Exponential convergence to equilibrium for coupled systems of nonlinear degenerate drift diffusion equations

Lisa Beck, Daniel Matthes, Martina Zizza

Angaben zur Veröffentlichung / Publication details:

Beck, Lisa, Daniel Matthes, and Martina Zizza. 2023. "Exponential convergence to equilibrium for coupled systems of nonlinear degenerate drift diffusion equations." *SIAM Journal on Mathematical Analysis* 55 (3): 1766–1809. https://doi.org/10.1137/21M1466980.

SIAM J. MATH. ANAL.
Vol. 55, No. 3, pp. 1766–1809

EXPONENTIAL CONVERGENCE TO EQUILIBRIUM FOR COUPLED SYSTEMS OF NONLINEAR DEGENERATE DRIFT DIFFUSION EQUATIONS*

LISA BECK†, DANIEL MATTHES‡, AND MARTINA ZIZZA§

Abstract. We study the existence and long-time asymptotics of weak solutions to a system of two nonlinear drift-diffusion equations that has a gradient flow structure in the Wasserstein distance. The two equations are coupled through a cross-diffusion term that is scaled by a parameter $\varepsilon \geq 0$. The nonlinearities and potentials are chosen such that in the decoupled system for $\varepsilon = 0$, the evolution is metrically contractive, with a global rate $\Lambda > 0\Lambda > 0$. The coupling is a singular perturbation in the sense that for any $\varepsilon > 0$, contractivity of the system is lost. Our main result is that for all sufficiently small $\varepsilon > 0$, the global attraction to a unique steady state persists, with an exponential rate $\Lambda_\varepsilon = \Lambda - K\varepsilon$ for some $k > 0$. The proof combines results from the theory of metric gradient flows with further variational methods and functional inequalities.

Key words. drift diffusion system, Wasserstein gradient flow, long time asymptotics, exponential convergence

MSC codes. 35K40, 35B40, 35A15, 35Q92

DOI. 10.1137/21M1466980

1. Introduction.

In this paper, we analyze existence and long-time asymptotics of nonnegative unit-mass solutions u and v of the following coupled system of two degenerate nonlinear drift-diffusion equations on $\mathbb{R}^d$:

$$\begin{aligned} \partial_t u &= \operatorname{div}\big(u\,\nabla[F'(u) + \varepsilon\partial_u h(u,v) + \Phi]\big),\\ \partial_t v &= \operatorname{div}\big(v\,\nabla[G'(v) + \varepsilon\partial_v h(u,v) + \Psi]\big). \end{aligned} \tag{1.1}$$

Notice that the diffusive contributions $\operatorname{div}(u\nabla F'(u))$ and $\operatorname{div}(v\nabla G'(v))$ of the system (1.1) can also be expressed as $\Delta f(u)$ and $\Delta g(v)$, respectively, by introducing functions f and g via the relations $f'(r) = rF''(r)$ and $g'(r) = rG''(r)$ for $r > 0$. The precise hypotheses on the various functions are formulated in Section 1.3 below. Briefly, the nonlinearities F, G for the individual components are smooth convex functions that degenerate at zero, i.e., with $F'(0) = G'(0) = 0$; the coupling is moderated by a nonlinear function h with quantified bounds on derivatives; the coupling strength $\varepsilon > 0$ is small and the potentials Φ, Ψ are Λ-convex, with some $\Lambda > 0$. We prove the global existence of transient solutions $(u_\varepsilon(t), v_\varepsilon(t))_{t\geq 0}$ to (1.1) for initial data of finite energy. We show existence and uniqueness of a stationary solution $(\bar u_\varepsilon, \bar v_\varepsilon)$, and analyze its regularity. Finally, we obtain convergence of $(u_\varepsilon(t), v_\varepsilon(t))$ to $(\bar u_\varepsilon, \bar v_\varepsilon)$ in

*Received by the editors December 22, 2021; accepted for publication (in revised form) October 10, 2022; published electronically June 2, 2023.
https://doi.org/10.1137/21M1466980
Funding: The work of the authors was supported by the DFG Collaborative Research Center TRR 109, "Discretization in Geometry and Dynamics."
†Institut für Mathematik, Universität Augsburg, D-86159 Augsburg, Germany (lisa.beck@math.uni-augsburg.de).
‡Zentrum Mathematik/M8, Technische Universität München, D-80538 Garching, Germany (matthes@ma.tum.de).
§SISSA-ISAS, 34136 Trieste TS, Italy (mzizza@sissa.it).

$L^1(\mathbb{R}^d)$ as $t \to \infty$, with an exponential rate of the form $\Lambda - K\varepsilon$, with a constant $K > 0$ independent of the solution.

1.1. Key ideas. Our approach is a variational one: we consider, at least formally, the system (1.1) as a metric gradient flow of the energy functional

$$\mathbf{E}_\varepsilon(u,v) = \int_{\mathbb{R}^d} \big[F(u) + G(v) + u\Phi + v\Psi + \varepsilon h(u,v)\big]\,\mathrm{d}x \tag{1.2}$$

on the cross product of two copies of the space $\mathcal{P}_2^r(\mathbb{R}^d)$ of probability densities of finite second moment, endowed with the L^2-Wasserstein distance; the definitions are recalled in section 2 below.

In the decoupled limit $\varepsilon = 0$, the evolution of u and that of v are independent in (1.1), and the energy is the sum of two functionals, $\mathbf{E}_0(u,v) - \mathbf{E}_0(\bar{u}_0, \bar{v}_0) = \mathbf{L}_1(u) + \mathbf{L}_2(v)$, depending only on u and v, respectively:

$$\mathbf{L}_1(u) = \int_{\mathbb{R}^d} \big[F(u) - F(\bar{u}_0) + (u - u_0)\Phi\big]\,\mathrm{d}x,$$
$$\mathbf{L}_2(v) = \int_{\mathbb{R}^d} \big[G(v) - G(\bar{v}_0) + (v - v_0)\Psi\big]\,\mathrm{d}x.$$

Our hypotheses on F, G and Φ, Ψ imply that $\mathbf{L}_1$ and $\mathbf{L}_2$ are uniformly displacement convex of modulus Λ. By the general theory of L^2-Wasserstein gradient flows, the flow defined by (1.1) is Λ-contractive. For any solution pair $(u(t), v(t))$, this implies convergence of $\mathbf{L}_1(u(t))$ and $\mathbf{L}_2(v(t))$ to zero at an exponential rate $\exp(-2\Lambda t)$, and a posteriori also convergence of the solutions to their respective stationary states $\bar{u}_0$ and $\bar{v}_0$ in L^1 at half that rate $\exp(-\Lambda t)$.

In the following, we are concerned with the long-time asymptotics of (1.1) with a small but positive coupling strength $\varepsilon > 0$. The aforementioned behavior at $\varepsilon = 0$ calls for a perturbative approach. Unfortunately, the perturbation induced by means of the coupling is singular: our first result (see Proposition 2.4) is that (unless $\partial_{uv} h \equiv 0$) the functional $\mathbf{E}_\varepsilon$ *loses uniform displacement convexity* for any $\varepsilon > 0$. In fact,[1] $\mathbf{E}_\varepsilon$ is not even uniformly displacement *semiconvex* of some negative modulus $\lambda < 0$. For this reason, the general machinery of metric gradient flows does not provide any result on exponential convergence anymore.

To make the long-time asymptotics accessible to perturbative methods, we blend the strong but elegant methods from gradient flow theory with more robust estimates related to the energy method. Our ansatz is to split $\mathbf{E}_\varepsilon$ in a particular way:

$$\mathbf{E}_\varepsilon(u,v) - \mathbf{E}_\varepsilon(\bar{u}_\varepsilon, \bar{v}_\varepsilon) = \mathbf{L}_\varepsilon(u,v) + \varepsilon \mathbf{A}_\varepsilon(u,v). \tag{1.3}$$

Here $\mathbf{L}_\varepsilon(u,v) = \mathbf{L}_{1,\varepsilon}(u) + \mathbf{L}_{2,\varepsilon}(v)$ is the sum of two "good" functionals, each depending on one density only,

$$\mathbf{L}_{1,\varepsilon}(u) = \int_{\mathbb{R}^d} \big[F(u) - F(\bar{u}_\varepsilon) + (u - \bar{u}_\varepsilon)\Phi_\varepsilon\big]\,\mathrm{d}x, \tag{1.4}$$
$$\mathbf{L}_{2,\varepsilon}(v) = \int_{\mathbb{R}^d} \big[G(v) - G(\bar{v}_\varepsilon) + (v - \bar{v}_\varepsilon)\Psi_\varepsilon\big]\,\mathrm{d}x,$$

[1] To avoid confusion, thanks to our hypotheses, $\mathbf{E}_\varepsilon$ remains convex in the usual flat sense for sufficiently small $\varepsilon > 0$—it is the (lack of) *displacement* convexity that is significant for the long-time asymptotics.

with perturbed potentials obtained from the stationary solution $(\bar{u}_\varepsilon, \bar{v}_\varepsilon)$ as follows:

$$\Phi_\varepsilon = \Phi + \varepsilon\partial_u h(\bar{u}_\varepsilon, \bar{v}_\varepsilon), \quad \Phi_\varepsilon = \Psi + \varepsilon\partial_v h(\bar{u}_\varepsilon, \bar{v}_\varepsilon),$$

chosen such that the nonnegative functionals $\mathbf{L}_{1,\varepsilon}$ and $\mathbf{L}_{2,\varepsilon}$ are zero precisely for $\bar{u}_\varepsilon$ and $\bar{v}_\varepsilon$, respectively, while $\mathbf{A}_\varepsilon$ is the "bad" functional that contains the coupling,

$$\mathbf{A}_\varepsilon(u,v) = \int_{\mathbb{R}^d} \big[h(u,v) - h(\bar{u}_\varepsilon, \bar{v}_\varepsilon) - (u - \bar{u}_\varepsilon)\,\partial_u h(\bar{u}_\varepsilon, \bar{v}_\varepsilon) - (v - \bar{v}_\varepsilon)\,\partial_v h(\bar{u}_\varepsilon, \bar{v}_\varepsilon)\big]\,\mathrm{d}x.$$

From a detailed analysis of the stationary solution $(\bar{u}_\varepsilon, \bar{v}_\varepsilon)$ by means of variational methods, we obtain bounds on $\partial_u h(\bar{u}_\varepsilon, \bar{v}_\varepsilon)$ and $\partial_v h(\bar{u}_\varepsilon, \bar{v}_\varepsilon)$ in $C^2(\mathbb{R}^d)$, uniformly for small $\varepsilon \geq 0$; see Theorem 1.3 and Corollary 3.9. This implies that Φ_ε and Ψ_ε above are still uniformly convex, with a diminished modulus $\Lambda - K_0\varepsilon$ for some $K_0 \geq 0$. By the general theory, $\mathbf{L}_{1,\varepsilon}$ and $\mathbf{L}_{2,\varepsilon}$ are uniformly displacement convex with the same modulus $\Lambda - K_0\varepsilon$, and so is their sum $\mathbf{L}_\varepsilon(u,v)$ on the product space.

The central step in the proof of our main result on the asymptotic behavior (see Theorem 1.5) is to show that $\mathbf{L}_\varepsilon$ is a Lyapunov functional that decays to zero at rate $\exp(-2[\Lambda - K\varepsilon]t)$, with some $K > K_0$. The dissipation of $\mathbf{L}_\varepsilon$ along the flow of $\mathbf{E}_\varepsilon$ consists of two contributions: the first is the auto-dissipation of $\mathbf{L}_\varepsilon$ by its own flow, that provides a Gronwall-type estimate thanks to uniform displacement convexity; see, e.g., [15],

(1.5)
$$|\partial\mathbf{L}_\varepsilon|^2 = \int_{\mathbb{R}^d} \Big(u\big|\nabla\big[F'(u) + \Phi_\varepsilon\big]\big|^2 + v\big|\nabla\big[G'(v) + \Psi_\varepsilon\big]\big|^2\Big)\,\mathrm{d}x \geq (1 - K_0\varepsilon)\mathbf{L}_\varepsilon(u,v).$$

The second contribution is the variation of $\mathbf{L}_\varepsilon$ along the flow of $\varepsilon\mathbf{A}_\varepsilon$. It is not of definite sign, in general, but can be controlled by an ε-amount of the integral expression for $|\partial\mathbf{L}_\varepsilon|^2$ above, at the price of reducing the rate of decay from $\Lambda - K_0\varepsilon$ to $\Lambda - K\varepsilon$. Obtaining the aforementioned control is a main technical challenge in the proof. It rests on a variety of elementary and functional analytic estimates that are established throughout sections 4.5 and 4.6.

The smallness of $\varepsilon > 0$ plays a role at various points of our considerations. Note that a result on global equilibration in the style of Theorem 1.5 cannot hold without any size restriction on ε, since for a generic choice of h, even the usual (flat) convexity is destroyed for sufficiently large values of ε, which could produce, e.g., multiple critical points. The threshold ε^* that we use is presumably significant smaller than the onset of flat nonconvexity: it is chosen to facilitate a variety of perturbative estimates, the main one being the aforementioned control of $\mathbf{L}_\varepsilon$'s variation along the flow of $\mathbf{A}_\varepsilon$ in terms of the integral expression in (1.5) above. Optimizing the admissible range of ε is a subject of future research.

We emphasize that the novelty of our result does not lie in the proof of convergence to equilibrium as such—a *qualitative* result could be obtained at almost no cost, e.g., from the LaSalle principle—but in the *quantitative* estimate on the convergence to equilibrium, with the exponential rate $\Lambda - K\varepsilon$. It is further significant that (1.1) is considered on $\mathbb{R}^d$, and that the steady state $(\bar{u}_\varepsilon, \bar{v}_\varepsilon)$ is compactly supported; hence, there is no standard inequality like Poincaré or log-Sobolev to conclude exponential convergence, not even at *some* smaller rate.

1.2. Positioning of our results. Coupled systems of nonlinear drift-diffusion equations are ubiquitous. They are used in the modeling of chemical reactions [39], flows in porous media [11], semiconductor devices [35], population dynamics [8], rival gangs in a city [4], segregation of species [9], just to name a few of the countless applications. The literature concerning the very natural question about long-time asymptotics is huge, albeit mostly focused on such systems with a particular rigid algebraic structure of the diffusion (being diagonal, or even linear) but with additional source terms, describing, e.g., reactions.

We briefly recall the situation for *scalar* drift-diffusion equations. The first proofs of exponential convergence to equilibrium in degenerate parabolic equations of the type $\partial_t u = \Delta f(u) + \operatorname{div}(u\nabla\Phi)$ has been given in the case $f(u) = u^m$ and $\Phi(x) = \frac{1}{2}|x|^2$ on $\mathbb{R}^d$ by a nonlinear extension of the Bakry–Emery method [14], by a variational proof of the entropy-dissipation inequality [18], and by virtue of gradient flows in the L^2-Wasserstein metric [36]. These methods have been extended later on to more general f's and Φ's, and also to bounded domains $\Omega \subset \mathbb{R}^d$; see, e.g., [13, 15]. The common fundamental result is that if f satisfies the McCann condition, and if Φ is uniformly convex of modulus $\Lambda > 0$, then solutions u converge to the unique equilibrium in L^1 at exponential rate Λ.

There appears to be no result of comparable simplicity and generality for *coupled systems* of parabolic equations. All of the aforementioned methods of proof break down as soon as multicomponent densities are considered, except in some particular systems with a very special algebraic structure; see, e.g., [30, 45]. There are numerous other applications of gradient flow methods for systems, like for studying the shape or qualitative stability of steady states; see, for instance, [29]. But that approach does not provide equilibration at exponential rates any longer.

Limited generalizations of the scalar theory have been developed for reaction-diffusion systems, and recently also for cross-diffusion systems (with or without reactions). Although many of these systems still bear a gradient flow structure [33], the more robust energy method has proven better adapted to study long-time asymptotics. In reaction-diffusion systems, the substantial challenge is in the control of the growth induced by the reactions, while the diffusion itself is typically decoupled, and frequently just linear. Prototypical results on exponential equilibration have been obtained in [19, 22, 34, 25] for systems with linear diffusion, and in [23] for componentwise nonlinear diffusion. In (reaction-)cross-diffusion systems, the diffusion matrix is nondiagonal, but usually subject to restrictive structural conditions. Recent results on exponential convergence to equilibrium have been obtained, e.g., for systems with volume filling [42], of Maxwell-Stefan type [16], or with SKT-structure [17].

None of the above results covers the exponential equilibration presented in Theorem 1.5 below, i.e., for a system which is fully nonlinear with a general (albeit small) coupling. Nonlinear diffusion on $\mathbb{R}^d$ with inhomogeneous, compactly supported steady states is apparently inaccessible by the commonly used methods, but calls for a different angle of attack, and adapted functional inequalities like (1.5). For comparison, we mention the recent result from [2] that is close in spirit to our approach: the authors treat a system with a small nonlinear coupling like (1.1); however, there, linearity of F and G, a bounded spatial domain, and an a priori L^∞-bound are assumed, which allows one to perform the estimates in a much simpler way, using Poincaré's inequality. We further mention two related results for the parabolic-parabolic Keller–Segel model [44] and the Nernst–Planck system [43]. There, the coupling is between a Wasserstein and an L^2-gradient flow, not between two Wasserstein gradient flows as here. A variant of the above has been explored in [46].

Finally, we briefly comment on the positioning of Theorem 1.4 on existence, which, as mentioned before, merely is an intermediate result on our route to the long-time asymptotics. We use the celebrated JKO scheme [26] to obtain solutions by a variational time-discrete approximation. This scheme has been used for proving the existence of various nonlinear parabolic equations like doubly degenerate parabolic PDEs [37], including the p-Laplace equation [1]; in nonlinear diffusion-aggregation equations [12], including the parabolic-elliptic Keller–Segel model [6], in fourth order quantum and thin film equations [24, 32, 31]; and in many further instances. Applications to coupled systems are numerous as well, including, for instance, systems with nonlocal aggregation [21, 20] and cross-diffusion [5, 10], and also combinations of Wasserstein and L^2-gradient flows, like the parabolic-parabolic Keller–Segel [44, 7] or the Nernst–Planck system [28]. In several of these cases, the existence proof could have also been obtained by more elementary methods. Also for (1.1), the *boundedness-by-entropy method* [27], albeit not directly applicable, would have paved an alternate way. For us, the time-discrete approximation via a minimizing movement with respect to the L^2-Wasserstein distance is crucial for making the long-time asymptotics fully rigorous. A closely related approach to existence for a system similar to (1.1), augmented with additional nonlocal interaction term, has been used in [20]. The hypotheses are complementary to ours, being more flexible on the coupling h, but more restrictive on F and G.

1.3. General hypotheses. Throughout this paper, we work under the following hypotheses. Several of them could be weakened, e.g., C^∞-regularity and normalization of Φ and Ψ are required for convenience only in (1.10), the limit could be replaced by bilateral bounds on $\liminf$ and $\limsup$, etc.

- **Potentials:** For $\Phi, \Psi \in C^\infty(\mathbb{R}^d)$, we assume that:
  - there are positive constants Λ and M such that

$$\Lambda \mathbf{1} \le \nabla^2 \Phi \le M\mathbf{1}, \quad \Lambda \mathbf{1} \le \nabla^2 \Psi \le M\mathbf{1}; \tag{1.6}$$

  - Φ and Ψ vanish at their respective minima $\underline{x}_\Phi, \underline{x}_\Psi \in \mathbb{R}^d$, i.e.,

$$0 = \inf_{\mathbb{R}^d} \Phi = \Phi\big(\underline{x}_\Phi\big), \quad 0 = \inf_{\mathbb{R}^d} \Psi = \Psi\big(\underline{x}_\Psi\big). \tag{1.7}$$

- **Nonlinearities:** We assume $F, G \in C^\infty(\mathbb{R}_{>0}) \cap C^1(\mathbb{R}_{\ge 0})$ such that
  - $F''(r) > 0$ and $G''(r) > 0$ for all $r > 0$, and

$$\liminf_{r\to\infty} F''(r) > 0, \quad \liminf_{r\to\infty} G''(r) > 0; \tag{1.8}$$

  - they degenerate at zero to first order, i.e.,

$$F(0) = G(0) = 0, \quad F'(0) = G'(0) = 0; \tag{1.9}$$

  - there are exponents $m, n \ge 2$ such that

$$\lim_{r\downarrow 0} r^{-(m-2)} F''(r) \in (0,\infty), \quad \lim_{r\downarrow 0} r^{-(n-2)} G''(r) \in (0,\infty); \tag{1.10}$$

  - they satisfy the (dimension-free) McCann condition, i.e., for all $r > 0$,

$$rF'(r) \le F(r) + r^2 F''(r), \quad rG'(r) \le G(r) + r^2 G''(r); \tag{1.11}$$

  - they satisfy the doubling condition, i.e., there is a constant D such that for all $r, s > 0$,

$$F(r+s) \le D(1 + F(r) + F(s)), \quad G(r+s) \le D(1 + G(r) + G(s)). \tag{1.12}$$

- **Coupling:** Concerning $h \in C^\infty(\mathbb{R}^2_{>0}) \cap C^1(\mathbb{R}^2_{\geq 0})$, we assume that
  - h vanishes to first order on $\partial\mathbb{R}^2_{\geq 0}$,

$$h = \partial_u h = \partial_v h \equiv 0 \quad \text{on } \partial\mathbb{R}^2_{\geq 0}; \tag{1.13}$$

  - there is an $\varepsilon^* > 0$ such that

$$(u,v) \mapsto F(u) + G(v) + 2\varepsilon^* h(u,v) \quad \text{is convex}; \tag{1.14}$$

  - with the same ε^*, there holds, for all $u, v > 0$,

$$2\varepsilon^* |h(u,v)| \leq F(u) + G(v). \tag{1.15}$$

- **Degeneracy, boundedness, and swap condition:** Define $\theta_u, \theta_v : \mathbb{R}^2_{\geq 0} \to \mathbb{R}$ by

$$\theta_u(\rho,\eta) := \partial_u h\big((F')^{-1}(\rho), (G')^{-1}(\eta)\big), \quad \theta_u(\rho,\eta) := \partial_u h\big((F')^{-1}(\rho), (G')^{-1}(\eta)\big). \tag{1.16}$$

  We say that the triple (F, G, h) ...
  - ... satisfies the *swap condition* if there is some constant W such that, for all $u, v > 0$,

$$\big|\partial_\eta \theta_u\big(F'(u), G'(v)\big)\big| \leq W\sqrt{v/u}, \quad \big|\partial_\rho \theta_v\big(F'(u), G'(v)\big)\big| \leq W\sqrt{u/v}; \tag{1.17}$$

  - ... is k-*bounded* for some $k \in \mathbb{N}$ if $\theta_u, \theta_v \in C^k(\mathbb{R}^2_{\geq 0})$, and if all partial derivatives of total order $\ell = 1, \dots, k$ are bounded on $\mathbb{R}^2_{\geq 0}$;
  - ... is k-*degenerate* for some $k \in \mathbb{N}$ if $\theta_u, \theta_v \in C^k(\mathbb{R}^2_{\geq 0})$, and if all partial derivatives of total order $\ell = 0, 1, \dots, k$ vanish on $\partial\mathbb{R}^2_{\geq 0}$.

Remark 1.1.

(1) Hypotheses (1.6) and (1.7) imply that Φ and Ψ are bounded from above and from below by parabolas:

$$\frac{\Lambda}{2}|x - \underline{x}_\Phi|^2 \leq \Phi(x) \leq \frac{M}{2}|x - \underline{x}_\Phi|^2, \quad \frac{\Lambda}{2}|x - \underline{x}_\Psi|^2 \leq \Psi(x) \leq \frac{M}{2}|x - \underline{x}_\Psi|^2. \tag{1.18}$$

These estimates are directly obtained by Taylor expansion about the respective minima. Similarly, one bounds the norm of the gradients and thus obtains in combination with (1.18)

$$\frac{2\Lambda^2}{M}\Phi(x) \leq |\nabla\Phi(x)|^2 \leq \frac{2M^2}{\Lambda}\Phi(x), \quad \frac{2\Lambda^2}{M}\Psi(x) \leq |\nabla\Psi(x)|^2 \leq \frac{2M^2}{\Lambda}\Psi(x). \tag{1.19}$$

(2) A consequence of the hypotheses on F and G is that both are *uniformly* convex on each interval of the form $[r, \infty)$ with $r > 0$. Further, in combination with the doubling condition, it follows that for all $r > 0$,

$$rF'(r) \leq D(1 + 2F(r)), \quad rG'(r) \leq D(1 + 2G(r)). \tag{1.20}$$

Indeed, convexity implies $F(2r) \geq F(r) + rF'(r)$, and (1.20) now follows via (1.12) for $s = r$.

(3) If (F,G,h) is 2-bounded and 2-degenerate, there exists a constant $A\geq 0$ such that

$$|\theta_u(\rho,\eta)|,\,|\theta_v(\rho,\eta)|\leq A\min\{\rho,\eta\},\quad |\omega(\rho,\eta)|\leq A\min\{1,\rho,\eta\}, \tag{1.21}$$

where $\omega:\mathbb{R}^2_{\geq 0}\to\mathbb{R}$ is any of the functions $\partial_\rho\theta_u$, $\partial_\eta\theta_u$, $\partial_\rho\theta_v$, or $\partial_\eta\theta_v$.

Example 1.2. Consider F, G, and h of the form

$$F(u)=\frac{u^m}{m},\ G(v)=\frac{v^n}{n},\ h(u,v)=u^p v^q\tilde{u}^\alpha\tilde{v}^\beta\quad\text{with}\quad \tilde{u}:=\frac{u}{1+u+v},\ \tilde{v}:=\frac{v}{1+u+v}$$

for nonnegative exponents m,n,p,q, and α,β. We claim that F, G and h satisfy their respective hypotheses (1.8)–(1.15) plus the swap condition (1.17), provided that

$$m,n\geq 2,\quad p+q\leq\min\{m,n\},\quad \alpha\geq m-p,\ \beta\geq n-q. \tag{1.22}$$

Moreover, given $k\in\mathbb{N}$, we claim that (F,G,h) is k-bounded and k-degenerate, if, additionally,

$$\alpha>k-\frac{p-1}{m-1},\quad \beta>k-\frac{q-1}{n-1}. \tag{1.23}$$

An admissible choice with a 2-bounded and 2-degenerate h is given, in particular, by $m=n=2$, $p=q=1$, and $\alpha=\beta=3$.

Note that, independently of the choice of $\alpha,\beta>0$, the function h behaves for large and comparable values of u and v very similar to u^pv^q. For small values of u, v, and also for ratios u/v that are very small or very large, $h(u,v)$ is significantly "flatter" than u^pv^q; that flatness is needed to guarantee k-boundedness, k-degeneracy, and the swap condition.

The verification of these claims is deferred to Appendix A.

1.4. Results. Our first result concerns the existence of stationary solutions to (1.1), characterized as minimizers of $\mathbf{E}_\varepsilon$ in the space $[L^2(\mathbb{R}^d)]^2\cap[\mathcal{P}_2^r(\mathbb{R}^d)]^2$ (see section 2 for the notation).

THEOREM 1.3. *For each $\varepsilon\in[0,\varepsilon^*]$, there is a unique minimizer $(\bar{u}_\varepsilon,\bar{v}_\varepsilon)$ of $\mathbf{E}_\varepsilon$ in $[L^2(\mathbb{R}^d)]^2\cap[\mathcal{P}_2^r(\mathbb{R}^d)]^2$. The densities $\bar{u}_\varepsilon$ and $\bar{v}_\varepsilon$ are continuous functions of compact support that are sublevels of Φ and Ψ, respectively, and satisfy, for suitable constants U_ε, $V_\varepsilon>0$,*

$$\begin{aligned} F'(\bar{u}_\varepsilon)+\varepsilon\partial_u h(\bar{u}_\varepsilon,\bar{v}_\varepsilon)&=(U_\varepsilon-\Phi)_+\,,\\ G'(\bar{v}_\varepsilon)+\varepsilon\partial_v h(\bar{u}_\varepsilon,\bar{v}_\varepsilon)&=(V_\varepsilon-\Psi)_+\,. \end{aligned} \tag{1.24}$$

Further, if h degenerates to order $k\in\mathbb{N}$, then the restrictions of $F'(\bar{u}_\varepsilon)$ and $G'(\bar{v}_\varepsilon)$ to their respective supports are bounded in C^k, uniformly with respect to $\varepsilon\in[0,\varepsilon^*]$.

A point of crucial importance is that for degeneracy of order two, the functions $\partial_u h(\bar{u}_\varepsilon,\bar{v}_\varepsilon)$ and $\partial_v h(\bar{u}_\varepsilon,\bar{v}_\varepsilon)$ are in $C^2(\mathbb{R}^d)$, with a global bound on second derivatives that is independent of $\varepsilon\in[0,\varepsilon^*]$. This is needed to establish the functional inequalities (1.5), which are essential for our proofs of the following results. Under an additional smallness assumption on ε (see (4.1)), we obtain the following result concerning existence of transient solutions.

THEOREM 1.4. *Assume, in addition, that* (F,G,h) *is* 2*-bounded and* 2*-degenerate, and that the swap condition holds. There is some* $\bar\varepsilon > 0$*, such that for each* $\varepsilon \in [0,\bar\varepsilon]$ *and any initial data* $(u_0,v_0) \in [\mathcal{P}_2^r(\mathbb{R}^d)]^2$ *of finite energy* $\mathbf{E}_\varepsilon(u_0,v_0) < \infty$*, there exists a transient weak solution* $(u_\varepsilon(t),v_\varepsilon(t))_{t\geq 0}$ *to the initial value problem for* (1.1)*, i.e., for arbitrary test functions* $\xi \in C_c^\infty((0,\infty)\times\mathbb{R}^d)$ *there holds*

$$
\begin{aligned}
0 &= \int_0^\infty \int_{\mathbb{R}^d} \big(u_\varepsilon \partial_t \xi - u_\varepsilon \nabla\big[F'(u_\varepsilon) + \Phi + \varepsilon\partial_u h(u_\varepsilon,v_\varepsilon)\big]\cdot\nabla\xi\big)\,\mathrm{d}x\,\mathrm{d}t,\\
0 &= \int_0^\infty \int_{\mathbb{R}^d} \big(v_\varepsilon \partial_t \xi - v_\varepsilon \nabla\big[G'(v_\varepsilon) + \Psi + \varepsilon\partial_v h(u_\varepsilon,v_\varepsilon)\big]\cdot\nabla\xi\big)\,\mathrm{d}x\,\mathrm{d}t.
\end{aligned} \tag{1.25}
$$

The initial data are attained in the L^2*-Wasserstein sense, i.e., as* $t\downarrow 0$*, one has weak-*$*$*-convergence of* $u_\varepsilon(t)$ *to* u_0*, and convergence of* $u_\varepsilon(t)$*'s second moment to that of* v_0*, and similarly for* v_ε*.*

Here the most significant point is not the mere existence but the way of construction, namely via the minimizing movement scheme for $\mathbf{E}_\varepsilon$ (starting from the given initial data) in the combined L^2-Wasserstein distances.

Finally, the main result of this paper is about the long-time asymptotics of transient solutions.

THEOREM 1.5. *Under the same conditions as in Theorem* 1.4 *above, there exist constants* $K>0$ *and* $C\geq 1$ *such that the following is true for all* $\varepsilon\in[0,\bar\varepsilon]$*: the transient solution* $(u_\varepsilon(t),v_\varepsilon(t))_{t\geq 0}$ *constructed in the proof of Theorem* 1.4 *converges to the unique global minimizer* $(\bar u_\varepsilon,\bar v_\varepsilon)$ *from Theorem* 1.3 *at exponential rate* $\Lambda_\varepsilon = \Lambda - K\varepsilon$*. More precisely, with* $\mathbf{L}_\varepsilon$ *being the Lyapunov functional defined in* (1.3)*, there holds*

$$
\mathbf{L}_\varepsilon\big(u_\varepsilon(t),v_\varepsilon(t)\big) \leq \mathbf{L}_\varepsilon(u_0,v_0)\exp\big(-2\Lambda_\varepsilon t\big), \tag{1.26}
$$

and in particular, $u_\varepsilon(t)$ *and* $v_\varepsilon(t)$ *converge in* $L^1(\mathbb{R}^d)$ *to* $\bar u_\varepsilon$ *and* $\bar v_\varepsilon$*, respectively, with*

$$
\|u_\varepsilon(t)-\bar u_\varepsilon\|_{L^1}^2 + \|v_\varepsilon(t)-\bar v_\varepsilon\|_{L^1}^2 \leq C\big(1+\mathbf{E}_\varepsilon(u_0,v_0)\big)\exp\big(-2\Lambda_\varepsilon t\big). \tag{1.27}
$$

2. Preliminaries.

2.1. Wasserstein distance. $\mathbb{B}_R := \{x\in\mathbb{R}^d : |x| < R\}$ is the ball of radius $R>0$. $\mathcal{L}^d$ denotes the standard Lebesgue measure on $\mathbb{R}^d$. For a probability measure μ on $\mathbb{R}^d$ and a measurable map $T:\mathbb{R}^d\to\mathbb{R}^d$, the *push-forward* of μ under T is the uniquely determined probability measure $T\#\mu$ such that

$$
\int_{\mathbb{R}^d} \omega(y)\,\mathrm{d}\big(T\#\mu\big)(y) = \int_{\mathbb{R}^d} \omega\circ T(x)\,\mathrm{d}\mu(x) \tag{2.1}
$$

for any test function $\omega\in C(\mathbb{R}^d)$. If both $\mu = u\mathcal{L}^d$ and $T\#\mu = \hat u\mathcal{L}^d$ are absolutely continuous, then we write $T\#u = \hat u$ for brevity.

$\mathcal{P}_2^r(\mathbb{R}^d)$ denotes the space of probability densities $u:\mathbb{R}^d\to\mathbb{R}_{\geq 0}$ of finite second moment. The natural notion of convergence on $\mathcal{P}_2^r(\mathbb{R}^d)$ is the *narrow* one, i.e., weak convergence in duality with bounded continuous functions. By Prokhorov's and by Alaoglu's theorem, subsets of densities with uniformly bounded second moment and L^p-norm (for some $p>1$) are sequentially compact in $\mathcal{P}_2^r(\mathbb{R}^d)$; boundedness of the L^p-norm is just needed to avoid concentrations.

The L^2-Wasserstein distance $\mathbf{W}_2$ is a metric on $\mathcal{P}_2^r(\mathbb{R}^d)$, see [40, Chapters 1 and 5] or [41, Chapters 1, 2, and 7] for an introduction. Convergence in $\mathbf{W}_2$ is equivalent to weak convergence and convergence of the second moments. Among the various possible definitions of $\mathbf{W}_2$ the following—known as the (pre-)dual Kantorovich formulation—is the most suitable one for our needs: for $u, \hat{u} \in \mathcal{P}_2^r(\mathbb{R}^d)$,

$$\frac{1}{2}\mathbf{W}_2(u,\hat{u})^2 := \sup\left\{\int_{\mathbb{R}^d}\varphi(x)u(x)\,\mathrm{d}x + \int_{\mathbb{R}^d}\psi(y)\hat{u}(y)\,\mathrm{d}y : \varphi(x)+\psi(y) \le \frac{1}{2}|x-y|^2\right\}. \tag{2.2}$$

(Note the square and the factor $1/2$ on the left-hand side.) A priori, the maximization above is carried out over all $\varphi \in L^1(\mathbb{R}^d; u\mathcal{L}^d)$ and $\psi \in L^1(\mathbb{R}^d; \hat{u}\mathcal{L}^d)$. However, it suffices to consider pairs (φ,ψ) from the class of *c-conjugate*[2] *potentials*. The latter means that the *auxiliary potentials* $\tilde{\varphi}, \tilde{\psi} : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ given by

$$\tilde{\varphi}(x) = \frac{1}{2}|x|^2 - \varphi(x), \quad \tilde{\psi}(y) = \frac{1}{2}|y|^2 - \psi(y)$$

are proper, lower semicontinuous, convex, and Legendre-dual to each other, $\tilde{\varphi}^* = \tilde{\psi}$ and $\tilde{\psi}^* = \tilde{\varphi}$. Note that knowledge of either φ or ψ determines the respective other. Further, note that $\varphi(x)+\psi(y) \le \frac{1}{2}|x-y|^2$ is automatically satisfied since $\tilde{\varphi}(x)+\tilde{\psi}(y) \ge x \cdot y$.

The supremum in (2.2) is attained by an optimal pair $(\varphi_{\mathrm{opt}}, \psi_{\mathrm{opt}})$ of c-conjugate potentials. Uniqueness of optimal pairs—beyond the global gauge invariance $(\varphi,\psi) \rightsquigarrow (\varphi + C, \psi - C)$—is delicate, in general. Fortunately, in the setting of absolutely continuous measures under consideration, Brenier's theorem (see [40, Theorem 1.22] or [41, Theorem 2.12]), implies that $\nabla\tilde{\varphi}_{\mathrm{opt}}$ is unique $u\mathcal{L}^d$-a.e., and that $\nabla\tilde{\psi}_{\mathrm{opt}}$ is unique $\hat{u}$-a.e. In section 4, the following consequence of this will be important if $\hat{u}$ has a uniform positive lower bound $\mathcal{L}^d$-a.e. on a ball $\mathbb{B}_R \subset \mathbb{R}^d$, and is zero outside, then the proper convex lower-semicontinuous function $\tilde{\psi}_{\mathrm{opt}}$ is unique on $\mathbb{B}_R$ up to a global constant and its Legendre-dual $\tilde{\varphi}_{\mathrm{opt}}$ is then unique up to a global constant on $\mathbb{B}_R$'s image under $\nabla\tilde{\psi}_{\mathrm{opt}}$, which is convex and has full $u\mathcal{L}^d$-measure.

For an optimal pair $(\varphi_{\mathrm{opt}}, \psi_{\mathrm{opt}})$, the *optimal transport map* $T : \mathbb{R}^d \to \mathbb{R}^d$ from u to $\hat{u}$ is given by

$$T(x) := x - \nabla\varphi_{\mathrm{opt}}(x), \tag{2.3}$$

which is well-defined $u\mathcal{L}^d$-a.e. It satisfies

$$\hat{u} = T\#u \tag{2.4}$$

and

$$\mathbf{W}_2(u,\hat{u})^2 = \int_{\mathbb{R}^d} |T(x) - x|^2 u(x)\,\mathrm{d}x. \tag{2.5}$$

By Brenier's theorem, T is $u\mathcal{L}^d$-a.e. unique for a given pair $(u,\hat{u})$, which implies the following converse: if $T = \nabla\tilde{\varphi} : \mathbb{R}^d \to \mathbb{R}^d$ is the $u\mathcal{L}^d$-a.e. defined gradient of a proper, lower semi-continuous, and convex function $\tilde{\varphi} : \mathbb{R}^d \to \mathbb{R}$, and satisfies (2.4), then T also satisfies (2.5), and $\varphi_{\mathrm{opt}}(x) := \frac{1}{2}|x|^2 - \tilde{\varphi}(x)$ gives rise to an optimal pair $(\varphi_{\mathrm{opt}}, \psi_{\mathrm{opt}})$ of c-conjugate potentials.

Finally, we recall a characterization of geodesics: define the interpolating maps $T_s : \mathbb{R}^d \to \mathbb{R}^d$ for all $s \in [0,1]$ by $T_s(x) := (1-s)x + sT(x) = x - s\nabla\varphi_u(x)$. Then the

[2]The c refers to the cost function, which is the standard one here, $c(x,y) = \frac{1}{2}|x-y|^2$.

curve $(u_s)_{s\in[0,1]}$ in $\mathcal{P}_2^r(\mathbb{R}^d)$ given by $u_s := T_s\#\rho$ is a geodesic joining $u=u_0$ to $\hat{u}=u_1$, that is,

$$\mathbf{W}_2(u,u_s)=s\mathbf{W}_2(u,\hat{u}),\quad \mathbf{W}_2(u_s,\hat{u})=(1-s)\mathbf{W}_2(u,\hat{u}).$$

The natural space for solutions (u,v) to (1.1) is the cross product $[\mathcal{P}_2^r(\mathbb{R}^d)]^2$. We endow it with a metric $\widetilde{\mathbf{W}}_2$ in the straightforward way:

$$\widetilde{\mathbf{W}}_2\big((u,v),(\hat{u},\hat{v})\big):=\sqrt{\mathbf{W}_2(u,\hat{u})^2+\mathbf{W}_2(v,\hat{v})^2}.$$

The following is easily seen.

LEMMA 2.1. *A curve $(u_s,v_s)_{0\le s\le 1}$ in $[\mathcal{P}_2^r(\mathbb{R}^d)]^2$ is a geodesic in $[\mathcal{P}_2^r(\mathbb{R}^d)]^2$ between (u_0,v_0) and (u_1,v_1) if and only if $(u_s)_{s\in[0,1]}$ and $(v_s)_{s\in[0,1]}$ are geodesics in $\mathcal{P}_2^r(\mathbb{R}^d)$ between u_0, u_1, and between v_0, v_1, respectively.*

2.2. Displacement convexity. See [40, Chapter 7] and [41, Chapter 5] for an introduction.

DEFINITION 2.2. *A functional $\mathbf{F}$ on $\mathcal{P}_2^r(\mathbb{R}^d)$ is λ-uniformly displacement convex with some modulus $\lambda\in\mathbb{R}$ if the real function*

$$[0,1]\ni s\mapsto \mathbf{F}\big(T_s\#u\big)-\frac{\lambda}{2}s(1-s)\mathbf{W}_2(u,\hat{u})^2$$

is convex for any family $(T_s)_{s\in[0,1]}$ realizing the geodesic between u and $\hat{u}=T_1\#u$.

Displacement convex functionals are rare. An important class of examples is given by the sum of internal and potential energy:

$$\mathbf{F}(u)=\int_{\mathbb{R}^d}\big[e(u)+Vu\big]\,\mathrm{d}x. \tag{2.6}$$

In this case $\mathbf{F}$ is λ-uniformly displacement convex provided that the convex function $e:\mathbb{R}_{\ge 0}\to\mathbb{R}$ satisfies McCann's condition, and that $V:\mathbb{R}^d\to\mathbb{R}$ is λ-convex in the usual sense. A consequence of that property is the validity of a functional inequality; see, e.g., [15, Theorem 2.1].

LEMMA 2.3. *Assume that the functional $\mathbf{F}$ of the type* (2.6) *is such that the convex function $e:\mathbb{R}_{\ge 0}\to\mathbb{R}$ satisfies McCann's condition, and such that $V:\mathbb{R}^d\to\mathbb{R}$ is λ-convex for $\lambda>0$. Then $\mathbf{F}$ possesses a unique minimizer $u_*\in\mathcal{P}_2^r(\mathbb{R}^d)$, and for all $u\in\mathcal{P}_2^r(\mathbb{R}^d)$, there holds*

$$2\lambda\big[\mathbf{F}(u)-\mathbf{F}(u_*)\big]\le\int_{\mathbb{R}^d}u\big|\nabla\big[e'(u)+V\big]\big|\,\mathrm{d}x. \tag{2.7}$$

Functionals of the type (2.6) with $\lambda\ge 0$ actually even enjoy the stronger property of being convex along generalized geodesics, which has a variety of consequences. The only consequence needed below is for the special case $h(u)=u\log u$ and $V\equiv 0$, when $\mathbf{F}=\mathbf{H}$ is the entropy functional,

$$\mathbf{H}(u)=\int_{\mathbb{R}^d}u\log u\,\mathrm{d}x. \tag{2.8}$$

The metric gradient flow of $\mathbf{H}$ is the heat equation $\partial_s U_s=\Delta U_s$, and (thanks to convexity along generalized geodesics) it satisfies the so-called *evolution variational inequality* (EVI_0) (see [3, Theorem 4.0.4]), which is

$$\frac{1}{2}\frac{\mathrm{d}^+}{\mathrm{d}s}\bigg|_{s=0^+}\mathbf{W}_2\big(U_s,w\big)^2\le\mathbf{H}(w)-\mathbf{H}(U_0) \tag{2.9}$$

for all $w\in\mathcal{P}_2^r(\mathbb{R}^d)$ and all solutions to $\partial_s U_s=\Delta U_s$.

2.3. Loss of displacement convexity for mixtures. We indicate why the aforementioned general theory of λ-uniformly displacement convex functionals does not apply to the energy functional $\mathbf{E}_\varepsilon$ for proving exponential convergence to equilibrium in (1.1). Specifically, we show that (the two-component analogue of) displacement convexity cannot be expected for a functional of the form $\mathbf{E}_\varepsilon$ on the space $[\mathcal{P}_2^r(\mathbb{R}^d)]^2$.

PROPOSITION 2.4. *Assume that h is not identically zero, and that $\varepsilon > 0$. Then, there is no $\lambda \in \mathbb{R}$ such that $\mathbf{E}_\varepsilon$ is λ-convex along geodesics in $[\mathcal{P}_2^r(\mathbb{R}^d)]^2$. More specifically, for each $\omega \in \mathbb{R}_{>0}$, there are functions $u^\omega, v^\omega \in \mathcal{P}_2^r(\mathbb{R}^d) \cap C^\infty(\mathbb{R}^d)$ such that*

$$\frac{\mathrm{d}^2}{\mathrm{d}s^2}\bigg|_{s=0} \int_{\mathbb{R}^d} H_\varepsilon(T_s \# u^\omega, v^\omega)\,\mathrm{d}x \le -C(\omega - 1), \tag{2.10}$$

where C is a positive constant, T_s is the translation by $s \ge 0$ in x_1-direction, i.e., $T_s(x) = x + s\mathbf{e}_1$, and H_ε is the function defined in (3.1).

Note that, by Lemma 2.1, the curve $(T_s \# u^\omega, v^\omega)_{0\le s\le 1}$ is a geodesic in $[\mathcal{P}_2^r(\mathbb{R}^d)]^2$.

Remark 2.5. With a little technical effort, the construction in the proof below can be used to show that such pairs (u_ω, v_ω) are actually dense in $[\mathcal{P}_2^r(\mathbb{R}^d)]^2$. For the sake of clarity, we only give the construction for one such pair.

Proof. We first notice that there exists some $(U,V) \in \mathbb{R}^2_{>0}$ with $\partial_{uv} h(U,V) \neq 0$, as h is not identically zero by assumption and satisfies $h \equiv 0$ on $\partial\mathbb{R}^2_{\ge 0}$ in view of the degeneracy condition (1.13). For the construction below, we assume $\partial_{uv} h(U,V) > 0$, and we comment on the other case at the end of the proof. Choose $u^0, v^0 \in \mathcal{P}_2^r(\mathbb{R}^d) \cap C_c^\infty(\mathbb{R}^d)$ such that $u^0(x) = U$ and $v^0(x) = V$ for all $|x| < r$, with some sufficiently small $r > 0$. For all sufficiently large $\omega > 0$, define $u^\omega, v^\omega \in \mathcal{P}_2^r(\mathbb{R}^d) \cap C_c^\infty(\mathbb{R}^d)$ by

$$u^\omega(x) = u^0(x) + \omega^{-1/2}\delta_r(x)\sin(\omega x_1), \quad v^\omega(x) = v^0(x) + \omega^{-1/2}\delta_r(x)\sin(\omega x_1), \tag{2.11}$$

where $\delta_r \in C_c^\infty(\mathbb{R}^d)$ is radially symmetric about the origin, with $\delta_r(x) = 1$ for $|x| < r/2$ and $\delta_r(x) = 0$ for $|x| > r$. For the integral in (2.10), we obtain via an integration by parts:

$$\begin{aligned}
&\frac{\mathrm{d}^2}{\mathrm{d}s^2}\bigg|_{s=0} \int_{\mathbb{R}^d} H_\varepsilon(T_s \# u^\omega, v^\omega)\,\mathrm{d}x \\
&\quad = \int_{\mathbb{R}^d} \left[\partial_u H_\varepsilon(u^\omega, v^\omega)\partial_{x_1x_1} u^\omega + \partial_{uu} H_\varepsilon(u^\omega, v^\omega)\big(\partial_{x_1} u^\omega\big)^2\right]\mathrm{d}x \\
&\quad = -\int_{\mathbb{R}^d} \partial_{uv} H_\varepsilon(u^\omega, v^\omega)\partial_{x_1} u^\omega \partial_{x_1} v^\omega\,\mathrm{d}x.
\end{aligned}$$

By construction, the contribution of this integral over $|x| < r$ is roughly proportional to ω, with a negative sign since $\partial_{uv} H_\varepsilon(U,V) > 0$. The contribution on $|x| > r$ has some finite value, independent of ω.

Now if $\partial_{uv} h(U,V)$ is negative instead of positive, we only change the definition of v^ω in (2.11) above into

$$v^\omega(x) = v^0(x) - \omega^{-1/2}\delta_r(x)\sin(\omega x_1),$$

which makes the product $\partial_{x_1} u^\omega \partial_{x_1} v^\omega$ negative instead of positive for $|x| < r/2$. ☐

3. Stationary solutions. In this section, Theorem 1.3 is proven. It is a consequence of the (more detailed) results stated in Propositions 3.2 and 3.5 below. For brevity, define $H_\varepsilon : \mathbb{R}^2_{\geq 0} \to \mathbb{R}$ by

$$H_\varepsilon(u,v) := F(u) + G(v) + \varepsilon h(u,v), \tag{3.1}$$

which allows one to write

$$\mathbf{E}_\varepsilon(u,v) = \int_{\mathbb{R}^d} \big[H_\varepsilon(u,v) + u\Phi + v\Psi \big]\, \mathrm{d}x.$$

Remark 3.1. We notice some important properties of the function H_ε and the energy $\mathbf{E}_\varepsilon$ for $\varepsilon \in [0, \varepsilon^*]$, which are used in this section.

(1) By hypothesis (1.14), H_ε is nonnegative, strictly convex., and satisfies the explicit convexity estimate

$$\mathrm{D}^2 H_\varepsilon(u,v) \geq \frac{1}{2}\begin{pmatrix} F''(u) & 0 \\ 0 & G''(v) \end{pmatrix} \quad \text{for all } (u,v) \in \mathbb{R}^2_{>0}. \tag{3.2}$$

(2) As a consequence of the strict convexity of H_ε, its differential $\mathrm{D}H_\varepsilon$ is a strictly monotone continuous map on the cone $\mathbb{R}^2_{\geq 0}$, i.e., it satisfies $(\mathrm{D}H_\varepsilon(u,v) - \mathrm{D}H_\varepsilon(\tilde{u},\tilde{v})) \cdot (u - \tilde{u}, v - \tilde{v}) > 0$ for all $(u,v), (\tilde{u},\tilde{v}) \in \mathbb{R}^2_{\geq 0}$ with $(u,v) \neq (\tilde{u},\tilde{v})$. In view of the identities $\mathrm{D}H_\varepsilon(u,0) = (F'(u), 0)$ and $\mathrm{D}H_\varepsilon(0,v) = (0, G'(v))$, and since F' and G' are monotone and unbounded with $F'(0) = G'(0) = 0$, the image of $\mathrm{D}H_\varepsilon$ is $\mathbb{R}^2_{\geq 0}$. Hence, $\mathrm{D}H_\varepsilon$ is a homeomorphism of $\mathbb{R}^2_{\geq 0}$ onto itself, and also a homeomorphism of $\mathbb{R}^2_{>0}$ onto itself.

(3) The qualified convexity (3.2), combined with the at least quadratic growth of F and G, and the Λ-convexity of Φ, Ψ imply

$$\mathbf{E}_\varepsilon(u,v) \geq c \int_{\mathbb{R}^d} (u^2 + v^2)\, \mathrm{d}x + \frac{\Lambda}{2} \int_{\mathbb{R}^d} |x|^2 (u+v)\, \mathrm{d}x - C$$

for all $(u,v) \in [L^2(\mathbb{R}^d)]^2 \cap [\mathcal{P}_2^r(\mathbb{R}^d)]^2$, with some constants C and $c > 0$.

In the following, $\bar{U} \geq 2$ denotes the smallest number such that, with M from (1.6), there hold

$$\begin{aligned} \frac{1}{2} F'(\bar{U}) &\geq F'(2) + dM + \varepsilon^* \sup_{u,v \leq 2} \partial_u h(u,v) + 1, \\ \frac{1}{2} G'(\bar{U}) &\geq G'(2) + dM + \varepsilon^* \sup_{u,v \leq 2} \partial_v h(u,v) + 1. \end{aligned} \tag{3.3}$$

PROPOSITION 3.2. *Let $\varepsilon \in [0, \varepsilon^*]$. There exists a unique global minimizer $(\bar{u}_\varepsilon, \bar{v}_\varepsilon)$ of $\mathbf{E}_\varepsilon$ in $[L^2(\mathbb{R}^d)]^2 \cap [\mathcal{P}_2^r(\mathbb{R}^d)]^2$. The components $\bar{u}_\varepsilon$, $\bar{v}_\varepsilon$ are continuous functions of compact support, bounded by $\bar{U}$ from* (3.3)*. Moreover, there are constants $U_\varepsilon, V_\varepsilon > 0$ such that $\bar{u}_\varepsilon, \bar{v}_\varepsilon$ satisfy the Euler–Lagrange equations in* (1.24)*. The supports of $\bar{u}_\varepsilon$ and $\bar{v}_\varepsilon$ are convex and given by the closures of the sublevel sets*

$$\Omega^u_\varepsilon := \{\Phi < U_\varepsilon\}, \qquad \Omega^v_\varepsilon := \{\Psi < V_\varepsilon\}, \tag{3.4}$$

respectively. Finally, there is an upper bound on U_ε and V_ε, and also on the diameters of Ω^u_ε and Ω^v_ε, uniformly for $0 \leq \varepsilon \leq \varepsilon^*$.

Remark 3.3. Thanks to hypothesis (1.13), we have $\partial_u h(u,0) = \partial_v h(0,v) = 0$, and thus the system (1.24) can be made a bit more explicit: On $\Omega_\varepsilon^u \setminus \Omega_\varepsilon^v$, one has $\bar{u}_\varepsilon = (F')^{-1}(U_\varepsilon - \Phi)$, on $\Omega_\varepsilon^v \setminus \Omega_\varepsilon^u$, one has $\bar{v}_\varepsilon = (G')^{-1}(V_\varepsilon - \Psi)$. Finally, on $\Omega_\varepsilon^u \cap \Omega_\varepsilon^v$, the values of $\bar{u}_\varepsilon$ and $\bar{v}_\varepsilon$ are obtained as pointwise solution of (1.24), with right-hand sides $U_\varepsilon - \Phi > 0$ and $V_\varepsilon - \Psi > 0$.

Remark 3.4. The explicit representation (3.4) of the supports is related to hypothesis (1.13), specifically to

$$\partial_u h(0,v) = \partial_v h(v,0) = 0. \tag{3.5}$$

If just the part (3.5) of our set of hypotheses was removed, then the conclusions of Proposition 3.2 are essentially still valid, but the supports of $\bar{u}_\varepsilon$ and $\bar{v}_\varepsilon$ are only *subsets* of the respective sublevel sets of Φ and Ψ, in general.

For an illustration of this situation, consider the choices $F(u) = \frac{u^2}{2}$, $G(v) = \frac{v^2}{2}$, and $h(u,v) = uv$, for which (3.5) is false. Proceeding as in the proof of Proposition 3.2 below, one obtains existence and uniqueness of a minimizer with densities $(\bar{u}_\varepsilon, \bar{v}_\varepsilon)$, and (1.24) turns into a *linear* system for the values of $\bar{u}_\varepsilon$ and $\bar{v}_\varepsilon$ on the intersection of their supports. Thereon, the explicit solution is given by

$$(1-\varepsilon^2)\bar{u}_\varepsilon = (U_\varepsilon - \Phi)_+ - \varepsilon(V_\varepsilon - \Psi)_+, \quad (1-\varepsilon^2)\bar{v}_\varepsilon = (V_\varepsilon - \Psi)_+ - \varepsilon(U_\varepsilon - \Phi)_+.$$

From this representation it is clear that if the respective sublevel sets of Φ and Ψ overlap, then the supports of $\bar{u}_\varepsilon$ and $\bar{v}_\varepsilon$ are genuinely smaller.

Proof of Proposition 3.2. By Remark 3.1 (3), the sublevel sets of $\mathbf{E}_\varepsilon$ are compact in $[\mathcal{P}_2^r(\mathbb{R}^d)]^2$, and weakly compact in $[L^2(\mathbb{R}^d)]^2$. Further, strict convexity of H_ε (see Remark 3.1 (1)) and nonnegativity of Φ and Ψ imply strict convexity of $\mathbf{E}_\varepsilon$, as well as its lower semicontinuity with respect to convergence in $[\mathcal{P}_2^r(\mathbb{R}^d)]^2$. Therefore, existence and uniqueness of the global minimizer $(\bar{u}_\varepsilon, \bar{v}_\varepsilon) \in [L^2(\mathbb{R}^d)]^2 \cap [\mathcal{P}_2^r(\mathbb{R}^d)]^2$ follow via the direct method from the calculus of variations.

Next, we verify that $\bar{U}$ is an upper bound by showing that if $\bar{u}_\varepsilon$ or $\bar{v}_\varepsilon$ would exceed $\bar{U}$, then there is a competitor $(\tilde{u}_\varepsilon, \tilde{v}_\varepsilon)$, bounded by $\bar{U}$, of a lower $\mathbf{E}_\varepsilon$-energy. Assume that $\bar{u}_\varepsilon > \bar{U}$ on a set $P \subset \mathbb{R}^d$ of positive Lebesgue measure. If $\bar{u}_\varepsilon \le \bar{U}$ a.e. but $\bar{v}_\varepsilon > \bar{U}$, the argument is analogous. Define

$$\sigma := \int_P (\bar{u}_\varepsilon - \bar{U})\,\mathrm{d}x \in (0,1). \tag{3.6}$$

Consider the cube $Q \subset \mathbb{R}^d$ of volume $V = 3$ (i.e., of side length $3^{1/d}$), centered around the minimum point $\underline{x}_\Phi$ of Φ. Since $\bar{u}_\varepsilon$ and $\bar{v}_\varepsilon$ are of unit mass, the subsets of Q on which $\bar{u}_\varepsilon \ge 1$ or $\bar{v}_\varepsilon \ge 1$, respectively, are of measure at most one. Hence, there is a set $S \subset Q$ of unit Lebesgue measure on which $\bar{u}_\varepsilon \le 1$ and $\bar{v}_\varepsilon \le 1$. Since $\bar{U} \ge 2$, the sets P and S are disjoint. We define $\tilde{u}_\varepsilon$ as a modification of $\bar{u}_\varepsilon$ as follows: we set $\tilde{u}_\varepsilon := \bar{U}$ on P, we set $\tilde{u}_\varepsilon := \bar{u}_\varepsilon + \sigma$ on S, and we set $\tilde{u}_\varepsilon := \bar{u}_\varepsilon$ otherwise. By definition of σ, and since S is of unit measure, $\tilde{u}_\varepsilon$ is a probability density, and thus $(\tilde{u}_\varepsilon, \bar{v}_\varepsilon)$ is an admissible competitor. On the one hand, a.e. on P, where $\tilde{u}_\varepsilon = \bar{U} \le \bar{u}_\varepsilon$,

$$H_\varepsilon(\bar{u}_\varepsilon, \bar{v}_\varepsilon) - H_\varepsilon(\tilde{u}_\varepsilon, \bar{v}_\varepsilon) \ge (\bar{u}_\varepsilon - \tilde{u}_\varepsilon)\partial_u H_\varepsilon(\tilde{u}_\varepsilon, \bar{v}_\varepsilon) \ge \frac{1}{2}(\bar{u}_\varepsilon - \bar{U})F'(\bar{U}),$$

using the convexity estimate (3.2), the degeneracy (1.9) of F, and (1.13) of h. Hence, recalling the definition (3.6) of σ and the nonnegativity of Φ,

$$\int_P \big(H_\varepsilon(\bar{u}_\varepsilon, \bar{v}_\varepsilon) + \bar{u}_\varepsilon\Phi + \bar{v}_\varepsilon\Psi\big)\,\mathrm{d}x \ge \int_P \big(H_\varepsilon(\tilde{u}_\varepsilon, \bar{v}_\varepsilon) + \tilde{u}_\varepsilon\Phi + \bar{v}_\varepsilon\Psi\big)\,\mathrm{d}x + \frac{1}{2}F'(\bar{U})\sigma.$$

On the other hand, a.e. on S, where $\tilde{u}_\varepsilon = \bar{u}_\varepsilon + \sigma \le 2$,

$$H_\varepsilon(\bar{u}_\varepsilon, \bar{v}_\varepsilon) - H_\varepsilon(\tilde{u}_\varepsilon, \bar{v}_\varepsilon) \ge \big(\bar{u}_\varepsilon - \tilde{u}_\varepsilon\big)\partial_u H_\varepsilon(\tilde{u}_\varepsilon, \bar{v}_\varepsilon) \ge -\Big[F'(2) + \varepsilon^* \sup_{a,b\le 2} \partial_u h(a,b)\Big]\sigma.$$

With S being of unit measure, and recalling that $0 \le \Phi(x) \le \frac{M}{2}|x - \underline{x}_\Phi|^2 \le (3/2)^{2/d} dM/2 \le dM$ for all $x \in S$ thanks to (1.18), it follows that

$$\begin{aligned}\int_S \big(H_\varepsilon(\bar{u}_\varepsilon, \bar{v}_\varepsilon) + \bar{u}_\varepsilon\Phi + \bar{v}_\varepsilon\Psi\big)\,\mathrm{d}x \ge{}& \int_S \big(H_\varepsilon(\tilde{u}_\varepsilon, \bar{v}_\varepsilon) + \tilde{u}_\varepsilon\Phi + \bar{v}_\varepsilon\Psi\big)\,\mathrm{d}x \\ &- \Big[dM + F'(2) + \varepsilon^* \sup_{a,b\le 2} \partial_u h(a,b)\Big]\sigma.\end{aligned}$$

In summary, recalling the implicit definition (3.3) of $\bar{U}$, we find

$$\mathbf{E}_\varepsilon(\bar{u}_\varepsilon, \bar{v}_\varepsilon) \ge \mathbf{E}_\varepsilon(\tilde{u}_\varepsilon, \bar{v}_\varepsilon) + \sigma.$$

This contradicts the minimality of $(\bar{u}_\varepsilon, \bar{v}_\varepsilon)$. Consequently, a.e. on $\mathbb{R}^d$ we have $\bar{u}_\varepsilon \le \bar{U}$ and $\bar{v}_\varepsilon \le \bar{U}$.

To characterize the minimizer, we perform variations of the form

$$\bar{u}_\varepsilon^s = (1 - \alpha s)\bar{u}_\varepsilon + s\xi, \qquad \bar{v}_\varepsilon^s = (1 - \beta s)\bar{v}_\varepsilon + s\eta,$$

with appropriate functions $\xi, \eta \in L^\infty(\mathbb{R}^d)$ of compact support, and parameters

$$\alpha = \int_{\mathbb{R}^d} \xi\,\mathrm{d}x, \qquad \beta = \int_{\mathbb{R}^d} \eta\,\mathrm{d}x.$$

We define

$$U_\varepsilon := \int_{\mathbb{R}^d} \big[\partial_u H_\varepsilon(\bar{u}_\varepsilon, \bar{v}_\varepsilon) + \Phi\big]\bar{u}_\varepsilon\,\mathrm{d}x, \quad V_\varepsilon := \int_{\mathbb{R}^d} \big[\partial_v H_\varepsilon(\bar{u}_\varepsilon, \bar{v}_\varepsilon) + \Psi\big]\bar{v}_\varepsilon\,\mathrm{d}x.$$

First, let $\xi, \eta \in C_c(\mathbb{R}^d)$ be nonnegative. Then $\bar{u}_\varepsilon^s$, $\bar{v}_\varepsilon^s \in \mathcal{P}_2^r(\mathbb{R}^d)$ for all $s \ge 0$ sufficiently small, and

$$\begin{aligned}0 &\le \lim_{s\downarrow 0} \frac{\mathbf{E}_\varepsilon(\bar{u}_\varepsilon^s, \bar{v}_\varepsilon^s) - \mathbf{E}_\varepsilon(\bar{u}_\varepsilon, \bar{v}_\varepsilon)}{s} \\ &= \int_{\mathbb{R}^d} \big[\partial_u H_\varepsilon(\bar{u}_\varepsilon, \bar{v}_\varepsilon) + \Phi\big]\xi\,\mathrm{d}x + \int_{\mathbb{R}^d} \big[\partial_v H_\varepsilon(\bar{u}_\varepsilon, \bar{v}_\varepsilon) + \Psi\big]\eta\,\mathrm{d}x - \alpha U_\varepsilon - \beta V_\varepsilon.\end{aligned}$$

This shows that, a.e. on $\mathbb{R}^d$,

$$\begin{aligned}F'(\bar{u}_\varepsilon) + \varepsilon\partial_u h(\bar{u}_\varepsilon, \bar{v}_\varepsilon) + \Phi &\ge U_\varepsilon, \\ G'(\bar{v}_\varepsilon) + \varepsilon\partial_v h(\bar{u}_\varepsilon, \bar{v}_\varepsilon) + \Psi &\ge V_\varepsilon.\end{aligned} \tag{3.7}$$

Consequently, with $\partial_u h(0, v) = 0$ for all $v \ge 0$ by hypothesis (1.13) and the fact that F is degenerate at zero to first order by hypothesis (1.9), we necessarily have $\bar{u}_\varepsilon > 0$ a.e. on $\{U_\varepsilon > \Phi\}$, and similarly $\bar{v}_\varepsilon > 0$ a.e. on $\{V_\varepsilon > \Psi\}$.

Next, let $\xi, \eta \in C_c(\mathbb{R}^d)$ be arbitrary. For any $\delta > 0$, we may perform the aforementioned variations with

$$\xi_\delta := \begin{cases}\xi & \text{where } \bar{u}_\varepsilon > \delta, \\ 0 & \text{otherwise,}\end{cases} \qquad \eta_\delta := \begin{cases}\eta & \text{where } \bar{v}_\varepsilon > \delta, \\ 0 & \text{otherwise}\end{cases}$$

even for all $-s \geq 0$ sufficiently small, and can thus conclude the opposite inequalities in (3.7). This means that equality holds in the first inequality in (3.7) a.e. on $\{\bar{u}_\varepsilon > 0\}$, and in the second inequality a.e. on $\{\bar{v}_\varepsilon > 0\}$.

We next consider the case that $\bar{u}_\varepsilon > 0$ and $\bar{v}_\varepsilon = 0$. Then $F'(\bar{u}_\varepsilon) > 0$, and since $\partial_u h(\bar{u}_\varepsilon, \bar{v}_\varepsilon) = 0$ by hypothesis (1.13), it follows that $\Phi < U_\varepsilon$. Analogously, $\bar{v}_\varepsilon > 0$ and $\bar{u}_\varepsilon = 0$ implies $\Psi < V_\varepsilon$. Finally, suppose $\bar{u}_\varepsilon > 0$ and $\bar{v}_\varepsilon > 0$, so that (3.7) becomes a system of two equations. Recalling the definition of H_ε above, that system can be written as a single vectorial equation as follows:

$$\mathrm{D}H_\varepsilon(\bar{u}_\varepsilon, \bar{v}_\varepsilon) = \begin{pmatrix} U_\varepsilon - \Phi \\ V_\varepsilon - \Psi \end{pmatrix}.$$

By Remark 3.1 (2), the positive cone $\mathbb{R}^2_{>0}$ is mapped into itself under $\mathrm{D}H_\varepsilon$. Therefore, $\bar{u}_\varepsilon > 0$ and $\bar{v}_\varepsilon > 0$ implies that $\Phi < U_\varepsilon$ and $\Psi < U_\varepsilon$, respectively. Consequently, the positivity sets of $\bar{u}_\varepsilon$ and $\bar{v}_\varepsilon$ are indeed given by the sublevel sets Ω^u_ε and Ω^v_ε from (3.4), respectively.

To sum up: a.e. on $\{\Phi < U_\varepsilon\}$, we have $\bar{u}_\varepsilon > 0$ and equality in the first inequality of (3.7), and a.e. on the complement $\{\Phi \geq U_\varepsilon\}$, we have $u_\varepsilon = 0$, which, thanks to $F'(0) = 0$ and $\partial_u h(0, \bar{v}_\varepsilon) = 0$, can be written as

$$F'(\bar{u}_\varepsilon) + \varepsilon \partial_u h(\bar{u}_\varepsilon, \bar{v}_\varepsilon) = 0.$$

This (and an analogous argument for $\bar{v}_\varepsilon$) implies (1.24).

Next, concerning the uniform boundedness of U_ε and V_ε, it suffices to observe that, in view of $\bar{u}_\varepsilon, \bar{v}_\varepsilon \leq \bar{U}$,

$$\partial_u H_\varepsilon\big(\bar{u}_\varepsilon, \bar{v}_\varepsilon\big) \leq C := F'(\bar{U}) + \varepsilon^* \sup_{a,b \leq \bar{U}} h(a,b),$$

and further that at $x = \underline{x}_\Phi$,

$$U_\varepsilon = U_\varepsilon - \Phi(\underline{x}_\Phi) = \partial_u H_\varepsilon\big(\bar{u}_\varepsilon(\underline{x}_\Phi), \bar{v}_\varepsilon(\underline{x}_\Phi)\big).$$

In combination, this yields $U_\varepsilon \leq C$, and we can argue analogously for V_ε. We further notice that the uniform boundedness of the sets Ω^u_ε and Ω^v_ε is a direct consequence of the estimates in (1.18) and the uniform bound on the constants U_ε and V_ε.

Finally, we verify that the L^2-representatives of $\bar{u}_\varepsilon$ and $\bar{v}_\varepsilon$ that are given as pointwise solution of (1.24) are continuous. To that end, we write (1.24) as

$$\mathrm{D}H_\varepsilon(u_\varepsilon, v_\varepsilon) = \begin{pmatrix} (U_\varepsilon - \Phi)_+ \\ (V_\varepsilon - \Psi)_+ \end{pmatrix}. \tag{3.8}$$

Since $\mathrm{D}H_\varepsilon$ has a continuous inverse (cf. Remark 3.1 (2)), and since the functions on the right-hand side of (3.8) are continuous on $\mathbb{R}^d$, so are the solutions $\bar{u}_\varepsilon$ and $\bar{v}_\varepsilon$. ◻

Proposition 3.5. *In addition to the general hypotheses, assume that (F, G, h) is k-degenerate for some $k \in \mathbb{N}$. Then $F'(\bar{u}_\varepsilon)$ and $G'(\bar{v}_\varepsilon)$ are k times continuously differentiable in Ω^u_ε and in Ω^v_ε, respectively. Moreover, all partial derivatives $\partial^\alpha F'(\bar{u}_\varepsilon)$ and $\partial^\alpha G'(\bar{v}_\varepsilon)$ of order $|\alpha| \leq k$ are bounded on Ω^u_ε and Ω^v_ε, respectively, uniformly in $\varepsilon \in [0, \varepsilon^*]$.*

Remark 3.6. The essential point of Proposition 3.5 is that $F'(\bar{u}_\varepsilon)$ is k times continuously differentiable across the boundary of the support $\partial\Omega^v_\varepsilon$ of *the other* component $\bar{v}_\varepsilon$, and vice versa. Across the boundary of *its own* support $\partial\Omega^u_\varepsilon$, the function $F'(\bar{u}_\varepsilon)$ is generally Lipschitz but no better; see Remark 3.7 after the proof. This translates into mere Hölder continuity for $\bar{u}_\varepsilon$, and in particular, one cannot expect the derivatives of $\bar{u}_\varepsilon$ itself to be bounded on Ω^u_ε.

Proof. The claim will follow by an application of the inverse function theorem. To that end, define the map $\Gamma_\varepsilon : \mathbb{R}^2 \to \mathbb{R}^2$ by

$$\Gamma_\varepsilon(\rho,\eta) = \mathrm{D}H_\varepsilon\big((F')^{-1}(\rho),(G')^{-1}(\eta)\big) = \begin{pmatrix} \rho + \varepsilon\theta_u(\rho,\eta) \\ \eta + \varepsilon\theta_v(\rho,\eta) \end{pmatrix},$$

with θ_u, θ_v from (1.16), and the convention that $\theta_u(\rho,\eta) = \theta_v(\rho,\eta) = 0$ if $\rho \le 0$ or $\eta \le 0$. By k-degeneracy, Γ_ε is C^k-regular.

Recall from Remark 3.1 (2) that $\mathrm{D}H_\varepsilon$ is a homeomorphism of $\mathbb{R}^2_{\ge 0}$. Since F' and G' are continuous and strictly monotone on $\mathbb{R}_{\ge 0}$, the restriction of Γ_ε to $\mathbb{R}^2_{\ge 0}$ possesses a continuous inverse as well. Moreover, on $\mathbb{R}^2 \setminus \mathbb{R}^2_{>0}$, the inverse of Γ_ε is simply the identity. In conclusion, Γ_ε has a global continuous inverse $\Gamma_\varepsilon^{-1} : \mathbb{R}^2 \to \mathbb{R}^2$ that is the identity on $\mathbb{R}^2 \setminus \mathbb{R}^2_{>0}$.

Next, we show that the inverses of the derivative matrices

$$\mathrm{D}\Gamma_\varepsilon = \begin{pmatrix} 1 + \varepsilon\partial_\rho\theta_u & \varepsilon\partial_\eta\theta_u \\ \varepsilon\partial_\rho\theta_v & 1 + \varepsilon\partial_\eta\theta_v \end{pmatrix} \tag{3.9}$$

are locally bounded. On $\mathbb{R}^2 \setminus \mathbb{R}^2_{>0}$, this is trivial since

$$\mathrm{D}\Gamma_\varepsilon(\rho,\eta) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{if } \rho \le 0 \text{ or } \eta \le 0. \tag{3.10}$$

On $\mathbb{R}^2_{>0}$, we use the alternative representation

$$\mathrm{D}\Gamma_\varepsilon(\rho,\eta) = \mathrm{D}^2 H_\varepsilon(u,v) \begin{pmatrix} \frac{1}{F''(u)} & 0 \\ 0 & \frac{1}{G''(v)} \end{pmatrix}, \quad \text{with} \quad u = (F')^{-1}(\rho),\, v = (G')^{-1}(\eta).$$

Since $u, v > 0$, the Hessian matrix $\mathrm{D}^2 H_\varepsilon(u,v)$ is positive definite. Since $R \ge M$ for symmetric, positive definite matrices implies $\det R \ge \det M > 0$, the estimate (3.2) shows that

$$\det \mathrm{D}\Gamma_\varepsilon(\rho,\eta) \ge \frac{1}{4} F''(u)G''(u) \cdot \frac{1}{F''(u)G''(v)} = \frac{1}{4}. \tag{3.11}$$

It follows that the inverses

$$(\mathrm{D}\Gamma_\varepsilon)^{-1} = \frac{1}{\det \mathrm{D}\Gamma_\varepsilon} \begin{pmatrix} 1 + \varepsilon\partial_\eta\theta_v & -\varepsilon\partial_\eta\theta_u \\ -\varepsilon\partial_\rho\theta_v & 1 + \varepsilon\partial_\rho\theta_u \end{pmatrix}$$

are bounded on any compact subset of $\mathbb{R}^2_{\ge 0}$, uniformly with respect to $\varepsilon \in [0,\varepsilon^*]$. The inverse function theorem is applicable and shows that Γ_ε^{-1} is of class C^k, with derivatives up to kth order ε-uniformly bounded on each compact set of $\mathbb{R}^2$.

To conclude regularity of $F'(\bar u_\varepsilon)$ and $G'(\bar v_\varepsilon)$ from here, we perform the following change of variables:

$$\rho(x) = F'(u(x)), \quad \eta(x) = G'(v(x)).$$

Written in terms of $\bar\rho_\varepsilon := F'(\bar u_\varepsilon)$ and $\bar\eta_\varepsilon := G'(\bar v_\varepsilon)$, the system (1.24) of Euler–Lagrange equations becomes

$$\Gamma_\varepsilon(\bar\rho_\varepsilon, \bar\eta_\varepsilon) = \begin{pmatrix} (U_\varepsilon - \Phi)_+ \\ (V_\varepsilon - \Psi)_+ \end{pmatrix}, \tag{3.12}$$

and its solution is given by

$$\begin{pmatrix}\bar\rho_\varepsilon\\ \bar\eta_\varepsilon\end{pmatrix} = (\Gamma_\varepsilon)^{-1}\left(\begin{pmatrix}(U_\varepsilon-\Phi)_+\\ (V_\varepsilon-\Psi)_+\end{pmatrix}\right). \tag{3.13}$$

On $\Omega_\varepsilon^u\cap\Omega_\varepsilon^v$, where we have

$$(U_\varepsilon-\Phi)_+ = U_\varepsilon-\Phi,\quad (V_\varepsilon-\Psi)_+ = V_\varepsilon-\Psi,$$

it now follows directly from (3.13) that $\bar\rho_\varepsilon$ and $\bar\eta_\varepsilon$ inherit the C^k-regularity of Φ and Ψ. Recalling that $\bar u_\varepsilon,\bar v_\varepsilon\le\bar U$, we conclude from the ε-uniform local boundedness of the derivatives of Γ_ε^{-1} that also the partial derivatives of $\bar\rho_\varepsilon$ and $\bar\eta_\varepsilon$ of order $\le k$ are bounded, uniformly in $\varepsilon\in[0,\varepsilon^*]$ on $\Omega_\varepsilon^u\cap\Omega_\varepsilon^v$.

Next, recalling that $\Gamma_\varepsilon(\rho,0)=(\rho,0)$, we observe that on $\Omega_\varepsilon^u\setminus\overline{\Omega_\varepsilon^v}$,

$$\bar\rho_\varepsilon = U_\varepsilon-\Phi. \tag{3.14}$$

Therefore, $\bar\rho_\varepsilon$ inherits the C^k-regularity and bounds from Φ. The analogous statement holds for $\bar\eta_\varepsilon$ on $\Omega_\varepsilon^v\setminus\overline{\Omega_\varepsilon^u}$.

It remains to verify the existence and continuity of all partial derivatives $\partial^\alpha\bar\rho_\varepsilon$ of order $|\alpha|\le k$ across the interfaces $\Omega_\varepsilon^u\cap\partial\Omega_\varepsilon^v$. We will do this by showing that at any point $x^*\in\Omega_\varepsilon^u\cap\partial\Omega_\varepsilon^v$, the limits of $\partial^\alpha\bar\rho_\varepsilon$ from the inside and from the outside of Ω_ε^v agree. Since, according to (3.14), the outside limit amounts to $-\partial^\alpha\Phi$, which is smooth on $\partial\Omega_\varepsilon^v$, this also proves continuity of $\bar\rho_\varepsilon$ across the boundary $\Omega_\varepsilon^u\cap\partial\Omega_\varepsilon^v$.

Thus, let $x^*\in\Omega_\varepsilon^u\cap\partial\Omega_\varepsilon^v$ be fixed. At points $x\in\Omega_\varepsilon^u\cap\Omega_\varepsilon^v$, (3.13) simplifies to

$$\begin{pmatrix}\bar\rho_\varepsilon(x)\\ \bar\eta_\varepsilon(x)\end{pmatrix} = \Gamma_\varepsilon^{-1}\left(\begin{pmatrix}U_\varepsilon-\Phi(x)\\ V_\varepsilon-\Psi(x)\end{pmatrix}\right).$$

We start by considering first derivatives. Since

$$\mathrm{D}\big(\Gamma_\varepsilon^{-1}\big) = (\mathrm{D}\Gamma_\varepsilon)^{-1}\circ\Gamma_\varepsilon^{-1} = \left[\frac{1}{\det\mathrm{D}\Gamma_\varepsilon}\begin{pmatrix}1+\varepsilon\partial_\eta\theta_v & -\varepsilon\partial_\eta\theta_u\\ -\varepsilon\partial_\rho\theta_v & 1+\varepsilon\partial_\rho\theta_u\end{pmatrix}\right]\circ\Gamma_\varepsilon^{-1},$$

it follows that

$$\partial_{x_k}\bar\rho_\varepsilon = \frac{1}{\det\mathrm{D}\Gamma_\varepsilon(\bar\rho_\varepsilon,\bar\eta_\varepsilon)}\big[-\big(1+\varepsilon\partial_\eta\theta_v(\bar\rho_\varepsilon,\bar\eta_\varepsilon)\big)\partial_{x_k}\Phi-\varepsilon\partial_\eta\theta_u(\bar\rho_\varepsilon,\bar\eta_\varepsilon)\partial_{x_k}\Psi\big]. \tag{3.15}$$

Since $\bar\rho_\varepsilon$ and $\bar\eta_\varepsilon$ are continuous at x^* with $\bar\eta_\varepsilon(x^*)=0$, by k-degeneracy we obtain $\partial_\rho\theta_u(\bar\rho_\varepsilon,\bar\eta_\varepsilon)\to0$ and $\partial_\eta\theta_u(\bar\rho_\varepsilon,\bar\eta_\varepsilon)\to0$ as $x\to x^*$, which via (3.15) implies $\partial_{x_k}\bar\rho_\varepsilon(x)\to-\partial_{x_k}\Phi(x^*)$, as desired. Higher-order partial derivatives can now be obtained by induction on the degree of differentiability. Assume that for $\ell<k$, uniform boundedness of all partial derivatives of order $\le\ell$ has been shown; note that the proof for $\ell=1$ is above. Application of $\partial^{\alpha'}$ with $|\alpha'|=\ell$ to (3.15) yields

$$\begin{aligned}\partial^{\alpha'}\partial_{x_k}\bar\rho_\varepsilon = &\sum_{\substack{\beta,\beta',\beta''\ge0\\ \beta+\beta'+\beta''=\alpha'}} C_{\beta,\beta',\beta''}\partial^\beta\left(\frac{1}{\det\mathrm{D}\Gamma_\varepsilon(\bar\rho_\varepsilon,\bar\eta_\varepsilon)}\right)\\ &\times\big[-\partial^{\beta'}\big(1+\varepsilon\partial_\eta\theta_v(\bar\rho_\varepsilon,\bar\eta_\varepsilon)\big)\partial^{\beta''}\partial_{x_k}\Phi-\varepsilon\partial^{\beta'}\partial_\eta\theta_u(\bar\rho_\varepsilon,\bar\eta_\varepsilon)\partial^{\beta''}\partial_{x_k}\Psi\big],\end{aligned}$$

with certain combinatorial coefficients $C_{\beta,\beta',\beta''}$. The partial derivatives of Φ and Ψ remain uniformly bounded as $x\to x^*$. The same is true for the partial derivatives of $1/\det \mathrm{D}\Gamma_\varepsilon$; the entries in $\mathrm{D}\Gamma_\varepsilon$ are C^{k-1}-regular functions by the k-degeneracy condition, and $\det \mathrm{D}\Gamma_\varepsilon$ is bounded below by virtue of (3.11). Further, observe that $\partial^{\beta'}\partial_\eta\theta_u(\bar\rho_\varepsilon,\bar\eta_\varepsilon)$ can be written as a finite weighted sum, where each term is a partial derivative (with respect to (ρ,η)) of order $\le \ell+1=k$ of θ_u at $(\bar\rho_\varepsilon(x),\bar\eta_\varepsilon(x))$—and hence continuous—multiplied by a product of partial derivatives (with respect to x) of $\bar\rho_\varepsilon$ and $\bar\eta_\varepsilon$ of order $\le \ell$—and hence uniformly bounded by induction hypothesis. As above, using that $\bar\eta_\varepsilon(x)\to 0$ as $x\to x^*$, by k-degeneracy we obtain $\partial^{\beta'}\partial_\eta\theta_u(\bar\rho_\varepsilon(x),\bar\eta_\varepsilon(x))\to 0$ and $\partial^{\beta'}\partial_\eta\theta_v(\bar\rho_\varepsilon(x),\bar\eta_\varepsilon(x))\to 0$ as $x\to x^*$, which implies $\partial^{\alpha'}\partial_{x_k}\bar\rho_\varepsilon(x)\to -\partial^{\alpha'}\partial_{x_k}\Phi(x^*)$ and thus finishes the proof. ∎

Remark 3.7. If (F,G,h) is 1-degenerate, then (3.10) implies that Γ_ε^{-1} is a differentiable perturbation of the identity near any point $(0,z)$ with $z\in\mathbb{R}$. Thus, the first component $\bar\rho_\varepsilon$ in (3.13) is of the form $(U_\varepsilon-\Phi)_+$ plus some differentiable perturbation near the boundary $\partial\Omega_\varepsilon^u$. This confirms Remark 3.6 that, in general, only Lipschitz regularity can be expected from $F'(\bar u_\varepsilon)$ across the boundary of its own support.

Example 3.8. The following example illustrates that without the hypothesis of 2-degeneracy, one cannot expect C^2-regularity of $F'(\bar u_\varepsilon)$ and $G'(\bar v_\varepsilon)$. We consider

$$F(u)=\frac{u^2}{2},\quad G(v)=\frac{v^2}{2},\quad h(u,v)=(uv)^2,$$

which is 1-degenerate, but not 2-degenerate. The system (1.24) attains the form

$$\bar u_\varepsilon(1+2\varepsilon\bar v_\varepsilon^2)=(U_\varepsilon-\Phi)_+,\quad \bar v_\varepsilon(1+2\varepsilon\bar u_\varepsilon^2)=(V_\varepsilon-\Psi)_+\,.$$

If $x^*\in\Omega_\varepsilon^u\cap\partial\Omega_\varepsilon^v$, then $\bar v_\varepsilon\approx(V_\varepsilon-\Psi)_+$ is Lipschitz but not differentiable at x^*; see Remark 3.7 above. Thus $\bar v_\varepsilon^2$ is once continuously differentiable, but fails to be twice differentiable at x^*. Consequently,

$$F'(\bar u_\varepsilon)=\bar u_\varepsilon=\frac{U_\varepsilon-\Phi}{1+2\varepsilon\bar v_\varepsilon^2}$$

fails to be twice differentiable at x^*.

COROLLARY 3.9. *Assume that (F,G,h) is 2-degenerate. Then $\partial_u h(\bar u_\varepsilon,\bar v_\varepsilon)$ and $\partial_v h(\bar u_\varepsilon,\bar v_\varepsilon)$ are ε-uniformly semiconvex, that is, there is a $K_0\ge 0$ such that*

$$\nabla^2\partial_u h(\bar u_\varepsilon,\bar v_\varepsilon)\ge -K_0\mathbf{1},\quad \nabla^2\partial_v h(\bar u_\varepsilon,\bar v_\varepsilon)\ge -K_0\mathbf{1} \tag{3.16}$$

on $\mathbb{R}^d$ for all $\varepsilon\in[0,\varepsilon^*]$.

Proof. By Proposition 3.5 above, it follows that the gradients and the Hessians of $F'(\bar u_\varepsilon)$ and $G'(\bar u_\varepsilon)$ are ε-uniformly bounded on the respective supports Ω_ε^u and Ω_ε^v. With

$$\partial_u h(\bar u_\varepsilon,\bar v_\varepsilon)=\theta_u\big(F'(\bar u_\varepsilon),G'(\bar v_\varepsilon)\big),$$

and with $\theta_u\in C^2(\mathbb{R}^2_{\ge 0})$, the first and second order derivatives of $\partial_u h(\bar u_\varepsilon,\bar v_\varepsilon)$ are ε-independently bounded on Ω_ε^u. In fact, omitting the arguments, we have the representation

$$\begin{aligned}\nabla^2\partial_u h&=\partial_\rho\theta_u\,\nabla^2F'+\partial_\eta\theta_u\,\nabla^2G'\\&\quad+\partial_{\rho\rho}\theta_u\,\nabla F'\otimes\nabla F'+\partial_{\eta\eta}\theta_u\,\nabla G'\otimes\nabla G'+\partial_{\rho\eta}\theta_u\,\big(\nabla F'\otimes\nabla G'+\nabla G'\otimes\nabla F'\big).\end{aligned}$$

Fix some x^* at the boundary of the support Ω_ε^u, and consider the expression above for x approaching x^* from inside Ω_ε^u. The gradients and Hessians of F' and G' remain bounded, whereas the partial derivatives of θ_u converge to zero because of 2-degeneracy, and because their argument $F'(\bar{u}_\varepsilon)$ converges to zero. Thus $\nabla^2\partial_u h(\bar{u}_\varepsilon,\bar{v}_\varepsilon)$ is continuous across the boundary of the support Ω_ε^u. This implies an ε-uniform bound on the second derivatives of $\partial_u h(\bar{u}_\varepsilon,\bar{v}_\varepsilon)$, and in particular, the semiconvexity estimate (3.16). ∎

4. Time-discrete variational approximation.

In this section, we assume all hypotheses on Φ, Ψ, F, G and h from section 1.3, and further that (F,G,h) is 2-bounded, 2-degenerate, and satisfies the swap condition. It will turn out that it is of crucial importance that ε is sufficiently small. To this end, we assume that $\varepsilon\in[0,\bar{\varepsilon}]$ is fixed, where $\bar{\varepsilon}\in(0,\varepsilon^*]$ is chosen such that

$$12\bar{\varepsilon}^2(A^2+W^2)\le 1 \quad \text{and} \quad 2K_0\bar{\varepsilon}\le\Lambda, \tag{4.1}$$

with ε^* from hypothesis (1.14), A the constant from Remark 1.1 related to the 2-boundedness and 2-degeneracy of (F,G,h), W the constant from the swap condition (1.17), K_0 the constant determined in Corollary 3.9, and Λ the lower ellipticity bound of Φ,Ψ from hypothesis (1.6). Note that this choice of $\bar{\varepsilon}$, in particular, implies non-negativity and strict convexity of $\mathbf{E}_\varepsilon$.

4.1. Yosida-regularization and results.

The fundamental object that we use for proving the results on existence and long-time asymptotics of solutions is the following Yosida-type regularization $\mathbf{E}_{\varepsilon,\tau}$ of $\mathbf{E}_\varepsilon$ with a time step $\tau>0$:

$$\mathbf{E}_{\varepsilon,\tau}\big((u,v)\big|(\bar{u},\bar{v})\big):=\frac{1}{2\tau}\widetilde{\mathbf{W}}_2\big((u,v),(\bar{u},\bar{v})\big)^2+\mathbf{E}_\varepsilon(u,v).$$

A time-discrete approximation of solutions to (1.1) will be obtained in JKO-style [26] by means of inductive minimization of $\mathbf{E}_{\varepsilon,\tau}$. The following certifies well-posedness of that induction.

LEMMA 4.1. *Given any pair $(\hat{u},\hat{v})\in[\mathcal{P}_2^r(\mathbb{R}^d)]^2$ of finite energy $\mathbf{E}_\varepsilon(\hat{u},\hat{v})<\infty$ with $\varepsilon\in[0,\varepsilon^*]$, there is a unique minimizing pair $(u^*,v^*)\in[\mathcal{P}_2^r(\mathbb{R}^d)]^2$ of $\mathbf{E}_{\varepsilon,\tau}(\cdot|(\hat{u},\hat{v}))$. Moreover, one has*

$$\mathbf{E}_\varepsilon(u^*,v^*)+\frac{\tau}{2}\left(\frac{\widetilde{\mathbf{W}}_2\big((u^*,v^*),(\hat{u},\hat{v})\big)}{\tau}\right)^2\le\mathbf{E}_\varepsilon(\hat{u},\hat{v}). \tag{4.2}$$

In particular, $\mathbf{E}_\varepsilon(u^*,v^*)\le\mathbf{E}_\varepsilon(\hat{u},\hat{v})$.

Proof. Existence and uniqueness of the minimizing pair $(u^*,v^*)\in[\mathcal{P}_2^r(\mathbb{R}^d)]^2$ follows from the direct methods in the calculus of variations, applied to the functional $\mathbf{E}_{\varepsilon,\tau}$ in $[\mathcal{P}_2^r(\mathbb{R}^d)]^2$. All three components $\mathbf{W}_2(u,\hat{u})^2$, $\mathbf{W}_2(v,\hat{v})^2$, and $\mathbf{E}_{\varepsilon,\tau}$ are non-negative (as $\varepsilon\in[0,\varepsilon^*]$), and are lower semicontinuous with respect to convergence in $[\mathcal{P}_2^r(\mathbb{R}^d)]^2$. Coercivity on $[\mathcal{P}_2^r(\mathbb{R}^d)]^2$ follows thanks to the control on the second moments of u and v by $\mathbf{E}_\varepsilon(u,v)$. This yields existence. For uniqueness, notice that $\mathbf{W}_2(u,\hat{u})^2$ is convex in u, that $\mathbf{W}_2(v,\hat{v})^2$ is convex in v, and that $\mathbf{E}_{\varepsilon,\tau}$ is strictly jointly convex in (u,v), again because of $\varepsilon\in[0,\varepsilon^*]$.

Inequality (4.2) now follows from (u^*,v^*) being a minimizer, which means that

$$\mathbf{E}_\varepsilon(u^*,v^*)+\frac{\widetilde{\mathbf{W}}_2\big((u^*,v^*),(\hat{u},\hat{v})\big)^2}{2\tau}=\mathbf{E}_{\varepsilon,\tau}\big((u^*,v^*)\big|(\hat{u},\hat{v})\big)\le\mathbf{E}_{\varepsilon,\tau}\big((\hat{u},\hat{v})\big|(\hat{u},\hat{v})\big)=\mathbf{E}_\varepsilon(\hat{u},\hat{v}).$$

∎

The goal of this section is to prove the three results in Propositions 4.2, 4.3, and 4.4 below on the minimizing pairs (u^*, v^*). In the statements of the propositions, it is understood that $\varepsilon \in [0, \bar{\varepsilon}]$ is fixed, that $(\hat{u}, \hat{v}) \in [\mathcal{P}_2^r(\mathbb{R}^d)]^2$ is a given datum of finite energy $\mathbf{E}_\varepsilon(\hat{u}, \hat{v}) < \infty$, and that (u^*, v^*) is the associated minimizer of $\mathbf{E}_{\varepsilon,\tau}(\cdot|(\hat{u}, \hat{v}))$ in $[\mathcal{P}_2^r(\mathbb{R}^d)]^2$.

The *first result* shows that (u^*, v^*) satisfies a time-discrete weak formulation of the evolution equations (1.1).

PROPOSITION 4.2. *For each $\zeta \in C_c^\infty(\mathbb{R}^d)$ there holds*

$$\begin{aligned} \int_{\mathbb{R}^d} \frac{u^* - \hat{u}}{\tau} \zeta \, \mathrm{d}x &= \int_{\mathbb{R}^d} u^* \nabla \big[F'(u^*) + \varepsilon \partial_u h(u^*, v^*) + \Phi \big] \cdot \nabla \zeta \, \mathrm{d}x + R_u, \\ \int_{\mathbb{R}^d} \frac{v^* - \hat{v}}{\tau} \zeta \, \mathrm{d}x &= \int_{\mathbb{R}^d} v^* \nabla \big[G'(v^*) + \varepsilon \partial_v h(u^*, v^*) + \Psi \big] \cdot \nabla \zeta \, \mathrm{d}x + R_v, \end{aligned} \tag{4.3}$$

with remainder terms R_u and R_v satisfying

$$|R_u| + |R_v| \le \|\zeta\|_{C^2} \big(\mathbf{E}_\varepsilon(\hat{u}, \hat{v}) - \mathbf{E}_\varepsilon(u^*, v^*) \big). \tag{4.4}$$

The *second result*, which is the key ingredient for our proof of Theorem 1.4, is an estimate on $F'(u^*)$ and $G'(v^*)$ in H^1. It is formulated with help of the entropy functional $\mathbf{H}$, introduced in (2.8). For brevity, we also define $\widetilde{\mathbf{H}}$ on $[\mathcal{P}_2^r(\mathbb{R}^d)]^2$ by $\widetilde{\mathbf{H}}(u, v) := \mathbf{H}(u) + \mathbf{H}(v)$. By Lemma B.1 from Appendix A, $\widetilde{\mathbf{H}}(u, v) > -\infty$.

PROPOSITION 4.3. *There is a constant C independent of $(\hat{u}, \hat{v})$ such that*

$$\begin{aligned} &\int_{\mathbb{R}^d} \big(|\nabla F'(u^*)|^2 + |\nabla G'(v^*)|^2 \big) \, \mathrm{d}x \\ &\quad \le C \left[1 + \mathbf{E}_\varepsilon(\hat{u}, \hat{v}) + \frac{\mathbf{E}_\varepsilon(\hat{u}, \hat{v}) - \mathbf{E}_\varepsilon(u^*, v^*)}{\tau} + \frac{\widetilde{\mathbf{H}}(\hat{u}, \hat{v}) - \widetilde{\mathbf{H}}(u^*, v^*)}{\tau} \right]. \end{aligned} \tag{4.5}$$

The *third result*, which contains the essence of the proof of Theorem 1.5, is concerned with proximity of u^* and v^* to the respective stationary solutions $\bar{u}_\varepsilon$ and $\bar{v}_\varepsilon$. Recall for a given strictly convex function $J : \mathbb{R}_{\ge 0} \to \mathbb{R}$ the definition of the Bregman divergence $d_J(\cdot|\cdot) : \mathbb{R}_{\ge 0} \times \mathbb{R}_{\ge 0} \to \mathbb{R}_{\ge 0}$ as

$$d_J(s|\bar{s}) := J(s) - \big[J(\bar{s}) + (s - \bar{s}) J'(\bar{s}) \big]. \tag{4.6}$$

By strict convexity of J, $d_J(s|\bar{s})$ is always nonnegative, and is zero if and only if $s = \bar{s}$. Next, introduce the relative entropy functionals $\mathbf{L}_1$ and $\mathbf{L}_2$ on $\mathcal{P}_2^r(\mathbb{R}^d)$ by

$$\mathbf{L}_1(u) := \int_{\mathbb{R}^d} \big[d_F(u|\bar{u}_\varepsilon) + u(\Phi - U_\varepsilon)_+ \big] \, \mathrm{d}x, \quad \mathbf{L}_2(v) := \int_{\mathbb{R}^d} \big[d_G(v|\bar{v}_\varepsilon) + v(\Psi - V_\varepsilon)_+ \big] \, \mathrm{d}x, \tag{4.7}$$

which are clearly nonnegative (as sums of nonnegative parts), and zero if and only if $u = \bar{u}_\varepsilon$ and $v = \bar{v}_\varepsilon$, respectively. Finally, define $\mathbf{L}$ on $[\mathcal{P}_2^r(\mathbb{R}^d)]^2$ by

$$\mathbf{L}(u, v) = \mathbf{L}_1(u) + \mathbf{L}_2(v). \tag{4.8}$$

The equivalence of these definitions of $\mathbf{L}$ and $\mathbf{L}_1$, $\mathbf{L}_2$ to the ones in (1.4) is shown in section 4.4.

PROPOSITION 4.4. *There is a constant $K>0$ independent of $(\hat u,\hat v)$ such that*

$$\mathbf{L}(\hat u,\hat v)-\mathbf{L}(u^*,v^*)\ge 2\tau(\Lambda-K\varepsilon)\mathbf{L}(u^*,v^*). \tag{4.9}$$

Our strategy for proving Propositions 4.2, 4.3, and 4.4 is the following. First, we prove these results *under the additional hypothesis on* $(\hat u,\hat v)$:

(4.10)
$\hat u$, $\hat v$ are positive a.e. on a ball $\mathbb{B}_R$ of some radius $R>0$, and vanish a.e. outside.

In section 4.7, we remove this additional hypothesis and generalize the propositions to arbitrary data $(\hat u,\hat v)$ of finite energy.

In sections 4.2–4.6 below, the datum $(\hat u,\hat v)$ is fixed and satisfies (4.10). Accordingly, (u^*,v^*) is the minimizer of $\mathbf{E}_{\varepsilon,\tau}(\cdot|(\hat u,\hat v))$ in $[\mathcal{P}_2^r(\mathbb{R}^d)]^2$. To fix notations, denote by (φ_u,ψ_u) and by (φ_v,ψ_v) the optimal pairs of c-conjugate potentials for the transports from u^* to $\hat u$, and from v^* to $\hat v$, respectively. By section 2.1, these potentials are $u^*\mathcal{L}^d$-a.e. uniquely determined, up to a global constant; we fix points $\bar x_u$ and $\bar x_v$ in the support of u and of v, respectively, and normalize $\varphi_u(\bar x_u)=0$ and $\varphi_v(\bar x_v)=0$.

4.2. Euler–Lagrange equation. In this section, we prove Proposition 4.2 under the additional hypothesis (4.10).

LEMMA 4.5. *There are constants $C_u,C_v\in\mathbb{R}$ such that*

$$\begin{aligned} C_u-\frac{\varphi_u}{\tau}&=F'(u^*)+\varepsilon\partial_u h(u^*,v^*)+\Phi, \qquad && u^*\mathcal{L}^d\text{-a.e.},\\ C_v-\frac{\varphi_v}{\tau}&=G'(v^*)+\varepsilon\partial_v h(u^*,v^*)+\Psi, \qquad && v^*\mathcal{L}^d\text{-a.e.} \end{aligned}\tag{4.11}$$

Proof. We only consider the u-component. Let $\rho\in L^\infty(\mathbb{R}^d)$ be given, which vanishes outside of some open set $\Omega\subset\mathbb{R}^d$ with compact closure. We assume that

$$\int_{\mathbb{R}^d}\rho u^*\,\mathrm{d}x=0, \tag{4.12}$$

which can always be achieved by addition of a suitable multiple of the indicator function of Ω to ρ. For $\delta\in\mathbb{R}$ such that $0<|\delta|\,\|\rho\|_{L^\infty}<1$, let $u^\delta:=(1+\delta\rho)u^*$, and note that $u^\delta\in\mathcal{P}_2^r(\mathbb{R}^d)$ thanks to identity (4.12). Next, let $(\varphi_u^\delta,\psi_u^\delta)$ be an optimal pair of c-conjugate potentials for the transport from u^δ to $\hat u$ with the normalization $\varphi_u^\delta(\bar x_u)=0$. Recalling that $\hat u$ satisfies the additional hypothesis (4.10), φ_u^δ is uniquely determined $u^*\mathcal{L}^d$-a.e. (note that $u^\delta\mathcal{L}^d$ and $u^*\mathcal{L}^d$ have the same negligible sets), up to a global constant. Since $\bar x_u$ is in support of u^δ, we can normalize φ_u^δ by $\varphi_u^\delta(\bar x_u)=0$.

For later reference, recall that the auxiliary potential $\tilde\varphi_u^\delta(x):=\frac12|x|^2-\varphi_u^\delta(x)$ is a proper, lower semicontinuous, and convex function. Moreover, as the associated optimal transport map $T_u^\delta=\nabla\tilde\varphi_u^\delta$ maps $u^*\mathcal{L}^d$-almost surely onto the support of $\hat u$, i.e., into $\mathbb{B}_R$, it follows that $|\nabla\tilde\varphi_u^\delta|\le R$ $u^*\mathcal{L}^d$-a.e. For convenience and without loss of generality, we may actually assume that on $u^*\mathcal{L}^d$-negligible sets, φ_u^δ is defined such that

$$|\nabla\tilde\varphi_u^\delta|\le R\quad\text{a.e. on }\mathbb{R}^d. \tag{4.13}$$

Clearly, $(\varphi_u^\delta,\psi_u^\delta)$ is a (in general, suboptimal) pair of c-conjugate potentials for the transport from u^* to $\hat u$. Recalling further the definition of (u^*,v^*) as the minimizer of $\mathbf{E}_{\varepsilon,\tau}(\cdot|(\hat u,\hat v))$, we conclude the following chain of inequalities:

$$\begin{aligned}
&\frac{1}{\tau}\bigg(\int_{\mathbb{R}^d}\varphi_u^\delta(x)u^*(x)\,\mathrm{d}x+\int_{\mathbb{R}^d}\psi_u^\delta(y)\hat{u}(y)\,\mathrm{d}y+\frac{1}{2}\mathbf{W}_2(v^*,\hat{v})^2\bigg)+\mathbf{E}_\varepsilon(u^*,v^*)\\
&\quad\le \mathbf{E}_{\varepsilon,\tau}\big((u^*,v^*)\big|(\hat{u},\hat{v})\big)\\
&\quad\le \mathbf{E}_{\varepsilon,\tau}\big((u^\delta,v^*)\big|(\hat{u},\hat{v})\big)\\
&\quad=\frac{1}{\tau}\bigg(\int_{\mathbb{R}^d}\varphi_u^\delta(x)u^\delta(x)\,\mathrm{d}x+\int_{\mathbb{R}^d}\psi_u^\delta(y)\hat{u}(y)\,\mathrm{d}y+\frac{1}{2}\mathbf{W}_2(v^*,\hat{v})^2\bigg)+\mathbf{E}_\varepsilon(u^\delta,v^*).
\end{aligned}$$

It thus follows that

$$0\le\int_{\mathbb{R}^d}\frac{\varphi_u^\delta}{\tau}\frac{u^\delta-u^*}{\delta}\,\mathrm{d}x+\frac{1}{\delta}\big(\mathbf{E}_\varepsilon(u^\delta,v^*)-\mathbf{E}_\varepsilon(u^*,v^*)\big). \tag{4.14}$$

We shall now pass to the limit $\delta\downarrow 0$. Since $\tilde{\varphi}_u^\delta(\bar{x}_u)=0$ is fixed, and since (4.13) gives a uniform Lipschitz bound, the Arzelà–Ascoli theorem yields local uniform convergence of $\tilde{\varphi}_u^{\delta_k}$ to a limit $\tilde{\varphi}_u^0$ along a suitable sequence $\delta_k\downarrow 0$. We wish to show that

$$\tilde{\varphi}_u^0(x)=\tilde{\varphi}_u(x):=\frac{1}{2}|x|^2-\varphi_u(x)\quad\text{for } u^*\mathcal{L}^d\text{-a.e. } x. \tag{4.15}$$

By convexity of the $\tilde{\varphi}_u^{\delta_k}$, local uniform convergence of the function values implies $\mathcal{L}^d$-a.e. convergence of the gradients. So, in particular, $T_u^{\delta_k}\to T_u^0:=\nabla\tilde{\varphi}_u^0$ $u^*\mathcal{L}^d$-a.e. But $T_u^0\#u^*=\hat{u}$, since for every $\omega\in C_c(\mathbb{R}^d)$

$$\begin{aligned}
\int_{\mathbb{R}^d}\omega\,\hat{u}\,\mathrm{d}y&=\int_{\mathbb{R}^d}\omega\,T_u^{\delta_k}\#u^\delta\,\mathrm{d}y=\int_{\mathbb{R}^d}\omega\circ T_u^{\delta_k}\,u^\delta\,\mathrm{d}x\\
&\to\int_{\mathbb{R}^d}\omega\circ T_u^0\,u^*\,\mathrm{d}x=\int_{\mathbb{R}^d}\omega\,T_u^0\#u^*\,\mathrm{d}x.
\end{aligned}$$

The limit in the chain above is justified by the dominated convergence theorem, since $\omega\circ T_u^{\delta_k}\to\omega\circ T_u^0$ $u^*\mathcal{L}^d$-a.e., since ω is bounded, and since $u^\delta-u^*=\delta\rho u^*$ by construction. This means that $\tilde{\varphi}_u^0$ is an auxiliary optimal potential for the transport from u^* to $\hat{u}$. Using again $u^*\mathcal{L}^d$-a.e. uniqueness of such an optimal potential up to a global constant thanks to (4.10), and observing that $\tilde{\varphi}_u^0(\bar{x}_u)=0$ by local uniform convergence, we conclude (4.15). Now, since $(u^\delta-u^*)/\delta=u^*\rho$ with ρ bounded, it follows that

$$\lim_{k\to\infty}\int_{\mathbb{R}^d}\frac{\varphi_u^{\delta_k}}{\tau}\frac{u^{\delta_k}-u^*}{\delta_k}\,\mathrm{d}x=\int_{\mathbb{R}^d}\frac{\varphi_u}{\tau}u^*\rho\,\mathrm{d}x.$$

Note that local uniform convergence has been sufficient here since ρ vanishes outside the compact set $\bar{\Omega}$.

Further, since $H_\varepsilon(u^*,v^*)$ is integrable on $\mathbb{R}^d$, so is $H_\varepsilon((1+\delta\rho)u^*,v^*)$ thanks to the doubling condition (1.12) and to (1.15), and the variational derivative of $\mathbf{E}_\varepsilon$ in direction ρu^* is readily computed by standard methods. With (4.14), this leads to

$$0\le\int_{\mathbb{R}^d}\Big[\frac{\varphi_u}{\tau}+F'(u^*)+\Phi+\varepsilon\partial_u h(u^*,v^*)\Big]\,u^*\rho\,\mathrm{d}x. \tag{4.16}$$

The same argument applies for $0>\delta>-1/\|\rho\|_{L^\infty}$, when the relation in (4.14) is reversed. Consequently, (4.16) holds with the reversed relation as well, i.e., it is an equality. Since ρ has been an arbitrary bounded function of compact support, only subject to the normalization (4.12), the term in square parenthesis above equals to a global constant C_u. The value of C_u is determined by our normalization $\varphi_u(\bar{x}_u)=0$. This finishes the proof of the lemma. ∎

Proof of Proposition 4.2 *under the additional hypothesis* (4.10). Let $T(x) = x - \nabla\varphi_u(x)$ be the optimal transport map from u^* to u, recalling that $\hat{u} = T\#u^*$ and the definition (2.1) of the push-forward, we obtain

$$\begin{aligned}
&\int_{\mathbb{R}^d} \frac{u^* - \hat{u}}{\tau}\zeta \,\mathrm{d}x \\
&= \frac{1}{\tau}\left(\int_{\mathbb{R}^d} \zeta(x)u^*(x)\,\mathrm{d}x - \int_{\mathbb{R}^d} \zeta\circ T(x)u^*(x)\,\mathrm{d}x\right) \\
&= \frac{1}{\tau}\int_{\mathbb{R}^d} \big(\zeta - \zeta\circ T\big)u^*\,\mathrm{d}x \\
&= \frac{1}{\tau}\int_{\mathbb{R}^d} \Big[\nabla\zeta(x)\cdot\big(x - T(x)\big) - \frac{1}{2}\big(x - T(x)\big)^T\nabla^2\zeta(m_x)\big(x - T(x)\big)\Big]u^*(x)\,\mathrm{d}x,
\end{aligned}$$

where m_x is a suitable intermediate point on the line connecting x to $T(x)$. In summary,

$$\int_{\mathbb{R}^d} \frac{u^* - \hat{u}}{\tau}\zeta\,\mathrm{d}x = \int_{\mathbb{R}^d} u^*\nabla\left(\frac{\varphi_u}{\tau}\right)\cdot\nabla\zeta\,\mathrm{d}x + R_u, \tag{4.17}$$

where, thanks to (2.5),

$$|R_u| \leq \frac{1}{2\tau}\|\nabla^2\zeta\|_\infty \int_{\mathbb{R}^d} |\nabla\varphi_u|^2 u^*\,\mathrm{d}x \leq \frac{\|\zeta\|_{C^2}}{2\tau}\mathbf{W}_2(u^*,\hat{u})^2. \tag{4.18}$$

Substitution of (4.11)—which is relying on the additional hypothesis (4.10)—into (4.17) above produces the first equation in (4.3).

The v-component is treated in a similar way, leading to the second equation in (4.3). Adding the two estimates of the form (4.18) and using (4.2) we obtain

$$|R_u| + |R_v| \leq \frac{\|\zeta\|_{C^2}}{2\tau}\big[\mathbf{W}_2(u^*,\hat{u})^2 + \mathbf{W}_2(v^*,\hat{v})^2\big] \leq \|\zeta\|_{C^2}\big[\mathbf{E}_\varepsilon(\hat{u},\hat{v}) - \mathbf{E}_\varepsilon(u^*,v^*)\big]. \qquad \square$$

4.3. Regularity estimates. In this section, we prove Proposition 4.3, subject to (4.10). The proof follows from Lemmas 4.6 and 4.7 below, where the first condition in (4.1) concerning the smallness of $\bar{\varepsilon}$ becomes relevant. The additional hypothesis (4.10) only enters indirectly, via Lemma 4.5.

LEMMA 4.6. *With a constant C independent of $(\hat{u},\hat{v})$, there holds*

$$\int_{\mathbb{R}^d} \big[u^*|\nabla F'(u^*)|^2 + v^*|\nabla G'(v^*)|^2\big]\,\mathrm{d}x \leq C\left(\mathbf{E}_\varepsilon(\hat{u},\hat{v}) + \frac{\mathbf{E}_\varepsilon(\hat{u},\hat{v}) - \mathbf{E}_\varepsilon(u^*,v^*)}{\tau}\right). \tag{4.19}$$

In particular, $\nabla F'(u^*) \in L^2(\mathbb{R}^d; u^*\mathcal{L}^d)$ and $\nabla G'(v^*) \in L^2(\mathbb{R}^d; v^*\mathcal{L}^d)$.

Proof. Thanks to Lemma 4.5 and the properties of c-conjugate potentials, the sum $F'(u^*) + \varepsilon\partial_u h(u^*,v^*) + \Phi$ is differentiable $u^*\mathcal{L}^d$-a.e. We apply the binomial theorem to the first equation in (4.11) and use (1.19), obtaining

$$\begin{aligned}
\frac{1}{3}|\nabla F'(u^*)|^2 &\leq \big|\nabla\big[F'(u^*) + \varepsilon\partial_u h(u^*,v^*) + \Phi\big]\big|^2 + \varepsilon^2|\nabla\partial_u h(u^*,v^*)|^2 + |\nabla\Phi|^2 \\
&\leq \left|-\nabla\frac{\varphi_u}{\tau}\right|^2 + 2\varepsilon^2\big(\partial_\rho\theta_u\big)^2|\nabla F'(u^*)|^2 + 2\varepsilon^2\big(\partial_\eta\theta_u\big)^2|\nabla G'(v^*)|^2 + \frac{2M^2}{\Lambda}\Phi.
\end{aligned}$$

By means of (2.5), the bound (1.21) thanks to 2-boundedness and 2-degeneracy of (F, G, h), and the swap condition (1.17), it follows that

$$\frac{1}{3}\int_{\mathbb{R}^d} u^*|\nabla F'(u^*)|^2\,\mathrm{d}x \le \left(\frac{\mathbf{W}_2(u^*,\hat{u})}{\tau}\right)^2 + 2\varepsilon^2 A^2 \int_{\mathbb{R}^d} u^*|\nabla F'(u^*)|^2\,\mathrm{d}x$$
$$+ 2\varepsilon^2 W^2 \int_{\mathbb{R}^d} v^*|\nabla G'(v^*)|^2\,\mathrm{d}x + \frac{2M^2}{\Lambda}\int_{\mathbb{R}^d}\Phi u^*\,\mathrm{d}x.$$

In combination with the analogous estimate for v^* in place of u^*, and observing that $H_\varepsilon \ge 0$, we obtain that

$$\left(\frac{1}{3} - 2\varepsilon^2(A^2+W^2)\right)\left(\int_{\mathbb{R}^d} u^*|\nabla F'(u^*)|^2\,\mathrm{d}x + \int_{\mathbb{R}^d} v^*|\nabla G'(v^*)|^2\,\mathrm{d}x\right)$$
$$\le \left(\frac{\mathbf{W}_2(u^*,\hat{u})}{\tau}\right)^2 + \left(\frac{\mathbf{W}_2(v^*,\hat{v})}{\tau}\right)^2 + \frac{2M^2}{\Lambda}\mathbf{E}_\varepsilon(u^*,v^*).$$

The result now follows using (4.2) and the choice of $\bar{\varepsilon}$ in (4.1). ∎

LEMMA 4.7. *With a constant C independent of $(\hat{u},\hat{v})$, there holds*

(4.20)
$$\int_{\mathbb{R}^d}\left(\left|\nabla[F'(u^*)]_{F'(1)}\right|^2 + \left|\nabla[G'(v^*)]_{G'(1)}\right|^2\right)\mathrm{d}x \le C\left(1 + \frac{\widetilde{\mathbf{H}}(\hat{u},\hat{v}) - \widetilde{\mathbf{H}}(u^*,v^*)}{\tau}\right),$$

where $[z]_k := \min\{k, z\}$ is the cut-off at the value k.

Proof. The proof uses the method of flow interchange, which estimates the effect of variations of a Yosida-regularized *nonconvex* functional along the gradient flow of an auxiliary *convex* functional. The method has been introduced in [32], unifying several similar ideas from the literature; see, e.g., [26, 24].

For all $s > 0$, define perturbations $(U_s, V_s) \in [\mathcal{P}_2^r(\mathbb{R}^d)]^2$ of $(U_0, V_0) := (u^*, v^*)$ as follows:

$$U_s := \mathcal{K}_s * u^*,\ V_s := \mathcal{K}_s * v^* \quad \text{with} \quad \mathcal{K}_s(z) = (4\pi s)^{-d/2}\exp(-(4s)^{-1}|z|^2).$$

Since $\mathcal{K}_s(z)$ is the fundamental solution of the heat equation, it is well known that $(s,x)\mapsto U_s(x)$ and $(s,x)\mapsto V_s(x)$ are C^∞-smooth on $(0,\infty)\times\mathbb{R}^d$, with

$$\partial_s U_s = \Delta U_s, \quad \partial_s V_s = \Delta V_s \tag{4.21}$$

in the classical sense, and that $U_s \to u^*$ and $V_s \to v^*$ in $L^1(\mathbb{R}^d)$. Moreover, $(U_s)_{s>0}$ and $(V_s)_{s>0}$—considered as flows on $\mathcal{P}_2^r(\mathbb{R}^d)$—satisfy the ($\mathrm{EVI}_0$); see (2.9).

We perform a detailed comparison of the $\mathbf{E}_{\varepsilon,\tau}$-scores of (U_s, V_s) and of (u^*, v^*). By minimality, we know that $\mathbf{E}_{\varepsilon,\tau}((u^*,v^*)|(\hat{u},\hat{v})) \le \mathbf{E}_{\varepsilon,\tau}((U_r,V_r)|(\hat{u},\hat{v}))$; consequently, for each $\sigma > 0$,

(4.22)
$$\frac{\mathbf{E}_\varepsilon(u^*,v^*) - \mathbf{E}_\varepsilon(U_\sigma,V_\sigma)}{\sigma}$$
$$\le \frac{1}{2\tau}\left(\frac{\mathbf{W}_2(U_\sigma,\hat{u})^2 - \mathbf{W}_2(u^*,\hat{u})^2}{\sigma} + \frac{\mathbf{W}_2(V_\sigma,\hat{v})^2 - \mathbf{W}_2(v^*,\hat{v})^2}{\sigma}\right).$$

We consider the limes superior as $\sigma \downarrow 0$ on both sides. On the right-hand side, the (EVI_0) from (2.9) is applicable and yields

$$\begin{aligned}\limsup_{\sigma\downarrow 0} &\frac{\mathbf{E}_\varepsilon(u^*,v^*)-\mathbf{E}_\varepsilon(U_\sigma,V_\sigma)}{\sigma} \\ &\le \frac{\mathbf{H}(\hat u)-\mathbf{H}(u^*)}{\tau}+\frac{\mathbf{H}(\hat v)-\mathbf{H}(v^*)}{\tau}=\frac{\widetilde{\mathbf{H}}(\hat u,\hat v)-\widetilde{\mathbf{H}}(u^*,v^*)}{\tau}.\end{aligned}\tag{4.23}$$

For estimation on the left-hand side, we use the heat equation (4.21). By regularity and convexity of H_ε, it easily follows that $s\mapsto \mathbf{E}_\varepsilon(U_s,V_s)$ is continuous at $\sigma=0^+$. Thanks to smoothness for $\sigma>0$, we can now write

$$\frac{\mathbf{E}_\varepsilon(u^*,v^*)-\mathbf{E}_\varepsilon(U_\sigma,V_\sigma)}{\sigma}=-\frac{1}{\sigma}\int_0^\sigma \partial_s\mathbf{E}_\varepsilon(U_s,V_s)\,\mathrm{d}s$$

by means of the fundamental theorem of calculus. Next, using also smoothness in x and the convexity estimate (3.2), we obtain

$$\begin{aligned}&-\partial_s\mathbf{E}_\varepsilon(U_s,V_s)\\ &\quad=-\int_{\mathbb{R}^d}\big([\partial_u H_\varepsilon(U_s,V_s)+\Phi]\Delta U_s+[\partial_v H_\varepsilon(U_s,V_s)+\Psi]\Delta V_s\big)\,\mathrm{d}x\\ &\quad=\int_{\mathbb{R}^d}\big(\nabla\partial_u H_\varepsilon(U_s,V_s)\cdot\nabla U_s+\nabla\partial_v H_\varepsilon(U_s,V_s)\cdot\nabla V_s-\Delta\Phi\,U_s-\Delta\Psi\,V_s\big)\,\mathrm{d}x\\ &\quad\ge\int_{\mathbb{R}^d}\sum_{j=1}^d\begin{pmatrix}\partial_{x_j}U_s\\ \partial_{x_j}V_s\end{pmatrix}\cdot\mathrm{D}^2H_\varepsilon(U_s,V_s)\cdot\begin{pmatrix}\partial_{x_j}U_s\\ \partial_{x_j}V_s\end{pmatrix}-dM\int_{\mathbb{R}^d}(U_s+V_s)\,\mathrm{d}x\\ &\quad\ge\frac12\sum_{j=1}^d\int_{\mathbb{R}^d}\begin{pmatrix}\partial_{x_j}U_s\\ \partial_{x_j}V_s\end{pmatrix}\cdot\begin{pmatrix}F''(U_s)&0\\0&G''(V_s)\end{pmatrix}\cdot\begin{pmatrix}\partial_{x_j}U_s\\ \partial_{x_j}V_s\end{pmatrix}-2dM\\ &\quad\ge\frac{1}{2B}\int_{\mathbb{R}^d}\Big[\big|\nabla[F'(U_s)]_{F'(1)}\big|^2+\big|\nabla[G'(V_s)]_{G'(1)}\big|^2\Big]\,\mathrm{d}x-2dM,\end{aligned}$$

where $B=\max_{0\le r\le 1}F''(r)$, since by monotonicity of F', we have a.e.

$$\begin{aligned}\big|\nabla[F'(U_s)]_{F'(1)}\big|^2&=\begin{cases}F''(U_s)^2|\nabla U_s|^2 & \text{if } 0\le U_s<1,\\ 0 & \text{if } U_s\ge 1\end{cases}\\ &\le BF''(U_s)|\nabla U_s|^2.\end{aligned}$$

In combination with (4.23), we have

$$\begin{aligned}\limsup_{\sigma\downarrow 0}&\frac{1}{2B\sigma}\int_0^\sigma\int_{\mathbb{R}^d}\Big[\big|\nabla[F'(U_s)]_{F'(1)}\big|^2+\big|\nabla[G'(V_s)]_{G'(1)}\big|^2\Big]\,\mathrm{d}x\,\mathrm{d}\sigma\\ &\le\frac{\widetilde{\mathbf{H}}(\hat u,\hat v)-\widetilde{\mathbf{H}}(u^*,v^*)}{\tau}+2dM.\end{aligned}\tag{4.24}$$

In addition, observe that, thanks to the at-most linear growth of F' near zero, $[F'(U_s)]_{F'(1)}$ is uniformly controlled in $L^2(\mathbb{R}^d)$ by the L^1-norm of U_s, which is one. With (4.24) it now follows that $[F'(U_\sigma)]_{F'(1)}$ is uniformly bounded in $H^1(\mathbb{R}^d)$ at least along some sequence with $\sigma\downarrow 0$. By Rellich's lemma, one may assume strong convergence of that sequence in $L^2(\mathbb{R}^d)$, and thus identify the limit as $[F'(u^*)]_{F'(1)}$, since

$U_s \to u^*$ strongly in $L^1(\mathbb{R}^d)$ for $s \downarrow 0$. By Alaoglu's theorem, $F'(U_\sigma)$ converges weakly in $H^1(\mathbb{R}^d)$ to $F'(u^*)$, and by lower semicontinuity of norms, we finally obtain from (4.24)

$$\frac{1}{2B}\int_{\mathbb{R}^d}\left[\left|\nabla[F'(u^*)]_{F'(1)}\right|^2+\left|\nabla[G'(v^*)]_{G'(1)}\right|^2\right]\mathrm{d}x \le \frac{\widetilde{\mathbf{H}}(\hat{u},\hat{v})-\widetilde{\mathbf{H}}(u^*,v^*)}{\tau}+2dM,$$

which immediately yields the claim (4.20). ∎

Proof of Proposition 4.3 *under the additional hypothesis* (4.10). Combine (4.19) and (4.20), using that

$$|\nabla F'(u^*)|^2 \le |\nabla[F'(u^*)]_{F'(1)}|^2 + u^*|\nabla F'(u^*)|^2$$

because $\nabla F'(u^*) = \nabla[F'(u^*)]_{F'(1)}$ if $u^* < 1$, and $|\nabla F'(u^*)|^2 \le u^*|\nabla F'(u^*)|^2$ if $u^* \ge 1$. ∎

4.4. Definition of auxiliary functionals. In this section and the next, we lay the basis for the proof of Proposition 4.4 in section 4.6. With $(\bar{u}_\varepsilon,\bar{v}_\varepsilon) \in [\mathcal{P}_2^r(\mathbb{R}^d)]^2$ denoting the unique stationary pair of densities and θ_u and θ_v introduced in (1.16), we shall use the abbreviation

$$\bar{\Theta}_u := \theta_u\big(F'(\bar{u}_\varepsilon),G'(\bar{v}_\varepsilon)\big) = \partial_u h(\bar{u}_\varepsilon,\bar{v}_\varepsilon), \quad \bar{\Theta}_v := \theta_v\big(F'(\bar{u}_\varepsilon),G'(\bar{v}_\varepsilon)\big) = \partial_v h(\bar{u}_\varepsilon,\bar{v}_\varepsilon). \tag{4.25}$$

This allows for an alternative representation of the functionals $\mathbf{L}_1$ and $\mathbf{L}_2$ defined in (4.7). As indicated in the introduction (see (1.4)),

$$\begin{aligned}\mathbf{L}_1(u) &= \int_{\mathbb{R}^d}\left[F(u)+(\Phi+\varepsilon\bar{\Theta}_u)u\right]\mathrm{d}x - \int_{\mathbb{R}^d}\left[F(\bar{u}_\varepsilon)+(\Phi+\varepsilon\bar{\Theta}_u)\bar{u}_\varepsilon\right]\mathrm{d}x,\\ \mathbf{L}_2(v) &= \int_{\mathbb{R}^d}\left[G(v)+(\Psi+\varepsilon\bar{\Theta}_v)v\right]\mathrm{d}x - \int_{\mathbb{R}^d}\left[G(\bar{v}_\varepsilon)+(\Psi+\varepsilon\bar{\Theta}_v)\bar{v}_\varepsilon\right]\mathrm{d}x\end{aligned} \tag{4.26}$$

for $u,v \in \mathcal{P}_2^r(\mathbb{R}^d)$. Notice that in both lines, the second integral is simply a normalization depending on ε but not on u or v. The equivalence of the first formula above to the definition of $\mathbf{L}_1$ in (4.7) is obtained by using, in that order, the fact that $\bar{u}_\varepsilon = 0$ on $\{\Phi \ge U_\varepsilon\}$, then the definition of d_F, next the first Euler–Lagrange equation from (1.24) in combination with the identity $\Phi - U_\varepsilon = (\Phi-U_\varepsilon)_+ - (U_\varepsilon-\Phi)_+$, which yields $F'(\bar{u}_\varepsilon)+\varepsilon\bar{\Theta}_u = (\Phi-U_\varepsilon)_+ - (\Phi-U_\varepsilon)$, and finally equality of mass of u and $\bar{u}_\varepsilon$. In this way, we find

$$\begin{aligned}\mathbf{L}_1(u) &= \int_{\mathbb{R}^d}\left[d_F(u|\bar{u}_\varepsilon)+(u-\bar{u}_\varepsilon)(\Phi-U_\varepsilon)_+\right]\mathrm{d}x\\ &= \int_{\mathbb{R}^d}\left[F(u)-F(\bar{u}_\varepsilon)+(u-\bar{u}_\varepsilon)\big((\Phi-U_\varepsilon)_+-F'(\bar{u}_\varepsilon)\big)\right]\mathrm{d}x\\ &= \int_{\mathbb{R}^d}\left[F(u)-F(\bar{u}_\varepsilon)+(u-\bar{u}_\varepsilon)\big(\Phi-U_\varepsilon+\varepsilon\bar{\Theta}_u\big)\right]\mathrm{d}x\\ &= \int_{\mathbb{R}^d}\left[F(u)-F(\bar{u}_\varepsilon)+(u-\bar{u}_\varepsilon)\big(\Phi+\varepsilon\bar{\Theta}_u\big)\right]\mathrm{d}x.\end{aligned}$$

The second formula in (4.26) is justified analogously. Note that in view of (4.26), the relation between $\mathbf{E}_\varepsilon$ and $\mathbf{L}$ is simply

$$\begin{aligned}&\mathbf{E}_\varepsilon(u,v)-\mathbf{E}_\varepsilon(\bar{u}_\varepsilon,\bar{v}_\varepsilon)\\ &\quad= \mathbf{L}(u,v)+\varepsilon\int_{\mathbb{R}^d}\left[h(u,v)-h(\bar{u}_\varepsilon,\bar{v}_\varepsilon)-\big(u-\bar{u}_\varepsilon\big)\bar{\Theta}_u-(v-\bar{v}_\varepsilon)\bar{\Theta}_v\right]\mathrm{d}x.\end{aligned} \tag{4.27}$$

For later reference, observe that thanks to the general hypotheses—in particular (1.15)—and to the ε-uniform estimates on the constants U_ε and V_ε from section 3, there is a constant C independent of $\varepsilon \in [0,\varepsilon^*]$ such that

$$\int_{\mathbb{R}^d} \Phi u \,\mathrm{d}x \le \mathbf{L}_1(u) + C, \quad \int_{\mathbb{R}^d} \Psi v \,\mathrm{d}x \le \mathbf{L}_2(v) + C, \tag{4.28}$$

and

$$\mathbf{L}(u,v) \le 2\mathbf{E}_\varepsilon(u,v) + C, \quad \mathbf{E}_\varepsilon(u,v) \le 2\mathbf{L}(u,v) + C. \tag{4.29}$$

Next, we introduce dissipation functionals that accompany $\mathbf{L}_1$ and $\mathbf{L}_2$:

$$\mathbf{D}_1(u) = \int_{\mathbb{R}^d} u \big|\nabla\big[F'(u) + \Phi + \varepsilon\bar\Theta_u\big]\big|^2 \,\mathrm{d}x, \quad \mathbf{D}_2(v) = \int_{\mathbb{R}^d} v \big|\nabla\big[G'(v) + \Psi + \varepsilon\bar\Theta_v\big]\big|^2 \,\mathrm{d}x. \tag{4.30}$$

In the language of subdifferential calculus in the L^2-Wasserstein metric (see, e.g., [3, Chapter 10]) and in view of the representation (4.26) above, one can characterize the functionals above as $\mathbf{D}_1 = |\partial\mathbf{L}_1|^2$ and $\mathbf{D}_2 = |\partial\mathbf{L}_2|^2$. Our method of proof does not require the full machinery of metric subdifferentials, but only the following consequence.

LEMMA 4.8. *There is a constant K_0 such that for all $\varepsilon \in [0,\varepsilon^*]$ with $K_0\varepsilon < \Lambda$, the functionals $\mathbf{L}_1$ and $\mathbf{L}_2$ are uniformly displacement convex of modulus $\Lambda - K_0\varepsilon$. In particular, for all $u, v \in \mathcal{P}_2^r(\mathbb{R}^d)$ with $\mathbf{D}_1(u) < \infty$ and $\mathbf{D}_2(v) < \infty$, there hold*

$$2(\Lambda - K_0\varepsilon)\mathbf{L}_1(u) \le \mathbf{D}_1(u), \quad 2(\Lambda - K_0\varepsilon)\mathbf{L}_2(v) \le \mathbf{D}_2(v). \tag{4.31}$$

Proof. By Corollary 3.9 (see (3.16)), the function $\bar\Theta_u = \partial_u h(\bar u_\varepsilon, \bar v_\varepsilon)$ is ε-uniformly semiconvex with some modulus $-K_0$. Recalling the Λ-uniform convexity of Φ, we conclude that the sum $\Phi + \varepsilon\bar\Theta_u$ is uniformly convex of modulus $\Lambda - K_0\varepsilon$ as long as $K_0\varepsilon < \Lambda$. By assumption, F satisfies McCann's condition (1.11). The result now follows from the general theory of displacement convexity; see Lemma 2.3. The argument for $\mathbf{L}_2$ is completely analogous. ∎

Notice that by the second condition in (4.1) for the choice of $\bar\varepsilon > 0$, we are able to bound for every $\varepsilon \in [0,\bar\varepsilon]$ the dissipation functionals $\mathbf{D}_1$ and $\mathbf{D}_2$ from below by some positive multiple of the relative entropy functionals $\mathbf{L}_1$ and $\mathbf{L}_2$, respectively.

4.5. An estimate by Bregman distances. The sole purpose of this section is to show Lemma 4.9 below, which becomes essential for the estimate of "garbage terms" in the proof of Proposition 4.4. The (technical) proof of the lemma heavily uses 2-degeneracy and 2-boundedness of (F,G,h).

LEMMA 4.9. *There is a constant κ independent of $\varepsilon \in [0,\varepsilon^*]$ such that the estimate*

$$\int_{\mathbb{R}^d} (u+v)\big[\omega\big(F'(u), G'(v)\big) - \omega\big(F'(\bar u_\varepsilon), G'(\bar v_\varepsilon)\big)\big]^2 \,\mathrm{d}x \le \kappa\mathbf{L}(u,v) \tag{4.32}$$

holds for all $(u,v) \in [\mathcal{P}_2^r(\mathbb{R}^d)]^2$ with $\mathbf{L}(u,v) < \infty$ and for $\omega : \mathbb{R}^2_{\ge 0} \to \mathbb{R}$ being any of the following four functions: $\partial_\rho\theta_u$, $\partial_\eta\theta_u$, $\partial_\rho\theta_v$, or $\partial_\eta\theta_v$.

Proof. We shall derive (4.32) as a consequence of a pointwise estimate on the integrand, namely

$$(u+v)\big[\omega\big(F'(u),G'(v)\big)-\omega\big(F'(\bar u_\varepsilon),G'(\bar v_\varepsilon)\big)\big]^2 \le \kappa\big(d_F(u|\bar u_\varepsilon)+d_G(v|\bar v_\varepsilon)\big), \tag{4.33}$$

where d_F and d_G were introduced in (4.6). In view of the definition of $\mathbf{L}$ in (4.7)–(4.8), an integration in x yields the claim (4.32).

We now prove (4.33), only with u instead of $(u+v)$ (as the case with v instead of $(u+v)$ is analogous). Let $\bar U$ be such that $\bar u_\varepsilon, \bar v_\varepsilon \le \bar U$ for all $\varepsilon \in [0,\varepsilon^*]$. We distinguish three cases.

Case 1*:* $u > 3\bar U$. By 2-boundedness and 2-degeneracy of (F,G,h), we have

$$u\big[\omega\big(F'(u),G'(v)\big)-\omega\big(F'(\bar u_\varepsilon),G'(\bar v_\varepsilon)\big)\big]^2 \le 4A^2 u;$$

see Remark 1.1 (3). The multiple of u on the right-hand side can be estimated by a multiple of $d_F(u|\bar u)$ as follows: thanks to convexity of F,

$$\begin{aligned} d_F(u|\bar u_\varepsilon) &\ge F(2\bar U)+(u-2\bar U)F'(2\bar U)-\big[F(\bar u_\varepsilon)+(u-\bar u_\varepsilon)F'(\bar u_\varepsilon)\big]\\ &\ge (u-2\bar U)F'(2\bar U)+(2\bar U-\bar u_\varepsilon)F'(\bar u_\varepsilon)-(u-\bar u_\varepsilon)F'(\bar u_\varepsilon)\\ &= (u-2\bar U)\big[F'(2\bar U)-F'(\bar u_\varepsilon)\big] \ge (u-2\bar U)\big[F'(2\bar U)-F'(\bar U)\big]. \end{aligned} \tag{4.34}$$

By strict convexity, $F'(2\bar U)-F'(\bar U)>0$, and with $u>3\bar U$, we clearly have $4A^2u \le \kappa d_F(u|\bar u_\varepsilon)$ for an appropriate constant κ independent of u and ε.

Case 2*:* $0 \le u \le 3\bar U$ and $v > 3\bar U$. Using again the global bound A on ω, we obtain

$$u\big[\omega\big(F'(u),G'(v)\big)-\omega\big(F'(\bar u_\varepsilon),G'(\bar v_\varepsilon)\big)\big]^2 \le 4A^2\bar U,$$

i.e., the left-hand side is bounded by an expression that is independent of u, v, and ε. Clearly, this expression is estimated by $\kappa d_G(v|\bar v_\varepsilon)$ with an appropriate κ independent of v; this follows in analogy to the estimate in the first case above.

Case 3*:* $0 \le u, v \le 3\bar U$. As a first step, we show that, for an appropriate constant L independent of ε, we have

$$[F'(u)-F'(\bar u_\varepsilon)]^2 \le L d_F(u|\bar u_\varepsilon) \tag{4.35}$$

and correspondingly

$$[G'(v)-G'(\bar v_\varepsilon)]^2 \le L d_G(v|\bar v_\varepsilon).$$

Considering both sides of (4.35) as functions in the variable u, we obviously have equality in the case $u=\bar u_\varepsilon$. Using the definition of d_F, we find that

$$\frac{\mathrm{d}}{\mathrm{d}u}[F'(u)-F'(\bar u_\varepsilon)]^2 = 2[F'(u)-F'(\bar u_\varepsilon)]F''(u) = 2F''(u)\frac{\mathrm{d}}{\mathrm{d}u}d_F(u|\bar u_\varepsilon)$$

for all $u>0$. Notice that this expression, by convexity of F, is negative for $u<\bar u_\varepsilon$ and positive for $u>\bar u_\varepsilon$. With the help of hypothesis (1.10) and the convexity of F, there exists a positive constant C such that

$$F''(r) \le C \quad \text{for all } r \in [0,3\bar U].$$

Hence, with the choice $L:=2C$, we conclude that the left-hand side of (4.35) decreases faster on $(0,\bar u_\varepsilon)$ and increases slower on $(\bar u_\varepsilon, 3\bar U)$ than the right-hand side. This suffices to have the validity of inequality (4.35) for all $u \in [0,3\bar U]$.

Next, define $\rho_s := sF'(u) + (1-s)F'(\bar{u}_\varepsilon)$ and $\eta_s := sG'(v) + (1-s)G'(\bar{v}_\varepsilon)$ for $s \in [0,1]$. Then

$$\begin{aligned}
&\big[\omega\big(F'(u), G'(v)\big) - \omega\big(F'(\bar{u}_\varepsilon), G'(\bar{v}_\varepsilon)\big)\big]^2 \\
&\quad \le \int_0^1 \big[\big(F'(u) - F'(\bar{u}_\varepsilon)\big)\partial_\rho\omega(\rho_s,\eta_s) + \big(G'(v) - G'(\bar{v}_\varepsilon)\big)\partial_\eta\omega(\rho_s,\eta_s)\big]^2\,\mathrm{d}s \\
&\quad \le 2\bigg(\sup_{0\le\mu,\nu\le 3\bar{U}} \big|\mathrm{D}\omega\big|\bigg)^2 \big[\big(F'(u) - F'(\bar{u}_\varepsilon)\big)^2 + \big(G'(v) - G'(\bar{v}_\varepsilon)\big)^2\big].
\end{aligned}$$

The supremum above is a finite quantity B, thanks to 2-boundedness of (F, G, h). Therefore, with L from (4.35), we find

$$u\big[\omega\big(F'(u), G'(v)\big) - \omega\big(F'(\bar{u}_\varepsilon), G'(\bar{v}_\varepsilon)\big)\big]^2 \le 6\bar{U}BL\big(d_F(u|\bar{u}_\varepsilon) + d_G(v|\bar{v}_\varepsilon)\big),$$

proving the pointwise estimate (4.33) also in the final case. ∎

4.6. Proof of the core inequality. Finally, we prove Proposition 4.4. Again, the additional hypothesis (4.10) enters only indirectly via Lemma 4.5.

Proof of Proposition 4.4 *under the additional hypothesis* (4.10). Let $P_F(r) = rF'(r) - F(r)$ for $r \ge 0$. Recall that (φ_u, ψ_u) is the optimal pair of c-conjugate potentials for the transport from u^* to $\hat{u}$. Since $\mathbf{L}_1$ is displacement convex, the following "above tangent formula" holds; see, e.g., [41, Proposition 5.29 and Theorem 5.30],

$$\mathbf{L}_1(\hat{u}) - \mathbf{L}_1(u^*) \ge \int_{\mathbb{R}^d} P_F(u^*)\Delta^{\mathrm{ac}}\varphi_u\,\mathrm{d}x - \int_{\mathbb{R}^d} u^*\nabla[\Phi + \varepsilon\bar{\Theta}_u]\cdot\nabla\varphi_u\,\mathrm{d}x,$$

where $\Delta^{\mathrm{ac}}\varphi_u$ is the absolutely continuous part of the signed measure defined by the distributional Laplacian $\Delta\varphi_u$. Thanks to the regularity of u^*, we may re-write the first integral on the right-hand side using integration by parts. Indeed, observe that

$$\nabla P_F(u^*)\cdot\nabla\varphi_u = u^*\nabla F'(u^*)\cdot\nabla\varphi_u \in L^1(\mathbb{R}^d)$$

since $\nabla F'(u^*) \in L^2(\mathbb{R}^d; u^*\mathcal{L}^d)$ by Lemma 4.6 and $\nabla\varphi_u \in L^2(\mathbb{R}^d; u^*\mathcal{L}^d)$ in view of (2.5). Now, since $P_F \ge 0$ by convexity of F, and since $\Delta^{ac}\varphi_u \ge \Delta\varphi_u$ (as measures) because φ_u is semiconcave, we have

$$\mathbf{L}_1(\hat{u}) - \mathbf{L}_1(u^*) \ge -\int_{\mathbb{R}^d} u^*\nabla\big[F'(u^*) + \Phi + \varepsilon\bar{\Theta}_u\big]\cdot\nabla\varphi_u\,\mathrm{d}x = \tau Z_1(u^*, v^*), \tag{4.36}$$

where Z_1 can be made more explicit by substitution of the potential φ_u from (4.11):

$$Z_1(u,v) := \int_{\mathbb{R}^d} u\,\nabla\big[F'(u) + \Phi + \varepsilon\bar{\Theta}_u\big]\cdot\nabla\big[F'(u) + \Phi + \varepsilon\partial_u h(u,v)\big]\,\mathrm{d}x.$$

We estimate $Z_1(u,v)$ using a combination of the previously shown results. For brevity, we use—only in the calculations below—in addition to $\bar{\Theta}_u$ and $\bar{\Theta}_v$ introduced in (4.25) the notations

$$\begin{aligned}
\Theta_u &:= \theta_u\big(F'(u), G'(v)\big) = \partial_u h(u,v), \\
\Theta_{u,\rho} &:= \partial_\rho\theta_u\big(F'(u), G'(v)\big) = \frac{\partial_{uu}h(u,v)}{F''(u)},
\end{aligned}$$

and so on. First, the Cauchy–Schwarz inequality yields

$$\begin{aligned} Z_1(u,v) &= \int_{\mathbb{R}^d} u \left|\nabla\left[F'(u)+\Phi+\varepsilon\bar\Theta_u\right]\right|^2 \mathrm{d}x \\ &\quad + \varepsilon \int_{\mathbb{R}^d} u\,\nabla\left[F'(u)+\Phi+\varepsilon\bar\Theta_u\right]\cdot\nabla\left[\Theta_u-\bar\Theta_u\right]\mathrm{d}x \\ &\geq \left(1-\frac{\varepsilon}{2}\right)\mathbf{D}_1(u) - \frac{\varepsilon}{2}\int_{\mathbb{R}^d} u\left|\nabla\left[\Theta_u-\bar\Theta_u\right]\right|^2 \mathrm{d}x. \end{aligned}$$

Inside the last integral, we have

$$\begin{aligned} \nabla\left[\Theta_u-\bar\Theta_u\right] &= \Theta_{u,\rho}\nabla F'(u)+\Theta_{u,\eta}\nabla G'(v)-\bar\Theta_{u,\rho}\nabla F'(\bar u_\varepsilon)-\bar\Theta_{u,\eta}\nabla G'(\bar v_\varepsilon) \\ &= \Theta_{u,\rho}\nabla\left[F'(u)+\Phi+\varepsilon\bar\Theta_u\right]+\Theta_{u,\eta}\nabla\left[G'(v)+\Psi+\varepsilon\bar\Theta_v\right] \\ &\quad -\Theta_{u,\rho}\nabla\left[F'(\bar u_\varepsilon)+\Phi+\varepsilon\bar\Theta_u\right]-\Theta_{u,\eta}\nabla\left[G'(\bar v_\varepsilon)+\Psi+\varepsilon\bar\Theta_v\right] \\ &\quad +\left(\Theta_{u,\rho}-\bar\Theta_{u,\rho}\right)\nabla F'(\bar u_\varepsilon)+\left(\Theta_{u,\eta}-\bar\Theta_{u,\eta}\right)\nabla G'(\bar v_\varepsilon). \end{aligned}$$

The third and the fourth term above can be simplified using that by combination of the Euler–Lagrange system (1.24) with the identity $\Phi-U_\varepsilon=(\Phi-U_\varepsilon)_+-(U_\varepsilon-\Phi)_+$, one has

$$\nabla\left[F'(\bar u_\varepsilon)+\Phi+\varepsilon\bar\Theta_u\right]=-\nabla(\Phi-U_\varepsilon)_+,\quad \nabla\left[G'(\bar v_\varepsilon)+\Psi+\varepsilon\bar\Theta_v\right]=-\nabla(\Psi-V_\varepsilon)_+.$$

This yields

$$\begin{aligned} &\int_{\mathbb{R}^d} u\left|\nabla\left[\Theta_u-\bar\Theta_u\right]\right|^2\mathrm{d}x \\ &\leq 6\int_{\mathbb{R}^d} u\Theta_{u,\rho}^2\left|\nabla\left[F'(u)+\Phi+\varepsilon\bar\Theta_u\right]\right|^2\mathrm{d}x+6\int_{\mathbb{R}^d} u\Theta_{u,\eta}^2\left|\nabla\left[G'(v)+\Psi+\varepsilon\bar\Theta_v\right]\right|^2\mathrm{d}x \quad (4.37)\\ &\quad +6\int_{\mathbb{R}^d} u\Theta_{u,\rho}^2\left|\nabla(\Phi-U_\varepsilon)_+\right|^2\mathrm{d}x+6\int_{\mathbb{R}^d} u\Theta_{u,\eta}^2\left|\nabla(\Psi-V_\varepsilon)_+\right|^2\mathrm{d}x \quad (4.38)\\ &\quad +6\int_{\mathbb{R}^d} u\left(\Theta_{u,\rho}-\bar\Theta_{u,\rho}\right)^2|\nabla F'(\bar u_\varepsilon)|^2\mathrm{d}x+6\int_{\mathbb{R}^d} u\left(\Theta_{u,\eta}-\bar\Theta_{u,\eta}\right)^2|\nabla G'(\bar v_\varepsilon)|^2\mathrm{d}x. \quad (4.39) \end{aligned}$$

For further estimation, we observe that 2-boundedness and 2-degeneracy of (F,G,h) imply

$$|\Theta_{u,\rho}|\leq A\min\left\{1,F'(u),G'(v)\right\},\quad |\Theta_{u,\eta}|\leq A\min\left\{1,F'(u),G'(v)\right\}; \tag{4.40}$$

see (1.21). The first integral in (4.37) is now easily estimated using that thanks to (4.40),

$$\int_{\mathbb{R}^d} u\Theta_{u,\rho}^2\left|\nabla\left[F'(u)+\Phi+\varepsilon\bar\Theta_u\right]\right|^2\mathrm{d}x\leq A^2\mathbf{D}_1(u).$$

For estimation of the second integral in (4.37), we use instead that (F,G,h) satisfies the swap condition (1.17): thus $u\Theta_{u,\eta}^2\leq W^2 v$, and consequently,

$$\int_{\mathbb{R}^d} u\Theta_{v,\eta}^2\left|\nabla\left[G'(v)+\Psi+\varepsilon\bar\Theta_v\right]\right|^2\mathrm{d}x\leq W^2\mathbf{D}_2(v).$$

For estimation of the first integral in (4.38), two ingredients are needed. First, recall that $|\nabla\Phi|^2 \le \frac{2M^2}{\Lambda}\Phi$ by (1.19), and conclude that on $\{\Phi > U_\varepsilon\}$,

$$\big|\nabla(\Phi - U_\varepsilon)_+\big|^2 = |\nabla\Phi|^2 \le \frac{2M^2}{\Lambda}\Phi = \frac{2M^2}{\Lambda}U_\varepsilon + \frac{2M^2}{\Lambda}(\Phi - U_\varepsilon)_+ .$$

Second, we claim that there is a constant B such that

$$u\Theta_{u,\rho}^2 \le BF(u).$$

For $u \ge 1$ this is a trivial consequence of the convexity of F. For $0 < u < 1$, we use that in view of hypotheses (1.9) and (1.10), there are constants c_0 and C_0 such that $F'(u) \le C_0 u^{m-1}$ and $F(u) \ge c_0 u^m$ are satisfied. Therefore, employing also (4.40), we have

$$u\Theta_{u,\rho}^2 \le uA^2F'(u) \le \frac{C_0A^2}{c_0}F(u).$$

Now we combine these ingredients, recalling again (4.40) and bearing in mind that the integral is actually an integral on $\{\Phi > U_\varepsilon\}$ only, where $d_F(u|\bar u_\varepsilon) = F(u)$ thanks to (1.9):

$$\begin{aligned}\int_{\mathbb{R}^d} u\Theta_{u,\rho}^2\big|\nabla(\Phi - U_\varepsilon)_+\big|^2\,\mathrm{d}x &\le \frac{2M^2BU_\varepsilon}{\Lambda}\int_{\mathbb{R}^d} d_F(u|\bar u_\varepsilon)\,\mathrm{d}x + \frac{2M^2A^2}{\Lambda}\int_{\mathbb{R}^d} u(\Phi - U_\varepsilon)_+\,\mathrm{d}x\\ &\le \frac{2M^2}{\Lambda}\max\big\{BU_\varepsilon, A^2\big\}\mathbf{L}_1(u).\end{aligned}$$

The second integral in (4.38) is estimated in a completely analogous manner.

Finally, the integrals in (4.39) are both estimated by means of Lemma 4.9. We combine this with the boundedness of $|\nabla F'(\bar u_\varepsilon)|$ and $|\nabla G'(\bar v_\varepsilon)|$, respectively: by Proposition 3.5, we have that $F'(\bar u_\varepsilon), G'(\bar u_\varepsilon) \in W^{1,\infty}(\mathbb{R}^d)$, and that

$$|\nabla F'(\bar u_\varepsilon)| \le \tilde B, \quad |\nabla G'(\bar v_\varepsilon)| \le \tilde B$$

a.e. on $\mathbb{R}^d$, with $\tilde B$ independent of ε. We thus obtain

$$\int_{\mathbb{R}^d} u\big(\Theta_{u,\rho} - \bar\Theta_{u,\rho}\big)^2\big|\nabla F'(\bar u_\varepsilon)\big|^2\,\mathrm{d}x + \int_{\mathbb{R}^d} u\big(\Theta_{u,\eta} - \bar\Theta_{u,\eta}\big)^2\big|\nabla G'(\bar v_\varepsilon)\big|^2\,\mathrm{d}x \le 2\tilde B^2\kappa\mathbf{L}(u,v).$$

To summarize so far, we have shown that, with a suitable constant C,

$$Z_1(u,v) \ge \Big(1 - \frac{\varepsilon}{2}\big[1 + A^2\big]\Big)\,\mathbf{D}_1(u) - \frac{\varepsilon}{2}W^2\mathbf{D}_2(v) - \frac{\varepsilon}{2}C\mathbf{L}(u,v).$$

This finishes our estimate on Z_1. The pendant of (4.36) for v in place of u is

$$\mathbf{L}_2(\hat v) - \mathbf{L}_2(v^*) \ge Z_2(u^*, v^*)$$

with

$$Z_2(u,v) := \int_{\mathbb{R}^d} v\,\nabla\big[G'(u) + \Psi + \varepsilon\bar\Theta_v\big]\cdot\nabla\big[G'(v) + \Psi + \varepsilon\partial_v h(u,v)\big]\,\mathrm{d}x.$$

Estimating Z_2 in analogy to Z_1 as above leads to

$$Z_1(u,v) + Z_2(u,v) \ge \Big(1 - \frac{\varepsilon}{2}\big[1 + A^2 + W^2\big]\Big)\,(\mathbf{D}_1(u) + \mathbf{D}_2(v)) - \varepsilon C\mathbf{L}(u,v).$$

With an application of (4.31) and an appropriate choice of $K > 0$, the claim (4.9) has been shown. ∎

4.7. Removal of the additional hypothesis on the datum. Below, Propositions 4.2, 4.3 and 4.4 are proven without the additional hypothesis (4.10) on $(\hat{u},\hat{v})$.

Given an arbitrary $(\hat{u},\hat{v})\in[\mathcal{P}_2^r(\mathbb{R}^d)]^2$ with $\mathbf{E}_\varepsilon(\hat{u},\hat{v})<\infty$, let $(u^*,v^*)\in[\mathcal{P}_2^r(\mathbb{R}^d)]^2$ be the unique minimizer of the Yosida-regularized energy, according to Lemma 4.1. We consider a sequence of radii $R\to\infty$ and corresponding approximating pairs $(\hat{u}_R,\hat{v}_R)\in[\mathcal{P}_2^r(\mathbb{R}^d)]^2$ that satisfy (4.10) for the corresponding R, and are such that $\hat{u}_R$, $F(\hat{u}_R)$, and $|x|^2\hat{u}_R$ converge to $\hat{u}$, $F(\hat{u})$ and $|x|^2\hat{u}$ in $L^1(\mathbb{R}^d)$, respectively, and likewise for $\hat{v}_R$. An immediate consequence is

$$\mathbf{E}_\varepsilon(\hat{u}_R,\hat{v}_R)\to\mathbf{E}_\varepsilon(\hat{u},\hat{v}),\quad \mathbf{L}(\hat{u}_R,\hat{v}_R)\to\mathbf{L}(\hat{u},\hat{v}),\quad \widetilde{\mathbf{H}}(\hat{u}_R,\hat{v}_R)\to\widetilde{\mathbf{H}}(\hat{u},\hat{v}). \tag{4.41}$$

By Lemma 4.1, there is a unique minimizer (u_R^*,v_R^*) for each functional $\mathbf{E}_{\varepsilon,\tau}(\cdot|(\hat{u}_R,\hat{v}_R))$. Since the minimizers satisfy $\mathbf{E}_\varepsilon(u_R^*,v_R^*)\le\mathbf{E}_\varepsilon(\hat{u}_R,\hat{v}_R)$, and the right-hand side converges as $R\to\infty$, we have an R-uniform bound on second moments and L^2-norms and may conclude that (u_R^*,v_R^*) converges to some limit $(\check{u},\check{v})$ in $[\mathcal{P}_2^r(\mathbb{R}^d)]^2$, at least along a subsequence $R\to\infty$. We verify that $(\check{u},\check{v})$ is a minimizer of $\mathbf{E}_{\varepsilon,\tau}(\cdot|(\hat{u},\hat{v}))$: first, for any fixed pair $(u,v)\in[\mathcal{P}_2^r(\mathbb{R}^d)]^2$ with $\mathbf{E}_\varepsilon(u,v)<\infty$, it immediately follows that $\mathbf{E}_{\varepsilon,\tau}((u,v)|(\hat{u}_R,\hat{v}_R))\to\mathbf{E}_{\varepsilon,\tau}((u,v)|(\hat{u},\hat{v}))$. And second, by lower semicontinuity of $\mathbf{E}_\varepsilon$ and continuity of $\mathbf{W}_2$ with respect to convergence in $[\mathcal{P}_2^r(\mathbb{R}^d)]^2$, one obtains that

$$\mathbf{E}_{\varepsilon,\tau}\big((\check{u},\check{v})\big|(\hat{u},\hat{v})\big)\le\liminf_{R\to\infty}\mathbf{E}_{\varepsilon,\tau}\big((u_R^*,v_R^*)\big|(\hat{u}_R,\hat{v}_R)\big).$$

In combination, this proves the desired minimality of $(\check{u},\check{v})$. By uniqueness of the minimizer (see again Lemma 4.1), we may identify $(\check{u},\check{v})=(u^*,v^*)$.

Proof of Proposition 4.4. Since the pair $(\hat{u}_R,\hat{v}_R)$ satisfies the additional hypothesis (4.10), inequality (4.9) is valid for (u_R^*,v_R^*) and $(\hat{u}_R,\hat{v}_R)$ in place of (u^*,v^*) and $(\hat{u},\hat{v})$, i.e.,

$$\mathbf{L}(\hat{u}_R,\hat{v}_R)\ge\big[1+2\tau(\Lambda-K\varepsilon)\big]\mathbf{L}(u_R^*,v_R^*).$$

By (4.41), the left-hand side converges to $\mathbf{L}(\hat{u},\hat{v})$, while we use lower semicontinuity of $\mathbf{L}$ with respect to narrow convergence on the right-hand side. This yields (4.9), as desired. ∎

For the remaining proof, additional information on the convergence $(u_R^*,v_R^*)\to(u^*,v^*)$ is needed. Specifically,

$$F'(u_R^*)\to F'(u^*),\ G'(v_R^*)\to G'(v^*)\quad \text{weakly in } H^1(\mathbb{R}^d),\ \text{strongly in } L^2(\mathbb{R}^d). \tag{4.42}$$

To see this (just for the u-component), first observe that $F'(s)\le C(s+F(s))$ with some constant C, which is true for small and for large values of $s\ge0$, respectively, because of (1.9) and (1.20). This implies an R-uniform L^1-bound on $F'(u_R^*)$ since

$$\int_{\mathbb{R}^d}F'(u_R^*)\,\mathrm{d}x\le C\bigg(\int_{\mathbb{R}^d}u_R^*\,\mathrm{d}x+\int_{\mathbb{R}^d}F(u_R^*)\,\mathrm{d}x\bigg)\le C\big(1+2\mathbf{E}_\varepsilon(u_R^*,v_R^*)\big).$$

Further, estimate (4.5) holds with (u_R^*,v_R^*) and $(\hat{u}_R,\hat{v}_R)$ in place of (u^*,v^*) and $(\hat{u},\hat{v})$, respectively, since $(\hat{u}_R,\hat{v}_R)$ satisfies hypothesis (4.10), that is

$$\begin{aligned}\mathbf{E}_\varepsilon(u_R^*,v_R^*)+\widetilde{\mathbf{H}}(u_R^*,v_R^*)+\frac{\tau}{C}\int_{\mathbb{R}^d}\big[|\nabla F'(u_R^*)|^2+|\nabla G'(v_R^*)|^2\big]\,\mathrm{d}x\\ \le\tau+(1+\tau)\mathbf{E}_\varepsilon(\hat{u}_R,\hat{v}_R)+\widetilde{\mathbf{H}}(\hat{u}_R,\hat{v}_R).\end{aligned} \tag{4.43}$$

By (4.41), the terms $\mathbf{E}_\varepsilon(\hat{u}_R, \hat{v}_R)$ and $\widetilde{\mathbf{H}}(\hat{u}_R, \hat{v}_R)$ are R-uniformly bounded from above, and by Lemma B.1, $\widetilde{\mathbf{H}}(u_R^*, v_R^*)$ is R-uniformly bounded from below. Together, this implies an R-uniform bound on $\nabla F'(u_R^*)$ in $L^2(\mathbb{R}^d)$. By interpolation with the bound in $L^1(\mathbb{R}^d)$ above, we obtain R-uniform boundedness of $F'(u_R^*)$ in $H^1(\mathbb{R}^d)$. The claim now follows by Alaoglu's theorem, and by Rellich's theorem, bearing in mind that $F'(u_R^*)$ converges narrowly to $F'(u^*)$.

Proof of Proposition 4.3. By means of (4.41), we can pass to the limit on the right-hand side of (4.43) above. And by means of (4.42) as well as lower semicontinuity of $\mathbf{E}_\varepsilon$ and $\widetilde{\mathbf{H}}$ with respect to narrow convergence, we can pass to the limit also on the left-hand side. This gives (4.5) with datum $(\hat{u}, \hat{v})$. ◻

Proof of Proposition 4.2. Fix $\zeta \in C_c^\infty(\mathbb{R}^d)$. Since (4.3) holds under the hypothesis (4.10), we have

$$\int_{\mathbb{R}^d} \frac{u_R^* - \hat{u}_R}{\tau} \zeta \, \mathrm{d}x = \int_{\mathbb{R}^d} u_R^* \nabla \big[F'(u_R^*) + \varepsilon \partial_u h(u_R^*, v_R^*) + \Phi \big] \cdot \nabla \zeta \, \mathrm{d}x + R_u.$$

We can pass to the limit $R \to \infty$ on the left-hand side by narrow and L^1-convergence of u_R^* and $\hat{u}_R$, respectively. For the integral on the right-hand side, notice that $u_R^* \nabla \zeta \to u^* \nabla \zeta$ in $L^2(\mathbb{R}^d)$ and that $\nabla F'(u_R^*) \rightharpoonup \nabla F'(u^*)$, thanks to (4.42). To conclude that also $\nabla \partial_u h(u_R^*, v_R^*) \rightharpoonup \nabla \partial_u h(u^*, v^*)$ in $L^2(\mathbb{R}^d)$, observe that

$$\nabla \partial_u h(u_R^*, v_R^*) = \partial_\rho \theta_u \big(F'(u_R^*), G'(v_R^*) \big) \nabla F'(u_R^*) + \partial_\eta \theta_u \big(F'(u_R^*), G'(v_R^*) \big) \nabla G'(v_R^*).$$

For both products on the right-hand side, weak convergence in $L^2(\mathbb{R}^d)$ is easily concluded from weak convergence of $\nabla F'(u_R^*)$ and of $\nabla G'(v_R^*)$, and from convergence in measure of the bounded functions $\partial_\rho \theta_u(F'(u_R^*), G'(v_R^*))$ and $\partial_\eta \theta_u(F'(u_R^*), G'(v_R^*))$, again thanks to (4.42), and to the 2-boundedness of (F, G, h).

In a completely analogous way, we can pass to the limit $R \to \infty$ in the v-equation in (4.10). The bound on $|R_u| + |R_v|$ is preserved thanks to (4.41) and lower semicontinuity of $\mathbf{E}_\varepsilon$ and $\widetilde{\mathbf{H}}$ with respect to narrow convergence. ◻

5. Existence of weak solutions. The Yosida-regularized energy functional $\mathbf{E}_{\varepsilon,\tau}$ is now used to obtain a time-discrete approximation $(u_\tau^n, v_\tau^n)_{n \in \mathbb{N}_0}$ of the solution to (1.1) for given initial data $u(0) = u_0$, $v(0) = v_0$ with finite energy $\mathbf{E}_\varepsilon(u_0, v_0) < \infty$ by means of the *minimizing movement scheme*. Inductively, define $(u_\tau^0, v_\tau^0) := (u_0, v_0)$, and for each $n \in \mathbb{N}$ let (u_τ^n, v_τ^n) be the minimizer—which exists and is unique by Lemma 4.1—of the functional

$$[\mathcal{P}_2^r(\mathbb{R}^d)]^2 \ni (u, v) \mapsto \mathbf{E}_{\varepsilon,\tau}\big((u, v) \big| (u_\tau^{n-1}, v_\tau^{n-1}) \big).$$

Further, define the piecewise constant "interpolations" $\tilde{u}_\tau, \tilde{v}_\tau : [0, \infty) \to \mathcal{P}_2^r(\mathbb{R}^d)$ (depending of course on ε) in the usual way:

$$\tilde{u}_\tau(0) = u_0,\ \tilde{v}_\tau(0) = v_0, \quad \text{and} \quad \tilde{u}_\tau(t) = u_\tau^n,\ \tilde{v}_\tau(t) = v_\tau^n \quad \text{for } (n-1)\tau < t \le n\tau \text{ with } n \in \mathbb{N}.$$

The result of this section is the following convergence.

PROPOSITION 5.1. *For every $\varepsilon \in [0, \varepsilon^*]$, the interpolations $\tilde{u}_\tau$, $\tilde{v}_\tau$ converge, for a suitable sequence $\tau \downarrow 0$, to Hölder-continuous limit curves $u_*, v_* : [0, \infty) \to \mathcal{P}_2^r(\mathbb{R}^d)$, weakly in $L^1(\mathbb{R}^d)$ at every $t \ge 0$. Moreover, $F'(\tilde{u}_\tau), G'(\tilde{v}_\tau)$ converge to the respective limits $F'(u_*), G'(v_*)$, weakly in $L^2(0, T; H^1(\mathbb{R}^d))$ and strongly in $L^2((0, T) \times \mathbb{R}^d)$, for* *any $T>0$. Furthermore, the limits are weak solutions to* (1.1)*, i.e., they satisfy* (1.25) *for arbitrary test functions $\xi\in C_c^\infty((0,\infty)\times\mathbb{R}^d)$.*

With the solution $u_*,v_*:[0,\infty)\to\mathcal{P}_2^r(\mathbb{R}^d)$ from Proposition 5.1 we have shown the existence of a transient weak solution to the initial value problem for (1.1), as stated in Theorem 1.4.

5.1. Multistep estimates. We next prove several τ-*independent* estimates for $(\tilde u_\tau,\tilde v_\tau)$, which in the subsequent section then allows us to establish convergence for a sequence $\tau\downarrow 0$. We start by recalling the classical estimate that follows directly from the variational construction.

LEMMA 5.2. *For each $n\in\mathbb{N}$, we have*

$$\mathbf{E}_\varepsilon(u_\tau^n,v_\tau^n)+\frac{1}{2\tau}\big(\mathbf{W}_2(u_\tau^n,u_\tau^{n-1})^2+\mathbf{W}_2(v_\tau^n,v_\tau^{n-1})^2\big)\le\mathbf{E}_\varepsilon(u_\tau^{n-1},v_\tau^{n-1}) \tag{5.1}$$

and for each $N\in\mathbb{N}$

$$\mathbf{E}_\varepsilon(u_\tau^N,v_\tau^N)+\frac{1}{2\tau}\sum_{n=1}^{N}\big(\mathbf{W}_2(u_\tau^n,u_\tau^{n-1})^2+\mathbf{W}_2(v_\tau^n,v_\tau^{n-1})^2\big)\le\mathbf{E}_\varepsilon(u_0,v_0). \tag{5.2}$$

Proof. The first inequality (5.1) rephrases (4.2). Summing these inequalities for $n=1,2,\dots,N$, we then end up with the second inequality (5.2). ∎

The following three conclusions of Lemma 5.2 are important in the following:

- The values of $\mathbf{E}_\varepsilon(u_\tau^n,v_\tau^n)$ are monotonically decreasing in n, and in particular, bounded by $\mathbf{E}_\varepsilon(u_0,v_0)$. For $\varepsilon\in[0,\varepsilon^*]$, the hypothesis (1.15) then implies a uniform bound on $F(u_\tau^n)$ and $G(v_\tau^n)$ in $L^1(\mathbb{R}^d)$,

$$\int_{\mathbb{R}^d}\big[F(u_\tau^n)+G(v_\tau^n)\big]\,\mathrm{d}x\le 2\mathbf{E}_\varepsilon(u_0,v_0). \tag{5.3}$$

- Another consequence of energy monotonicity: for $\varepsilon\in[0,\varepsilon^*]$ we obtain, thanks to nonnegativity of H_ε, and to the lower bounds on Φ and Ψ by quadratic functions (see (1.18)), a uniform bound on the second moments of u_τ^n and v_τ^n,

$$\int_{\mathbb{R}^d}|x|^2\big(u_\tau^n+v_\tau^n\big)\,\mathrm{d}x\le 2|\underline{x}_\Phi|^2+2|\underline{x}_\Psi|^2+\frac{4}{\Lambda}\mathbf{E}_\varepsilon(u_0,v_0). \tag{5.4}$$

- By nonnegativity of $\mathbf{E}_\varepsilon$, one can pass for $\varepsilon\in[0,\varepsilon^*]$ to the limit $N\to\infty$ in (5.2) to obtain

$$\frac{1}{\tau}\sum_{n=1}^{\infty}\big(\mathbf{W}_2(u_\tau^n,u_\tau^{n-1})^2+\mathbf{W}_2(v_\tau^n,v_\tau^{n-1})^2\big)\le 2\mathbf{E}_\varepsilon(u_0,v_0). \tag{5.5}$$

  This gives rise to the following uniform estimate on the modulus of quasi-continuity.

LEMMA 5.3. *There is a τ-independent constant C such that for any $s,t\ge 0$,*

$$\mathbf{W}_2\big(\tilde u_\tau(t),\tilde u_\tau(s)\big)\le C\sqrt{|t-s|+\tau},\quad \mathbf{W}_2\big(\tilde v_\tau(t),\tilde v_\tau(s)\big)\le C\sqrt{|t-s|+\tau}. \tag{5.6}$$

Proof. Assume $0 \leq s < t$, and let $\underline{n}, \overline{n} \in \mathbb{N}_0$ be such that $(\underline{n}-1)\tau < s \leq \underline{n}\tau$ and $(\overline{n}-1)\tau < t \leq \overline{n}\tau$, i.e., $\tilde{u}_\tau(t) = u_\tau^{\overline{n}}$ and $\tilde{u}_\tau(s) = u_\tau^{\underline{n}}$, with $(\overline{n}-\underline{n})\tau \leq (t-s)+\tau$. If $\overline{n} = \underline{n}$, then (5.6) trivially holds. Otherwise, it follows from (5.5) via the triangle inequality for $\mathbf{W}_2$ and Hölder's inequality for sums that

$$\begin{aligned}\mathbf{W}_2\big(\tilde{u}_\tau(t), \tilde{u}_\tau(s)\big) &\leq \sum_{n=\underline{n}+1}^{\overline{n}} \mathbf{W}_2(u_\tau^n, u_\tau^{n-1}) \\ &\leq \left(\frac{1}{\tau}\sum_{n=1}^{\infty}\big(\mathbf{W}_2(u_\tau^n, u_\tau^{n-1})\big)^2\right)^{1/2}\left(\sum_{n=\underline{n}+1}^{\overline{n}}\tau\right)^{1/2} \\ &\leq \sqrt{2\mathbf{E}_\varepsilon(u_0,v_0)}\sqrt{(t-s)+\tau}.\end{aligned}$$

This proves the first inequality in (5.6), the second follows in the analogous way. ◻

LEMMA 5.4. *There is a τ-independent constant C such that, for each $T > 0$,*

$$\int_0^T \int_{\mathbb{R}^d} \big(|\nabla F'(\tilde{u}_\tau)|^2 + |\nabla G'(\tilde{v}_\tau)|^2\big)\,\mathrm{d}x \leq C(1 + T + \mathbf{E}_\varepsilon(u_0, v_0)). \tag{5.7}$$

Proof. Assume $T = N\tau$ for simplicity. Apply estimate (4.5) to $(\hat{u}, \hat{v}) = (u_\tau^{n-1}, v_\tau^{n-1})$ and $((u^*, v^*) = (u_\tau^n, v_\tau^n))$, and sum over $n = 1, \ldots, N$. This yields

$$\begin{aligned}&\tau\sum_{n=1}^{N}\int_{\mathbb{R}^d}\big[|\nabla F'(u_\tau^n)|^2 + |\nabla G'(v_\tau^n)|^2\big]\,\mathrm{d}x \\ &\quad\leq CN\tau\big(1+\mathbf{E}_\varepsilon(u_0,v_0)\big) + C\big[\mathbf{E}_\varepsilon(u_0,v_0) - \mathbf{E}_\varepsilon(u_\tau^N, v_\tau^N) + \widetilde{\mathbf{H}}(u_0,v_0) - \widetilde{\mathbf{H}}(u_\tau^N, v_\tau^N)\big].\end{aligned}$$

The left-hand side of this inequality coincides with the left-hand side of (5.7). On the right-hand side, first observe that $CN\tau = CT$, and that $\mathbf{E}_\varepsilon(u_\tau^N, v_\tau^N)$ is positive and thus negligible. To arrive at (5.7), it suffices to show that

$$-C\big(1+\mathbf{E}_\varepsilon(u,v)\big) \leq \widetilde{\mathbf{H}}(u,v) \leq C\big(1+\mathbf{E}_\varepsilon(u,v)\big).$$

The lower bound is easily obtained by combination of Lemma B.1 from Appendix A with the following estimate, that is a consequence of (1.18) and (4.28):

$$\frac{\Lambda}{2}\int_{\mathbb{R}^d}\big[|x-\underline{x}_\Phi|^2 u + |x-\underline{x}_\Psi|^2 v\big]\,\mathrm{d}x \leq \int_{\mathbb{R}^d}\big[\Phi u + \Psi v\big]\,\mathrm{d}x \leq C + \mathbf{H}_\varepsilon(u,v).$$

The control of $\widetilde{\mathbf{H}}$ from above is a simple consequence of $u\log u \leq C(1+F(u))$ and $v\log v \leq C(1+G(v))$ thanks to the at-least-quadratic growth of F and G, combined with (1.15). ◻

LEMMA 5.5. *Let $p, q > 1$ be such that*

$$(d-2)p < 2d \quad \text{and} \quad \frac{q}{p'} < 1 + \frac{2}{d}. \tag{5.8}$$

For each $T > 0$, there is a τ-independent constant C_T such that

$$\int_0^T \|F'(\tilde{u}_\tau)\|_{L^p}^q\,\mathrm{d}t \leq C_T. \tag{5.9}$$

Proof. Thanks to (5.8), we have that

$$\theta := \frac{2d}{d+2}\frac{1}{p'} < 1 \quad \text{and} \quad q\theta < 2.$$

Therefore, by the Gagliardo–Nirenberg interpolation inequality,

$$\int_0^T \|F'(\tilde{u}_\tau)\|_{L^p}^q \,\mathrm{d}t \le CT^{1-q\theta/2} \sup_{0<t<T} \|F'(\tilde{u}_\tau(t))\|_{L^1}^{q(1-\theta)} \left(\int_0^T \|\nabla F'(\tilde{u}_\tau)\|_{L^2}^2 \,\mathrm{d}t \right)^{q\theta/2}.$$

From (5.7) the τ-uniform boundedness of the term with $\nabla F'(\tilde{u}_\tau)$ follows. For the other term, we first observe that $F'(s) \le C(s + F(s))$ which for $s \ge 1$ is a direct consequence of (1.20), while for $s \le 1$, it is obtained from (1.9) in combination with the uniform boundedness of $F''(t)$ for $t \in (0,1]$. Since $\tilde{u}_\tau$ is of unit mass, and because of the uniform bound (5.3) on $F(\tilde{u}_\tau)$ in $L^1(\mathbb{R}^d)$, also $F'(\tilde{u}_\tau)$ is bounded in $L^1(\mathbb{R}^d)$, uniformly in $t \in [0,T]$ and in τ. ∎

5.2. Convergence proofs. The statements of Proposition 5.1 are proven in Lemmas 5.6, 5.7, and 5.8 below.

LEMMA 5.6. *For every $\varepsilon \in [0,\varepsilon^*]$, there are curves $u_*, v_* : [0,\infty) \to \mathcal{P}_2^r(\mathbb{R}^d)$, Hölder continuous in $\mathbf{W}_2$, such that, along a suitable sequence $\tau \downarrow 0$, the interpolations $\tilde{u}_\tau(t)$ and $\tilde{v}_\tau(t)$ converge to $u_*(t)$ and $v_*(t)$, respectively, weakly in $L^1(\mathbb{R}^d)$, at every $t \ge 0$.*

Proof. This lemma is a consequence of the generalized Arzelà–Ascoli theorem [3, Proposition 3.3.1]. Lemma 5.3 above provides a uniform modulus of (quasi-)continuity for $\tilde{u}_\tau$ and $\tilde{v}_\tau$ in $\mathbf{W}_2$; the topology induced by $\mathbf{W}_2$ is stronger than the narrow one. Further, the values of $\tilde{u}_\tau$ and $\tilde{v}_\tau$ belong to a narrowly compact set thanks to the uniform boundedness of second moments (5.4). The aforementioned proposition yields the narrow convergence along a sequence $\tau \downarrow 0$ of $\tilde{u}_\tau(t)$ and $\tilde{v}_\tau(t)$ to respective limits $u_*(t)$ and $v_*(t)$ for each $t \ge 0$, and u_*, v_* are continuous with respect to $\mathbf{W}_2$. By lower semicontinuity of $\mathbf{W}_2$ under narrow convergence, the estimate (5.6) is inherited by the limits u_*, v_* in the form

$$\mathbf{W}_2\big(u_*(t), u_*(s)\big) \le C|t-s|^{1/2}, \quad \mathbf{W}_2\big(v_*(t), v_*(s)\big) \le C|t-s|^{1/2},$$

which is the claimed Hölder continuity of u_*, v_* with respect to $\mathbf{W}_2$. Finally, the upgrade from narrow convergence of $\tilde{u}_\tau(t)$ and $\tilde{v}_\tau(t)$ to weak convergence in $L^1(\mathbb{R}^d)$ after passage to a suitable subsequence $\tau \downarrow 0$ is obtained from the boundedness of $F(\tilde{u}_\tau(t))$ and $G(\tilde{v}_\tau(t))$ in $L^1(\mathbb{R}^d)$; see (5.3). Indeed, since F and G are superlinear at infinity in view of (1.8), the Dunford–Pettis criterion applies. ∎

LEMMA 5.7. *For every $\varepsilon \in [0,\varepsilon^*]$ and every $T > 0$, we have along a suitable sequence $\tau \downarrow 0$,*

$$\tilde{u}_\tau \to u_* \quad \textit{strongly in } L^2([0,T]\times\mathbb{R}^d), \tag{5.10}$$

$$F'(\tilde{u}_\tau) \to F'(u_*) \quad \textit{strongly in } L^2_{loc}([0,T]\times\mathbb{R}^d), \tag{5.11}$$

$$\nabla F'(\tilde{u}_\tau) \rightharpoonup \nabla F'(u_*) \quad \textit{weakly in } L^2([0,T]\times\mathbb{R}^d). \tag{5.12}$$

Proof. For the proof of (5.10), we apply the generalized version [38, Theorem 2] of the Aubin–Lions lemma. On the Banach space $L^2(\mathbb{R}^d)$, define the *normal coercive integrand* by

$$\mathcal{F}(u) = \|F'(u)\|_{H^1}^2 + \int_{\mathbb{R}^d} |x|^2 u \,\mathrm{d}x,$$

and the compatible map $g\colon L^2(\mathbb{R}^d)\times L^2(\mathbb{R}^d)\to[0,\infty]$ by

$$g(u,u') := \mathbf{W}_2(u,u'),$$

with the conventions that $\mathcal{F}(u)=\infty$ unless $u\in\mathcal{P}_2^r(\mathbb{R}^d)$ with $F'(u)\in H^1(\mathbb{R}^d)$, and that $g(u,u')=\infty$ unless $u,u'\in\mathcal{P}_2^r(\mathbb{R}^d)$. Below, we verify lower semicontinuity of $\mathcal{F}$ and compactness of sublevel sets; we further show that

$$\int_0^{T-h} g\big(\tilde{u}_\tau(t+h),\tilde{u}_\tau(t)\big)\,\mathrm{d}t\to 0 \quad \text{uniformly in } \tau \text{ as } h\downarrow 0, \tag{5.13}$$

and that

$$\int_0^T \mathcal{F}(\tilde{u}_\tau(t))\,\mathrm{d}t\le C_T \tag{5.14}$$

with some constant C_T depending only on $T>0$. In the language of [38], the family $(\tilde{u}_\tau)_{\tau>0}$ is called *tight* because of the uniform bound (5.14) with the normal coercive integrand $\mathcal{F}$. In conclusion, [38, Theorem 2] yields that $\tilde{u}(t)$ converges in $L^2(\mathbb{R}^d)$, in measure with respect to $t\in(0,T)$, along a sequence $\tau\downarrow 0$. That limit necessarily coincides with u_* obtained in the proof of Lemma 5.6 above.

To prove lower semicontinuity of $\mathcal{F}$, consider a sequence (u_n) converging to u_* in $L^2(\mathbb{R}^d)$ such that $(\mathcal{F}(u_n))$ has a finite limit. Without loss of generality, we may assume that u_n even converges pointwise a.e. in $\mathbb{R}^d$. Fatou's lemma directly yields

$$\int_{\mathbb{R}^d}|x|^2u\,\mathrm{d}x\le\liminf_{n\to\infty}\int_{\mathbb{R}^d}|x|^2u_n\,\mathrm{d}x. \tag{5.15}$$

Further, by boundedness of the sequence $(F'(u_n))$ in the reflexive space $H^1(\mathbb{R}^d)$ and Rellich's theorem, a suitable subsequence converges to some limit ν_*, weakly in $H^1(\mathbb{R}^d)$, strongly in $L^2_{\mathrm{loc}}(\mathbb{R}^d)$, and pointwise a.e. in $\mathbb{R}^d$. By a.e. pointwise convergence of (u_n) and the continuity of F', we conclude $\nu_*=F'(u_*)$. Weak lower semicontinuity of norms implies

$$\|F'(u_*)\|_{H^1}^2=\|\nu_*\|_{H^1}^2\le\liminf_{n\to\infty}\|F'(u_n)\|_{H^1}^2,$$

and thus with (5.15) we conclude that

$$\mathcal{F}(u_*)\le\lim_{n\to\infty}\mathcal{F}(u_n).$$

Concerning compactness, consider the $\bar{\mathcal{F}}$-sublevel set S of $\mathcal{F}$. By boundedness of the H^1-norm of $F'(u)$ for all $u\in S$, there is a sequence (u_n) in S for which $F'(u_n)$ converges strongly in $L^2_{\mathrm{loc}}(\mathbb{R}^d)$. Since F' has at least linear growth, (u_n) itself also converges strongly in $L^2_{\mathrm{loc}}(\mathbb{R}^d)$. To show that this convergence is not just locally, observe that by interpolation there is a $p>2$ such that $F'(u)$ is uniformly bounded in $L^p(\mathbb{R}^d)$ for all u in the sublevel S. Using that $s^p\le C(s+F'(s)^p)$ for a suitable constant C thanks to the at-least-linear growth of F', it follows that there is a L^p-uniform bound for the $u_n\in S$ as well. Hölder's inequality yields

$$\int_{\mathbb{R}^d}|x|^{2(p-2)/(p-1)}u_n^2\,\mathrm{d}x\le\|u_n\|_{L^p}^{p/(p-1)}\left(\int_{\mathbb{R}^d}|x|^2u_n\,\mathrm{d}x\right)^{(p-2)/(p-1)},$$

and thus (u_n) is tight and converges in $L^2(\mathbb{R}^d)$. This proves compactness.

The property (5.13) follows from (5.6). More precisely, we have

$$\int_0^{T-h} g\big(\tilde{u}_\tau(t+h), \tilde{u}_\tau(t)\big)\,\mathrm{d}t \le CT\big(\sqrt{h}+\sqrt{\tau}\big).$$

Next, the estimate (5.14) is a consequence of the a priori bound (5.7) on $\nabla F'(\tilde{u}_\tau)$, estimate (5.9) with $p = q = 2$, and the moment control (5.4). By means of [38, Theorem 2], we obtain convergence of $\tilde{u}_\tau(t)$ to $u_*(t)$ in $L^2(\mathbb{R}^d)$, in measure with respect to $t \in [0,T]$, along a sequence $\tau \downarrow 0$. Estimate (5.9) with $2 = p < q < 2 + 4/d$ further yields a τ-uniform control on $\|\tilde{u}_\tau(t)\|_{L^2}$ in $L^q(0,T)$, since F' has at least linear growth, and thus uniform integrability to exponent two in time. This finishes the proof of (5.10).

Next, we show the convergence (5.11) of $(F'(\tilde{u}_\tau))$ to $F'(u_*)$ locally in $L^2([0,T] \times \mathbb{R}^d)$. By (5.10), we may assume without loss of generality that the chosen sequence $(\tilde{u}_\tau)$ converges pointwise to u_* a.e. on $[0,T] \times \mathbb{R}^d$. By continuity of F', the sequence $(F'(\tilde{u}_\tau))$ converges to $F'(u_*)$ pointwise almost everywhere. Moreover, estimate (5.9) with $2 < p = q < 2 + 2/d$ provides τ-uniform integrability of $F'(\tilde{u}_\tau)$ on $[0,T] \times \mathbb{R}^d$ to exponent two. Hence, by Vitali's convergence theorem, we get convergence of $(F'(\tilde{u}_\tau))$ to $F'(u_*)$ in $L^2_{loc}([0,T] \times \mathbb{R}^d)$.

It remains to verify (5.12), i.e., the weak convergence of $(\nabla F'(\tilde{u}_\tau))$ to $\nabla F'(u_*)$ in $L^2([0,T] \times \mathbb{R}^d)$. In fact, weak L^2-convergence to *some* limit $\nabla\zeta$ follows immediately from the boundedness (5.7) and the local convergence (5.11) via Alaoglu's theorem. Using once again the local convergence of $(F'(\tilde{u}_\tau))$ to $F'(u_*)$ we can identify the limit ζ as $F'(u_*)$. ∎

LEMMA 5.8. *The limits (u_*, v_*) obtained in Lemmas* 5.6 *and* 5.7 *satisfy the weak formulations* (1.25) *for every test function $\xi \in C_c^\infty((0,\infty) \times \mathbb{R}^d)$.*

Proof. By abuse of notation, τ will always denote an element of the sequence $\tau \downarrow 0$ along which $(\tilde{u}_\tau, \tilde{v}_\tau)$ converges to (u_*, v_*) in the sense of Lemmas 5.6 and 5.7.

Assume that the support of ξ lies in $(0,T) \times \Omega$, for some bounded open set $\Omega \subset \mathbb{R}^d$. For each $t \in (\tau, T)$, let n be such that $(n-1)\tau < t \le n\tau$, and use $\zeta :\,= \xi(t;\cdot)$ as test function in the first equation of (4.3) for that n. Integrate these equations in $t \in (\tau, T)$. The result can be written as

(5.16)
$$\begin{aligned}&\int_\tau^T \int_{\mathbb{R}^d} \frac{\tilde{u}_\tau(t) - \tilde{u}_\tau(t-\tau)}{\tau} \xi(t)\,\mathrm{d}x\,\mathrm{d}t \\ &\quad = \int_\tau^T \int_{\mathbb{R}^d} \tilde{u}_\tau(t) \nabla\big[F'(\tilde{u}_\tau(t)) + \Phi + \varepsilon\partial_u h(\tilde{u}_\tau(t), \tilde{v}_\tau(t))\big] \cdot \nabla\xi(t)\,\mathrm{d}x\,\mathrm{d}t + \int_\tau^T R_u(t)\,\mathrm{d}t,\end{aligned}$$

where, thanks to (4.4),

$$\begin{aligned}\left|\int_\tau^T R_u(t)\,\mathrm{d}t\right| &\le \tau \sum_{n=1}^\infty \sup_{0<t<T} \|\xi(t;\cdot)\|_{C^2} \big(\mathbf{E}_\varepsilon(u_\tau^{n-1}, v_\tau^{n-1}) - \mathbf{E}_\varepsilon(u_\tau^n, v_\tau^n)\big) \\ &\le \tau \sup_{0<t<T} \|\xi(t;\cdot)\|_{C^2} \mathbf{E}_\varepsilon(u_0, v_0),\end{aligned}$$

and so $\int_0^T R_u(t)\,\mathrm{d}t \to 0$ as $\tau \downarrow 0$. For the integral on the left-hand side of (5.16), we obtain

$$\int_\tau^T \int_{\mathbb{R}^d} \frac{\tilde{u}_\tau(t) - \tilde{u}_\tau(t-\tau)}{\tau} \xi(t)\,\mathrm{d}x\,\mathrm{d}t = \int_0^T \int_{\mathbb{R}^d} \tilde{u}_\tau(t) \frac{\xi(t+\tau) - \xi(t)}{\tau}\,\mathrm{d}x\,\mathrm{d}t$$

for $\tau > 0$ sufficiently small (recall that the support of ξ is contained in $(0,T)\times\Omega$), which, in turn, implies

$$\int_\tau^T\int_{\mathbb{R}^d}\frac{\tilde u_\tau(t)-\tilde u_\tau(t-\tau)}{\tau}\xi(t)\,\mathrm{d}x\,\mathrm{d}t \to \int_0^T\int_{\mathbb{R}^d}u_*(t)\partial_t\xi(t)\,\mathrm{d}x\,\mathrm{d}t,$$

thanks to the convergence of $\tilde u_\tau$ to u_* in $L^1((0,T)\times\Omega)$, and the uniform convergence of difference quotients of ξ.

It remains to verify the convergence of the integral on the right-hand side of (5.16). By the strong L^2-convergence (5.10) of $\tilde u_\tau$ to u_*, and since $\nabla\xi$ is smooth and has compact support inside $(0,T)\times\Omega$, it suffices to verify weak convergence of

$$\nabla\big[F'(\tilde u_\tau)+\varepsilon\partial_u h(\tilde u_\tau,\tilde v_\tau)\big] \rightharpoonup \nabla\big[F'(u_*)+\varepsilon\partial_u h(u_*,v_*)\big]$$

in $L^2((0,T)\times\Omega)$. But this is clear: on the one hand, $F'(\tilde u_\tau)$ and $G'(\tilde v_\tau)$ converge weakly in $L^2(0,T;H^1(\mathbb{R}^d))$; see (5.12). On the other hand, recalling the 2-boundedness of (F,G,h), the local L^2-convergence (5.11) of $F'(\tilde u_\tau)$ and $G'(\tilde v_\tau)$ implies convergence in measure of $\partial_\rho\theta_u(F'(\tilde u_\tau),G'(\tilde v_\tau))$ and of $\partial_\eta\theta_u(F'(\tilde u_\tau),G'(\tilde v_\tau))$. By boundedness and continuity of the derivatives of θ_u, the weak convergence of $\nabla F'(\tilde u_\tau)$ and $\nabla G'(\tilde v_\tau)$ in $[L^2((0,T)\times\mathbb{R}^d)]^d$ is inherited by

$$\nabla\partial_u h(\tilde u_\tau,\tilde v_\tau) = \partial_\rho\theta_u\big(F'(\tilde u_\tau),G'(\tilde v_\tau)\big)\nabla F'(\tilde u_\tau)+\partial_\eta\theta_u\big(F'(\tilde u_\tau),G'(\tilde v_\tau)\big)\nabla G'(\tilde v_\tau). \qquad \square$$

6. Convergence to equilibrium. In preparation of the proof of Theorem 1.5, we provide an adapted version of the Csiszar–Kullback inequality for $\mathbf{L}$.

LEMMA 6.1. *There is a constant C, independent of $\varepsilon\in[0,\bar\varepsilon]$, such that for all $u,v\in\mathcal{P}_2^r(\mathbb{R}^d)$ with $\mathbf{L}_1(u)<\infty$ and $\mathbf{L}_2(v)<\infty$, there hold*

$$\|u-\bar u_\varepsilon\|_{L^1}^2 \le C\mathbf{L}_1(u), \quad \|v-\bar v_\varepsilon\|_{L^1}^2\le C\mathbf{L}_2(v). \tag{6.1}$$

Proof. It suffices to prove the first inequality in (6.1). The point of departure is that both u and $\bar u_\varepsilon$ have unit mass, and therefore,

$$\|u-\bar u_\varepsilon\|_{L^1} = 2\int_{\{u<\bar u_\varepsilon\}}(\bar u_\varepsilon-u)\,\mathrm{d}x. \tag{6.2}$$

It is thus sufficient to estimate the integral of $\bar u_\varepsilon - u$ on $\{u<\bar u_\varepsilon\}$, which is a subset of Ω^u_ε. Let $\bar U$ be an upper bound on $\bar u_\varepsilon$, uniformly in $\varepsilon\in[0,\bar\varepsilon]$; see Proposition 3.2. By hypothesis (1.10) there is a constant $c_0>0$ such that $F''(r)\ge c_0 r^{m-2}$ for all $r\le\bar U$, and thus we have that

$$\begin{aligned} d_F(u|\bar u_\varepsilon) &= \int_u^{\bar u_\varepsilon}(r-u)F''(r)\,\mathrm{d}r \ge c_0\int_{\frac{u+\bar u_\varepsilon}{2}}^{\bar u_\varepsilon}(r-u)r^{m-2}\,\mathrm{d}r\\ &\ge \frac{c_0}{2^{m-2}}\bar u_\varepsilon^{m-2}\int_{\frac{u+\bar u_\varepsilon}{2}}^{\bar u_\varepsilon}(r-u)\,\mathrm{d}r\\ &= \frac{3c_0}{2^{m+1}}\bar u_\varepsilon^{m-2}(u-\bar u_\varepsilon)^2. \end{aligned}$$

This implies, by means of the Cauchy–Schwarz inequality, that

$$\begin{aligned} \int_{\{u<\bar u_\varepsilon\}}(u-\bar u_\varepsilon)\,\mathrm{d}x &\le \left(\int_{\{u<\bar u_\varepsilon\}}\bar u_\varepsilon^{-(m-2)}\,\mathrm{d}x\right)^{1/2}\left(\int_{\{u<\bar u_\varepsilon\}}\bar u_\varepsilon^{m-2}(u-\bar u_\varepsilon)^2\,\mathrm{d}x\right)^{1/2}\\ &\le \sqrt{2^{m+1}/(3c_0)}\left(\int_{\Omega^u_\varepsilon}\bar u_\varepsilon^{-(m-2)}\,\mathrm{d}x\right)^{1/2}\left(\int_{\mathbb{R}^d}d_F(u|\bar u_\varepsilon)\,\mathrm{d}x\right)^{1/2}. \end{aligned} \tag{6.3}$$

It remains to be shown that the integral of $\bar{u}_\varepsilon^{-(m-2)}$ over Ω_ε^u is finite. For the estimation of the integrand, we obtain thanks to (1.21),

$$\partial_u h(\bar{u}_\varepsilon, \bar{v}_\varepsilon) = \theta_u\big(F'(\bar{u}_\varepsilon), G'(\bar{v}_\varepsilon)\big) \leq AF'(\bar{u}_\varepsilon),$$

and therefore, the first Euler–Lagrange equation in (1.24) implies that

$$(1 + A\bar{\varepsilon})F'(\bar{u}_\varepsilon) \geq (U_\varepsilon - \Phi)_+ .$$

Using further that $F'(\bar{u}_\varepsilon) \leq K\bar{u}_\varepsilon^{m-1}$, again thanks to (1.9) and (1.10), we conclude that on Ω_ε^u,

$$\bar{u}_\varepsilon \geq c(U_\varepsilon - \Phi)^{1/(m-1)} \quad \text{with} \quad c := \big(K(1 + A\bar{\varepsilon})\big)^{-1/(m-1)}.$$

We can now estimate the integral of $\bar{u}_\varepsilon^{-(m-2)}$ by means of the coarea formula. Two observations: first, $|\nabla\sqrt{\Phi}| \geq \sqrt{2\Lambda/M}$ by (1.19), and second, the diameter of Ω_ε^u is bounded uniformly in $\varepsilon \in [0, \bar{\varepsilon}]$; see Proposition 3.2. Hence, the $(d-1)$-dimensional Hausdorff measures $\mathcal{H}^{d-1}$ of the surfaces of the convex sets $\{\Phi < r^2\}$ are uniformly bounded by some S for every r with $r^2 \leq U_\varepsilon$. Now we estimate

$$\begin{aligned}
\int_{\Omega_\varepsilon^u} \bar{u}_\varepsilon^{-(m-2)} \,\mathrm{d}x &\leq \frac{M^{1/2}}{(2\Lambda)^{1/2} c^{m-2}} \int_{\{\sqrt{\Phi} < \sqrt{U_\varepsilon}\}} (U_\varepsilon - \Phi)^{-(m-2)/(m-1)} |\nabla\sqrt{\Phi}| \,\mathrm{d}x \\
&\leq \frac{M^{1/2}}{(2\Lambda)^{1/2} c^{m-2}} \int_0^{\sqrt{U_\varepsilon}} (U_\varepsilon - r^2)^{-(m-2)/(m-1)} \mathcal{H}^{d-1}\big(\partial\big\{\sqrt{\Phi} < r\big\}\big) \,\mathrm{d}r \\
&\leq \frac{M^{1/2} S}{(2\Lambda)^{1/2} c^{m-2}} \int_0^{\sqrt{U_\varepsilon}} (U_\varepsilon - r^2)^{-(m-2)/(m-1)} \,\mathrm{d}r,
\end{aligned}$$

and this integral has a finite value since $(m-2)/(m-1) < 1$, which is bounded independently of $\varepsilon \in [0, \bar{\varepsilon}]$. Combining this with (6.3) and (6.2), we obtain

$$\|u - \bar{u}_\varepsilon\|_{L^1} \leq C\bigg(\int_{\mathbb{R}^d} d_F(u|\bar{u}_\varepsilon) \,\mathrm{d}x\bigg)^{1/2}$$

for some constant C, which is uniform in $\varepsilon \in [0, \bar{\varepsilon}]$. With the definition of $\mathbf{L}_1(u)$ from (4.7), the proof of the first claim in (6.1) is complete. ∎

Proof of Theorem 1.5. Apply Proposition 4.4 with $(u^*, v^*) = (u_\tau^n, v_\tau^n)$ and $(\hat{u}, \hat{v}) = (u_\tau^{n-1}, v_\tau^{n-1})$, i.e.,

$$\mathbf{L}(u_\tau^{n-1}, v_\tau^{n-1}) \geq \big(1 + 2\Lambda_\varepsilon\tau\big)\mathbf{L}(u_\tau^n, v_\tau^n)$$

with $\Lambda_\varepsilon = \Lambda - K\varepsilon$, and then, after iteration on $n = 1, 2, \ldots, N$ and for τ sufficiently small,

$$\mathbf{L}(u_\tau^n, v_\tau^n) \leq \big(1 + 2\Lambda_\varepsilon\tau\big)^{-n}\mathbf{L}(u_0, v_0).$$

Since $\mathbf{L}$ is a convex functional and thus lower semicontinuous with respect to convergence in $\mathbf{W}_2$, it follows in the limit $\tau \downarrow 0$ for the limiting curve from Theorem 1.4 that

$$\mathbf{L}(u_t, v_t) \leq \exp\big(-2\Lambda_\varepsilon t\big)\mathbf{L}(u_0, v_0).$$

The $\mathbf{L}(u_0, v_0)$ on the right-hand side is easily estimated in terms of $\mathbf{E}_\varepsilon(u_0, v_0)$; see (4.29). Thanks to (6.1), the left-hand side controls the L^1-norms of $u_t - \bar{u}_\varepsilon$ and $v_t - \bar{v}_\varepsilon$. ∎

Appendix A. Verification of Example 1.2. The properties (1.8)–(1.12) for the nonlinearities F and G are immediately checked. With

$$\partial_u \tilde{u} = u^{-1}\tilde{u}(1-\tilde{u}), \quad \partial_v \tilde{u} = v^{-1}\tilde{u}\tilde{v}, \tag{A.1}$$

and similar formulas for $\partial_u \tilde{v}$, $\partial_v \tilde{v}$, we find for $i,j = 0,1,2,\dots$, that

$$\partial_u^i \partial_v^j h(u,v) = u^{p-i} v^{q-j} \tilde{u}^\alpha \tilde{v}^\beta P(\tilde{u},\tilde{v}),$$

where here and below, P is some polynomial in two variables that may change from line to line. For later reference, note that an expression of the form

$$u^{p'} v^{q'} \tilde{u}^\alpha \tilde{v}^\beta P(\tilde{u},\tilde{v}) = \tilde{u}^{p'+\alpha} \tilde{v}^{q'+\beta} (1+u+v)^{p'+q'} P(\tilde{u},\tilde{v}) \tag{A.2}$$

with arbitrary $p', q' \in \mathbb{R}$ is uniformly bounded for $(u,v) \in \mathbb{R}^2_{>0}$ if

$$p' + q' \le 0 \quad \text{and} \quad p' + \alpha \ge 0,\ q' + \beta \ge 0, \tag{A.3}$$

and vanishes for $(u,v) \in \partial\mathbb{R}^2_{\ge 0}$ if

$$p' + \alpha > 0,\ q' + \beta > 0. \tag{A.4}$$

We verify the hypotheses for the coupling h. Concerning (1.13), criterion (A.4) yields the vanishing of h on $\partial\mathbb{R}^2_{\ge 0}$ since $\alpha > -p$ and $\beta > -q$ thanks to hypotheses (1.22), and similarly the vanishing of $\partial_u h$ and of $\partial_v h$ follows from $\alpha > 1-p$ and $\beta > 1-q$, respectively.

Concerning the convexity condition (1.14), we denote by $W(u,v)$ the Hessian of $F(u)+G(v)$—that is, the diagonal matrix with entries $(m-1)u^{m-2}$ and $(n-1)v^{n-2}$—and then we need to verify that $W(u,v) + 2\varepsilon^* \nabla^2 h(u,v) \ge 0$ for a small $\varepsilon^* > 0$. A sufficient criterion is the uniform boundedness of the matrix

$$W(u,v)^{-1/2} \nabla^2 h(u,v) W(u,v)^{-1/2} = \begin{pmatrix} \frac{\partial_{uu} h(u,v)}{(m-1)u^{m-2}} & \frac{\partial_{uv} h(u,v)}{\sqrt{(m-1)(n-1)u^{m-2}v^{n-2}}} \\ \frac{\partial_{uv} h(u,v)}{\sqrt{(m-1)(n-1)u^{m-2}v^{n-2}}} & \frac{\partial_{vv} h(u,v)}{(n-1)u^{n-2}} \end{pmatrix}$$

in $u,v \in \mathbb{R}_{>0}$. The top-left entry of this matrix has the form

$$\frac{\partial_{uu} h(u,v)}{(m-1)u^{m-2}} = u^{-(m-p)} v^q \tilde{u}^\alpha \tilde{v}^\beta P(\tilde{u},\tilde{v}).$$

Criterion (A.3) for boundedness is satisfied since $p+q \le m$ and $\alpha \ge m-p$, $\beta \ge -q$ by hypotheses (1.22). The corresponding conditions for the bottom right entry are $p+q \le n$ and $\alpha \ge -p$, $\beta \ge n-q$. For the off-diagonal entries, we find

$$\frac{\partial_{uv} h(u,v)}{\sqrt{(m-1)(n-1)u^{m-2}v^{n-2}}} = u^{p-m/2} v^{q-n/2} \tilde{u}^\alpha \tilde{v}^\beta P(\tilde{u},\tilde{v}),$$

and this is bounded since thanks to (1.22) we have $p+q \le (m+n)/2$, and $\alpha \ge m/2 - p$, $\beta \ge n/2 - q$.

For the verification of hypothesis (1.15), we show uniform boundedness of

$$\frac{h(u,v)}{F(u)+G(v)} \le \max(m,n)\ u^{p-m/2} \tilde{u}^\alpha v^{q-n/2} \tilde{v}^\beta$$

in $u, v \in \mathbb{R}_{>0}$. Indeed, according to (A.3), $p+q \le (m+n)/2$ and $\alpha \ge m/2-p, \beta \ge n/2-q$ suffice, which is again implied by hypotheses (1.22).

We turn to the swap condition (1.17) that we reformulate directly in terms of h as

$$\left|\frac{\sqrt{u}\partial_{uv}h(u,v)}{\sqrt{v}G''(v)}\right| \le W, \quad \left|\frac{\sqrt{v}\partial_{uv}h(u,v)}{\sqrt{u}F''(u)}\right| \le W.$$

For the first expression, we obtain

$$\frac{\sqrt{u}\partial_{uv}h(u,v)}{\sqrt{v}G''(v)} = u^{p-1/2}v^{q-n+1/2}\tilde{u}^\alpha \tilde{v}^\beta P(\tilde{u},\tilde{v}),$$

and this is bounded according to (A.3) since $p+q \le n$ and $\alpha \ge 1/2-p$, $\beta \ge n-1/2-q$ hold by (1.22). Boundedness of the second term is shown in an analogous way.

To discuss k-boundedness and k-degeneracy, first note that since $F'(u) = u^{m-1}$ and $G'(v) = v^{n-1}$,

$$\theta_u(\rho,\eta) = \rho^{(p-1)/(m-1)}\eta^{q/(n-1)}\tilde{\rho}^\alpha\tilde{\eta}^\beta P(\tilde{\rho},\tilde{\eta}),$$
$$\theta_v(\rho,\eta) = \rho^{p/(m-1)}\eta^{(q-1)/(n-1)}\tilde{\rho}^\alpha\tilde{\eta}^\beta P(\tilde{\rho},\tilde{\eta}),$$

where the quotients

$$\tilde{\rho} = \frac{\rho^{1/(m-1)}}{1+\rho^{1/(m-1)}+\eta^{1/(n-1)}} \quad \text{and} \quad \tilde{\eta} = \frac{\eta^{1/(n-1)}}{1+\rho^{1/(m-1)}+\eta^{1/(n-1)}}$$

are still positive and globally bounded by 1. Combining (A.1) with the chain rule, we obtain for the partial derivatives

$$\partial_\rho\tilde{\rho} = \frac{\tilde{\rho}(1-\tilde{\rho})}{(m-1)\rho}, \quad \partial_\eta\tilde{\rho} = \frac{\tilde{\rho}\tilde{\eta}}{(n-1)\eta},$$

and similarly for $\partial_\rho\tilde{\rho}$, $\partial_\eta\tilde{\eta}$, so that we arrive at

$$\partial_\rho^i\partial_\eta^j\theta_u(\rho,\eta) = \rho^{(p-1)/(m-1)-i}\eta^{q/(n-1)-j}\tilde{\rho}^\alpha\tilde{\eta}^\beta P(\tilde{\rho},\tilde{\eta}),$$

and similarly for $\partial_\rho^i\partial_\eta^j\theta_v$. Applying (A.3), we find that all of θ_u's partial derivatives of total order $\ell \ge 1$ are bounded if

$$\frac{p-1}{m-1} + \frac{q}{n-1} \le \ell \quad \text{and} \quad \alpha + \frac{p-1}{m-1} \ge \ell, \ \beta + \frac{q}{n-1} \ge \ell.$$

It is easily checked that the first condition is implied by $p+q \le \min(m,n)$, for all $\ell \ge 1$, while the second condition follows—for all $\ell = 1, 2, \ldots, k$—from the hypotheses $\alpha > k-(p-1)/(m-1)$ and $\beta > k-(q-1)/(n-1)$. Actually, these hypotheses also imply vanishing of $\partial_\rho^i\partial_\eta^j\theta_u$ on $\partial\mathbb{R}^2_{\ge 0}$ via (A.4). The discussion of the partial derivatives of θ_v is analogous. We thus have verified k-boundedness and k-degeneracy of the triple (F, G, h) under the hypotheses (1.23).

Appendix B. A lower bound on the entropy. The following has been obtained, e.g., in [26]; we recall the proof for convenience.

Lemma B.1. *For any $u \in \mathcal{P}_2^r(\mathbb{R}^d)$, any $\beta > 0$, and any $\underline{x} \in \mathbb{R}^d$,*

$$\mathbf{H}^u(u) = \int u\log u \ge 1 - \left(\pi/\beta\right)^{d/2} - \beta\int_{\mathbb{R}^d}|x-\underline{x}|^2 u\,\mathrm{d}x \tag{B.1}$$

(with $z\log z$ interpreted as zero for $z=0$). In particular, $\mathbf{H}^u$ is nowhere $-\infty$ on $\mathcal{P}_2^r(\mathbb{R}^d)$.

Proof. By Legendre duality, $zv \leq z \log z - z + e^v$ for all $z \geq 0$ and $v \in \mathbb{R}$. With the choices $z := u(x)$ and $v := -\beta|x - \underline{x}|^2$ this gives

$$-\beta \int_{\mathbb{R}^d} |x - \underline{x}|^2 u \, \mathrm{d}x \leq \int_{\mathbb{R}^d} u \log u \, \mathrm{d}x - \int_{\mathbb{R}^d} u \, \mathrm{d}x + \int_{\mathbb{R}^d} e^{-\beta|x-\underline{x}|^2} \, \mathrm{d}x,$$

which is just (B.1). ∎

Acknowledgment. The authors thank Filippo Santambrogio for significant help on the rigorous justification of the Euler–Lagrange equations.

REFERENCES

[1] M. Agueh, *Existence of solutions to degenerate parabolic equations via the Monge-Kantorovich theory*, Adv. Differential Equations, 10 (2005), pp. 309–360.

[2] L. Alasio, H. Ranetbauer, M. Schmidtchen, and M.-T. Wolfram, *Trend to Equilibrium for Systems with Small Cross-Diffusion*, preprint, https://arxiv.org/abs/1906.08060, 2019.

[3] L. Ambrosio, N. Gigli, and G. Savaré, *Gradient Flows in Metric Spaces and in the Space of Probability Measures*, 2nd ed., Lectures in Mathematics ETH Zürich, Birkhäuser Verlag, Basel, 2008.

[4] A. B. T. Barbaro, N. Rodriguez, H. Yoldaş, and N. Zamponi, *Analysis of a Cross-Diffusion Model for Rival Gangs Interaction in a City*, preprint, https://arxiv.org/abs/2009.04189, 2020.

[5] J. Berendsen, M. Burger, and J.-F. Pietschmann, *On a cross-diffusion model for multiple species with nonlocal interaction and size exclusion*, Nonlinear Anal., 159 (2017), pp. 10–39.

[6] A. Blanchet, V. Calvez, and J. A. Carrillo, *Convergence of the mass-transport steepest descent scheme for the subcritical Patlak–Keller–Segel model*, SIAM J. Numer. Anal., 46 (2008), pp. 691–721, https://doi.org/10.1137/070683337.

[7] A. Blanchet, J. A. Carrillo, D. Kinderlehrer, M. Kowalczyk, P. Laurençot, and S. Lisini, *A hybrid variational principle for the Keller-Segel system in $\mathbb{R}^2$*, ESAIM Math. Model. Numer. Anal., 49 (2015), pp. 1553–1576.

[8] M. Bruna and S. J. Chapman, *Diffusion of multiple species with excluded-volume effects*, J. Chem. Phys., 137 (2012), 204116.

[9] M. Burger, M. Di Francesco, S. Fagioli, and A. Stevens, *Sorting phenomena in a mathematical model for two mutually attracting/repelling species*, SIAM J. Math. Anal., 50 (2018), pp. 3210–3250, https://doi.org/10.1137/17M1125716.

[10] M. Burger, M. Di Francesco, J.-F. Pietschmann, and B. Schlake, *Nonlinear cross-diffusion with size exclusion*, SIAM J. Math. Anal., 42 (2010), pp. 2842–2871, https://doi.org/10.1137/100783674.

[11] C. Cancès, T. O. Gallouët, and L. Monsaingeon, *Incompressible immiscible multiphase flows in porous media: A variational approach*, Anal. PDE, 10 (2017), pp. 1845–1876.

[12] J. A. Carrillo, M. Di Francesco, A. Figalli, T. Laurent, and D. Slepčev, *Global-in-time weak measure solutions and finite-time aggregation for nonlocal interaction equations*, Duke Math. J., 156 (2011), pp. 229–271.

[13] J. A. Carrillo, A. Jüngel, P. A. Markowich, G. Toscani, and A. Unterreiter, *Entropy dissipation methods for degenerate parabolic problems and generalized Sobolev inequalities*, Monatsh. Math., 133 (2001), pp. 1–82.

[14] J. A. Carrillo and G. Toscani, *Asymptotic L^1-decay of solutions of the porous medium equation to self-similarity*, Indiana Univ. Math. J., 49 (2000), pp. 113–142.

[15] J. A. Carrillo, R. J. McCann, and C. Villani, *Kinetic equilibration rates for granular media and related equations: Entropy dissipation and mass transportation estimates*, Rev. Mat. Iberoamericana, 19 (2003), pp. 971–1018.

[16] E. S. Daus, A. Jüngel, and B. Q. Tang, *Exponential time decay of solutions to reaction-cross-diffusion systems of Maxwell-Stefan type*, Arch. Ration. Mech. Anal., 235 (2020), pp. 1059–1104.

[17] E. S. Daus and B. Q. Tang, *Trend to equilibrium of renormalized solutions to reaction-cross-diffusion systems*, Appl. Math. Lett., 88 (2019), pp. 81–89.

[18] M. Del Pino and J. Dolbeault, *Best constants for Gagliardo-Nirenberg inequalities and applications to nonlinear diffusions*, J. Math. Pures Appl., 81 (2002), pp. 847–875.

[19] L. Desvillettes and K. Fellner, *Exponential decay toward equilibrium via entropy methods for reaction-diffusion equations*, J. Math. Anal. Appl., 319 (2006), pp. 157–176.

[20] M. Di Francesco, A. Esposito, and S. Fagioli, *Nonlinear degenerate cross-diffusion systems with nonlocal interaction*, Nonlinear Anal., 169 (2018), pp. 94–117.

[21] M. Di Francesco and S. Fagioli, *Measure solutions for non-local interaction PDEs with two species*, Nonlinearity, 26 (2013), pp. 2777–2808.

[22] M. Di Francesco, K. Fellner, and P. A. Markowich, *The entropy dissipation method for spatially inhomogeneous reaction-diffusion-type systems*, Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci., 464 (2008), pp. 3273–3300.

[23] K. Fellner, E. Latos, and B. Q. Tang, *Global regularity and convergence to equilibrium of reaction-diffusion systems with nonlinear diffusion*, J. Evol. Equ., 20 (2020), pp. 957–1003.

[24] U. Gianazza, G. Savaré, and G. Toscani, *The Wasserstein gradient flow of the Fisher information and the quantum drift-diffusion equation*, Arch. Ration. Mech. Anal., 194 (2009), pp. 133–220.

[25] J. Haskovec, S. Hittmeir, P. Markowich, and A. Mielke, *Decay to equilibrium for energy-reaction-diffusion systems*, SIAM J. Math. Anal., 50 (2018), pp. 1037–1075, https://doi.org/10.1137/16M1062065.

[26] R. Jordan, D. Kinderlehrer, and F. Otto, *The variational formulation of the Fokker–Planck equation*, SIAM J. Math. Anal., 29 (1998), pp. 1–17, https://doi.org/10.1137/S0036141096303359.

[27] A. Jüngel, *The boundedness-by-entropy method for cross-diffusion systems*, Nonlinearity, 28 (2015), pp. 1963–2001.

[28] D. Kinderlehrer, L. Monsaingeon, and X. Xu, *A Wasserstein gradient flow approach to Poisson-Nernst-Planck equations*, ESAIM Control Optim. Calc. Var., 23 (2017), pp. 137–164.

[29] P. Laurençot and B.-V. Matioc, *Self-similarity in a thin film Muskat problem*, SIAM J. Math. Anal., 49 (2017), pp. 2790–2842, https://doi.org/10.1137/16M1055335.

[30] M. Liero and A. Mielke, *Gradient structures and geodesic convexity for reaction-diffusion systems*, Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci., 371 (2013), 20120346.

[31] S. Lisini, D. Matthes, and G. Savaré, *Cahn-Hilliard and thin film equations with nonlinear mobility as gradient flows in weighted-Wasserstein metrics*, J. Differential Equations, 253 (2012), pp. 814–850.

[32] D. Matthes, R. J. McCann, and G. Savaré, *A family of nonlinear fourth order equations of gradient flow type*, Comm. Partial Differential Equations, 34 (2009), pp. 1352–1397.

[33] A. Mielke, *A gradient structure for reaction-diffusion systems and for energy-drift-diffusion systems*, Nonlinearity, 24 (2011), pp. 1329–1346.

[34] A. Mielke, J. Haskovec, and P. A. Markowich, *On uniform decay of the entropy for reaction-diffusion systems*, J. Dynam. Differential Equations, 27 (2015), pp. 897–928.

[35] R. Natalini, *The bipolar hydrodynamic model for semiconductors and the drift-diffusion equations*, J. Math. Anal. Appl., 198 (1996), pp. 262–281.

[36] F. Otto, *The geometry of dissipative evolution equations: The porous medium equation*, Comm. Partial Differential Equations, 26 (2001), pp. 101–174.

[37] R. Rossi, A. Mielke, and G. Savaré, *A metric approach to a class of doubly nonlinear evolution equations and applications*, Ann. Sc. Norm. Super. Pisa Cl. Sci., 7 (2008), pp. 97–169.

[38] R. Rossi and G. Savaré, *Tightness, integral equicontinuity and compactness for evolution problems in Banach spaces*, Ann. Sc. Norm. Super. Pisa Cl. Sci., 2 (2003), pp. 395–431.

[39] V. K. Vanag and I. R. Epstein, *Cross-diffusion and pattern formation in reaction-diffusion systems*, Phys. Chem. Chem. Phys., 11 (2009), pp. 897–912.

[40] F. Santambrogio, *Optimal Transport for Applied Mathematicians*, Prog. Nonlinear Differ. Equ. Appl. 87, Birkhäuser, Cham, 2015.

[41] C. Villani, *Topics in Optimal Transportation*, Grad. Stud. Math. 58, American Mathematical Society, Providence, RI, 2003.

[42] N. Zamponi and A. Jüngel, *Analysis of degenerate cross-diffusion population models with volume filling*, Ann. Inst. H. Poincaré Anal. Non Linéaire, 34 (2017), pp. 1–29.

[43] J. Zinsl, *Exponential convergence to equilibrium in a Poisson-Nernst-Planck-type system with nonlinear diffusion*, Discrete Contin. Dyn. Syst., 36 (2016), pp. 2915–2930.

[44] J. Zinsl and D. Matthes, *Exponential convergence to equilibrium in a coupled gradient flow system modeling chemotaxis*, Anal. PDE, 8 (2015), pp. 425–466.

[45] J. Zinsl and D. Matthes, *Transport distances and geodesic convexity for systems of degenerate diffusion equations*, Calc. Var. Partial Differential Equations, 54 (2015), pp. 3397–3438.

[46] M. Zizza, *Weak Solutions of a Two Coupled Fokker-Planck Equations and Its Large Time Behaviour by Means of Optimal Transport*, Master's thesis, Universitá di Pavia, Pavia, Italy, 2018/2019.