k)$. Note $\{\gamma_k\} \subset \Pi(\mu,\{\nu_k\})$ (see notations Section 2.1). Since $\{\nu_k\}$ is weakly convergent then it is tight, and [Vil08, Lemma 4.4] implies that $\Pi(\mu,\{\nu_k\})$ is so too. Extracting (and relabelling) a subsequence $\{\gamma_k\}$, we know that (as $k\to\infty$) $\gamma_k \rightharpoonup \gamma \in \mathcal{P}(\mathbb{R}^{2d})$. In fact $\gamma \in \Pi(\mu,\nu)$ since the weak convergence of γ_k implies the weak convergence of its marginals (and we know $\nu_k \rightharpoonup \nu$). Now, the lower semi-continuity described in Lemma A.3 implies that

$$\liminf_{k\to\infty} W_{c_h}(\mu,\nu_k) = \liminf_{k\to\infty} \frac{1}{2h}(c_h,\gamma_k) \geq \frac{1}{2h}(c_h,\gamma) \geq W_{c_h}(\mu,\nu).$$

The proof for $\mathcal{F}$ is identical to [ADdR21, Lemma 3.8]. □

A.1 Well-Posedness

Proof of Proposition 3.4. Let $0<h<1$ and $\mu,\nu\in\mathcal{P}_2^r(\mathbb{R}^d)$. Define $J_{c_h}(\mu,\nu) := \frac{1}{2h}W_{c_h}(\mu,\nu)+\mathcal{F}(\nu)$, then we have

$$\begin{aligned} J_{c_h}(\mu,\nu) = \frac{1}{2h}W_{c_h}(\mu,\nu) + M(\mu) + \mathcal{F}(\nu) - M(\mu) \geq & W_{c_h}(\mu,\nu) + M(\mu) + \mathcal{F}(\nu) - M(\mu) && \text{(A.6)}\\ \geq & C_1 M(\nu) + H(\nu) - M(\mu) && \text{(A.7)}\\ \geq & C_1 M(\nu) - C_2(1+M(\nu))^\alpha + C_\mu, && \text{(A.8)} \end{aligned}$$

where in (A.6) we have used that $h<1$, in (A.7) we used Lemma A.1 and the non-negativity of f, and in (A.8) we used Lemma A.2. We emphasize that the constants $C_1, C_2 > 0$ are independent of μ,ν and $C_\mu > 0$ is independent of ν. Inequality (A.8) implies that $\nu \mapsto J(\mu,\nu)$ is bounded from below. Note that there exists a $\nu\in\mathcal{P}_2^r(\mathbb{R}^d)$ such that $J_{c_h}(\mu,\nu)<\infty$, for example, take $\nu=\mu$ (and the product plan).

Let $\{\nu_k\}$ be a minimising sequence and note that this implies $M(\nu_k), H(\nu_k)$ are uniformly bounded. The uniform boundedness of $M(\nu_k)$ implies tightness of $\{\nu_k\}$, and hence extracting a subsequence we have $\nu_k \rightharpoonup \nu^* \in \mathcal{P}(\mathbb{R}^d)$. Moreover, $\nu^*\in\mathcal{P}_2(\mathbb{R}^d)$ since uniformly bounded 2nd moments and weak convergence of $\{\nu_k\}$ implies that the limit has a bounded 2nd moment as well. Additionally, $\nu^*\in\mathcal{P}_2^r(\mathbb{R}^d)$ by the lower semi-continuity of H, see Lemma A.2. That ν^* is indeed the minimiser of (3.5) follows from the lower semi-continuity in Lemma A.4. Finally, the uniqueness follows by linearity of $F(\cdot)$, convexity of $W_2(\mu,\cdot)$, and the strict convexity of $H(\cdot)$. □

Acknowledgment

We would like to thank the anonymous referees for useful comments and suggestions which have helped us to improve the presentation of the article.

References

[ADdR21] D. Adams, M. H. Duong, and G. dos Reis, *Entropic regularisation of non-gradient systems*, 2021. arXiv:2104.04372, To appear in SIAM Journal on Mathematical Analysis.

[ADPZ11] S. Adams, N. Dirr, M. A. Peletier, and J. Zimmer, *From a large-deviations principle to the Wasserstein gradient flow: a new micro-macro passage*, Comm. Math. Phys. **307** (2011), no. 3, 791–815. MR2842966

[AGG+12] A. Arnold, I. M. Gamba, M. P. Gualdani, S. Mischler, C. Mouhot, and C. Sparber, *The Wigner-Fokker-Planck equation: stationary states and large time behavior*, Math. Models Methods Appl. Sci. **22** (2012), no. 11, 1250034, 31. MR2974172

[AGS08] L. Ambrosio, N. Gigli, and G. Savaré, *Gradient flows: in metric spaces and in the space of probability measures*, Springer Science & Business Media, 2008.

[BA15] M. Bowles and M. Agueh, *Weak solutions to a fractional Fokker-Planck equation via splitting and Wasserstein gradient flow*, Appl. Math. Lett. **42** (2015), 30–35. MR3294365

[Ber18] E. Bernton, *Langevin Monte Carlo and JKO splitting*, Conference on learning theory, 2018, pp. 1777–1798.

[BL08] T. Bodineau and R. Lefevere, *Large deviations of lattice Hamiltonian dynamics coupled to stochastic thermostats*, J. Stat. Phys. **133** (2008), no. 1, 1–27. MR2438895

[BS18] T. Bühler and D. A. Salamon, *Functional analysis*, Graduate Studies in Mathematics, vol. 191, American Mathematical Society, Providence, RI, 2018. MR3823238

[CDPS17] G. Carlier, V. Duval, G. Peyré, and B. Schmitzer, *Convergence of entropic schemes for optimal transport and gradient flows*, SIAM Journal on Mathematical Analysis **49** (2017), no. 2, 1385–1418.

[CG04] E. A. Carlen and W. Gangbo, *Solution of a model Boltzmann equation via steepest descent in the 2-Wasserstein metric*, Arch. Ration. Mech. Anal. **172** (2004), no. 1, 21–64. MR2048566

[CL17] G. Carlier and M. Laborde, *A splitting method for nonlinear diffusions with nonlocal, nonpotential drifts*, Nonlinear Analysis: Theory, Methods & Applications **150** (2017), 1–18.

[CM17] E. A. Carlen and J. Maas, *Gradient flow and entropy inequalities for quantum Markov semigroups with detailed balance*, Journal of Functional Analysis **273** (2017), no. 5, 1810–1869.

[CS95] J. A. Carrillo and J. Soler, *On the initial value problem for the Vlasov-Poisson-Fokker-Planck system with initial data in L^p spaces*, Math. Methods Appl. Sci. **18** (1995), no. 10, 825–839. MR1343393

[Cut13] M. Cuturi, *Sinkhorn distances: Lightspeed computation of optimal transport*, Advances in neural information processing systems **26** (2013), 2292–2300.

[DL19] M. H. Duong and Y. Lu, *An operator splitting scheme for the fractional kinetic Fokker-Planck equation*, Discrete & Continuous Dynamical Systems **39** (2019), no. 10, 5707–5727.

[DL89] R. J. DiPerna and P.-L. Lions, *Ordinary differential equations, transport theory and Sobolev spaces*, Invent. Math. **98** (1989), no. 3, 511–547. MR1022305

[DM10] F. Delarue and S. Menozzi, *Density estimates for a random noise propagating through a chain of differential equations*, Journal of functional analysis **259** (2010), no. 6, 1577–1630.

[DMC20] S. Di Marino and L. Chizat, *A tumor growth model of Hele-Shaw type as a gradient flow*, ESAIM: COCV **26** (2020), 103.

[DPZ13] M. H. Duong, M. A. Peletier, and J. Zimmer, *GENERIC formalism of a Vlasov-Fokker-Planck equation and connection to large-deviation principles*, Nonlinearity **26** (2013), no. 11, 2951–2971. MR3129075

[DPZ14] M. H. Duong, M. A. Peletier, and J. Zimmer, *Conservative-dissipative approximation schemes for a generalized Kramers equation*, Math. Methods Appl. Sci. **37** (2014), no. 16, 2517–2540.

[DT17] M. H. Duong and H. M. Tran, *Analysis of the mean squared derivative cost function*, Mathematical Methods in the Applied Sciences **40** (2017), no. 14, 5222–5240.

[DT18] M. H. Duong and H. M. Tran, *On the fundamental solution and a variational formulation for a degenerate diffusion of Kolmogorov type*, Discrete Contin. Dyn. Syst. **38** (2018), no. 7, 3407–3438. MR3809088

[Duo15] M. H. Duong, *Long time behaviour and particle approximation of a generalised Vlasov dynamic*, Nonlinear Anal. **127** (2015), 1–16. MR3392354

[EPSS21] A. Esposito, F. S. Patacchini, A. Schlichting, and D. Slepčev, *Nonlocal-interaction equation on graphs: Gradient flow structure and continuum limit*, Archive for Rational Mechanics and Analysis **240** (2021May), no. 2, 699–760.

[GO16] R. Glowinski and S. J. Osher (eds.), *Splitting methods in communication, imaging, science, and engineering*, Scientific Computation, Springer, Cham, 2016. MR3587821

[HJ00] C. Huang and R. Jordan, *Variational formulations for Vlasov-Poisson-Fokker-Planck systems*, Math. Methods Appl. Sci. **23** (2000), no. 9, 803–843. MR1763126

[HJ13] H. J. Hwang and J. Jang, *On the Vlasov-Poisson-Fokker-Planck equation near Maxwellian*, Discrete Contin. Dyn. Syst. Ser. B **18** (2013), no. 3, 681–691. MR3007749

[Hua00] C. Huang, *A variational principle for the Kramers equation with unbounded external forces*, J. Math. Anal. Appl. **250** (2000), no. 1, 333–367. MR1893894

[JKO98] R. Jordan, D. Kinderlehrer, and F. Otto, *The variational formulation of the Fokker-Planck equation*, SIAM J. Math. Anal. **29** (1998), no. 1, 1–17. MR1617171

[JW18] P.-E. Jabin and Z. Wang, *Quantitative estimates of propagation of chaos for stochastic systems with $W^{-1,\infty}$ kernels*, Inventiones mathematicae **214** (2018Oct), no. 1, 523–591.

[Kra40] H. A. Kramers, *Brownian motion in a field of force and the diffusion model of chemical reactions*, Physica **7** (1940), 284–304. MR0002962 (2,140d)

[KS67] P. Knopp and R. Sinkhorn, *Concerning nonnegative matrices and doubly stochastic matrices.*, Pacific Journal of Mathematics **21** (1967), no. 2, 343 –348.

[Kup04] R. Kupferman, *Fractional kinetics in Kac-Zwanzig heat bath models*, J. Statist. Phys. **114** (2004), no. 1-2, 291–326. MR2032133

[Lab17] M. Laborde, *On some nonlinear evolution systems which are perturbations of Wasserstein gradient flows*, Topological optimization and optimal transport, 2017, pp. 304–332. MR3729381

[Lis09] S. Lisini, *Nonlinear diffusion equations with variable coefficients as gradient flows in Wasserstein spaces*, ESAIM Control Optim. Calc. Var. **15** (2009), no. 3, 712–740. MR2542579

[LWW21] C. Liu, C. Wang, and Y. Wang, *A structure-preserving, operator splitting scheme for reaction-diffusion equations with detailed balance*, Journal of Computational Physics **436** (2021), 110253.

[Maa11] J. Maas, *Gradient flows of the entropy for finite Markov chains*, Journal of Functional Analysis **261** (2011), no. 8, 2250–2292.

[MPR14] A. Mielke, M. A. Peletier, and D. R. M. Renger, *On the relation between gradient flows and the large-deviation principle, with applications to Markov chains and diffusion*, Potential Analysis **41** (2014Nov), no. 4, 1293–1327.

[MRS90] P. A. Markowich, C. A. Ringhofer, and C. Schmeiser, *Semiconductor equations*, Springer-Verlag, Vienna, 1990. MR1063852

[MS20a] A. Marcos and A. Soglo, *Solutions of a class of degenerate kinetic equations using steepest descent in Wasserstein space*, Journal of Mathematics **2020** (2020).

[MS20b] D. Matthes and B. Söllner, *Discretization of flux-limited gradient flows: γ-convergence and numerical schemes*, Mathematics of Computation **89** (2020), no. 323, 1027–1057.

[OP11] M. Ottobre and G. A. Pavliotis, *Asymptotic analysis for the generalized Langevin equation*, Nonlinearity **24** (2011), no. 5, 1629–1653. MR2793823

[Ött18] H. C. Öttinger, *GENERIC: Review of successful applications and a challenge for the future*, arXiv preprint arXiv:1810.08470 (2018).

[PC19] G. Peyré and M. Cuturi, *Computational optimal transport: with applications to data science*, Foundations and Trends® in Machine Learning **11** (2019), no. 5-6, 355–607.

[Pey15] G. Peyré, *Entropic approximation of wasserstein gradient flows*, SIAM Journal on Imaging Sciences **8** (2015), no. 4, 2323–2351.

[RB06] L. Rey-Bellet, *Open classical systems*, Open quantum systems. II, 2006, pp. 41–78. MR2248987

[Ris89] H. Risken, *The Fokker-Planck equation*, Second, Springer Series in Synergetics, vol. 18, Springer-Verlag, Berlin, 1989. Methods of solution and applications. MR987631

[San15] F. Santambrogio, *Optimal transport for applied mathematicians*, Birkäuser, NY **55** (2015), no. 58-63, 94.

[Ser20] S. Serfaty, *Mean field limit for Coulomb-type flows*, Duke Math. J. **169** (2020), no. 15, 2887–2935. With an appendix by Mitia Duerinckx and Serfaty. MR4158670

[Vil08] C. Villani, *Optimal transport: old and new*, Vol. 338, Springer Science & Business Media, 2008.

[Vil09] C. Villani, *Hypocoercivity*, Vol. 202, American Mathematical Soc., 2009.

[Vil21] C. Villani, *Topics in optimal transportation*, Vol. 58, American Mathematical Soc., 2021.

[YB13] Y. Yao and A. L Bertozzi, *Blow-up dynamics for the aggregation equation with degenerate diffusion*, Physica D: Nonlinear Phenomena **260** (2013), 77–89.