

# A PDE approach to Benamou–Brenier formula for the Schrödinger problem

M. Garatti<sup>\*</sup>, L. Nenna<sup>†‡</sup>, S. Rota Nodari<sup>§‡</sup>, L. Tamanini<sup>¶</sup>

Tuesday 28<sup>th</sup> April, 2026

We studied the Benamou–Brenier formulation of the Schrödinger problem, focusing on a gap between theoretical results and applications, that often involve measures with unbounded support. While the existing proof in the literature relies on the compactness of the marginals’ supports to ensure the necessary regularity of the Schrödinger potentials, we extend the validity of the Benamou–Brenier formula to the larger class of sub-Gaussian probability measures. Exploiting fine estimates on the Hessian of the potentials and the entropic interpolation, we provide an almost self-contained proof that establishes the existence of a velocity field with the appropriate polynomial growth that ensures the right integrability. This result justifies the use of the dynamic formulation in more general settings, such as Gaussian and mixture-of-Gaussians models, important also for the applications.

*Keywords.* Benamou–Brenier formula, Schrödinger problem.

*2020 Mathematics Subject Classification.* Primary: 49Q22; Secondary: 49N15, 94A17, 49K40.

## 1 Introduction

The *Schrödinger problem* originated in a series of papers by Erwin Schrödinger in 1931–1932 (see [Sch31] and [Sch32]), motivated by the following basic question: given the empirical distribution of a Brownian particle cloud at two distinct times, what is the most likely evolution connecting these two observations?

Following [FG97], [Léo14], this naturally leads to the following variational problem: given two probability measures  $\mu_0, \mu_1 \in \mathcal{P}(\mathbb{R}^n)$  and a parameter  $\varepsilon > 0$ , one aims to solve

$$\inf \{ H(\gamma \mid R_\varepsilon) : \gamma \in \Gamma(\mu_0, \mu_1) \},$$

---

<sup>\*</sup>Université Paris-Saclay, CNRS, Laboratoire de mathématiques d’Orsay, ParMA, Inria Saclay, 91405, Orsay, France. e-mail: [mattia.garatti@universite-paris-saclay.fr](mailto:mattia.garatti@universite-paris-saclay.fr)

<sup>†</sup>Université Paris-Saclay, CNRS, Laboratoire de mathématiques d’Orsay, ParMA, Inria Saclay, 91405, Orsay, France. e-mail: [luca.nenna@universite-paris-saclay.fr](mailto:luca.nenna@universite-paris-saclay.fr)

<sup>‡</sup>Institut Universitaire de France (IUF)

<sup>§</sup>Université Côte d’Azur, CNRS, Laboratoire J.A. Dieudonné, 06108 Nice, France. e-mail: [simona.rotanodari@univ-cotedazur.fr](mailto:simona.rotanodari@univ-cotedazur.fr)

<sup>¶</sup>Università Cattolica del Sacro Cuore, Dipartimento di Matematica e Fisica “Niccolò Tartaglia”, I-25133, Brescia, Italy. e-mail: [luca.tamanini@unicatt.it](mailto:luca.tamanini@unicatt.it)

where the reference measure  $R_\varepsilon = r_\varepsilon \mathcal{L}^n \otimes \mathcal{L}^n$  (see (1) for the definition of  $r_\varepsilon$ ) is the joint law at times  $t = 0$  and  $t = \varepsilon$  of the (reversible) Wiener measure,  $H(\cdot | \cdot)$  denotes the relative entropy, and  $\Gamma(\mu_0, \mu_1)$  is the set of transport plans with marginals  $\mu_0$  and  $\mu_1$ .

Although originally introduced in statistical mechanics, the Schrödinger problem has acquired renewed importance over the last two decades because of its deep connection with optimal transport. Indeed, as  $\varepsilon \rightarrow 0$ , the Schrödinger problem converges to the quadratic optimal transport problem, both at the level of costs and minimizers; for this reason, it can be interpreted as an entropic regularization of the quadratic Monge–Kantorovich problem.

This qualitative result was initially established by Mikami [Mik04] for the quadratic cost on  $\mathbb{R}^n$ , subsequently extended by Mikami–Thieullen [MT08] to more general cost functions and ultimately generalized by Léonard [Léo12] to Polish spaces and general diffusion processes. From a more quantitative point of view, a huge literature on convergence rates is available (see [CPT23; EMR15; Ada+13; CT21; Chi+20; Pal24; NP25; NPT25; EN24]). Understood the importance of the temperature parameter  $\varepsilon$ , it will be clear the decision of making it explicit throughout the paper. A key feature of the Schrödinger problem is that, under mild assumptions, its unique minimizer  $\gamma^\varepsilon$  admits the factorization

$$\gamma^\varepsilon = (f^\varepsilon \otimes g^\varepsilon)R_\varepsilon,$$

where the pair  $(f^\varepsilon, g^\varepsilon)$  solves the so-called Schrödinger system (see, for details, [Rüs95] and (2) below). This decomposition plays a major role both in theory and in computation. On the one hand, it provides the natural analogue of Kantorovich potentials in the entropic framework; on the other hand, it allows to apply powerful iterative scaling procedures such as the Sinkhorn algorithm: see, for details, from the very first contributions [Sin64], [Sin67] and [SK67] up to the more recent [Cut13], in which the entropic regularization and the Sinkhorn algorithm have really been introduced in the optimal transport theory as a computational tool, [Ben+15], in which they show the applicability of the algorithm also to fluid-dynamics problems and [GT25], in which they establish new Hessian stability results and convergence rates for Sinkhorn potentials, using the semiconcavity of the potentials to refine the analysis of the algorithm’s performance. In recent years, significant progress has been made in understanding the regularity, stability, and convexity properties of the corresponding Schrödinger potentials

$$\varphi_t^\varepsilon = \varepsilon \log f_t^\varepsilon, \quad \psi_t^\varepsilon = \varepsilon \log g_t^\varepsilon,$$

as well as their relation to Hamilton–Jacobi–Bellman and Fokker–Planck equations; see for instance [Chi+23], in which they provide fundamental gradient estimates for Schrödinger potentials, which rigorously hold up the stability and convergence behavior of the Sinkhorn approximations toward Monge’s problem; [Con24], in which he derives weak semiconvexity estimates for Schrödinger potentials, finding a direct link between these ones and the log-Sobolev inequality for Schrödinger bridges; and, again, [GT25].

The analogy with optimal transport becomes particularly striking at the dynamical level. In the quadratic transport setting, the celebrated Benamou–Brenier formula pro-

vides a dynamic reformulation of the Monge–Kantorovich problem:

$$\inf_{\gamma \in \Gamma(\mu_0, \mu_1)} \int |x - y|^2 d\gamma(x, y) = \inf \left\{ \int_0^1 \int_{\mathbb{R}^n} |v_t|^2 \eta_t d\mathcal{L}^n dt : \frac{\partial}{\partial t} \eta_t + \operatorname{div}(v_t \eta_t) = 0, \eta_0 = \mu_0, \eta_1 = \mu_1 \right\}.$$

First proved in [BB00], this formula is now one of the cornerstones of the theory of optimal transport; see, for example, [Vil08; San15; AGS05]. Its importance lies in the fact that it transforms a purely static minimization problem into a fluid-dynamical one, revealing the geometric structure of Wasserstein space and making possible a PDE-based approach to transport problems. Because the Schrödinger problem is an entropic counterpart of optimal transport, it is natural to ask whether a similar dynamical reformulation holds in this setting. In the presence of diffusion encoded by the Wiener measure, the continuity equation should be replaced by the Fokker–Planck equation

$$\frac{\partial}{\partial t} \eta_t - \frac{\varepsilon}{2} \Delta \eta_t + \operatorname{div}(v_t \eta_t) = 0.$$

One is then led to expect a Benamou–Brenier-type formula for the entropic cost, in which the relative entropy with respect to  $R_\varepsilon$  is represented as the minimum of a kinetic action under the Fokker–Planck constraint. Such a result was established by Gigli and Tamanini in [GT20] in the setting of  $\operatorname{RCD}^*(K, N)$  spaces and under compactness assumptions on the marginals, namely bounded densities with bounded support. We also refer to [Léo14; GLR17], where the equivalence between the two formulations with an only finite entropy assumption on the marginals is suggested via probabilistic arguments. However, the proof contains some flaws which can be fixed by using the same arguments as in [CL95]. Here we want to give an alternative, more explicit, and self-contained proof by only means of PDE tools.

However, many natural examples fall outside this compact framework. Gaussian measures, strongly log-concave measures, and various models used in applications have unbounded support. This is not merely a technical issue. In the Gaussian case, explicit computations are available and strongly suggest that the dynamic representation should remain valid beyond compactly supported marginals: see [MGM22]. More generally, probability distributions with sub-Gaussian tails arise naturally in stochastic analysis, Bayesian inference, diffusion models, and entropy-regularized transport between statistical distributions. Therefore, extending the Benamou–Brenier formulation to non-compact settings is both mathematically natural and relevant for applications.

The main difficulty is that the proof of [GT20] relies on regularity and integrability properties that are automatic under compact support but become delicate when the marginals are supported on the whole space. In particular, one needs to justify differentiation under the integral sign and repeated applications of Gauss–Green formulas for quantities involving the Schrödinger potentials and the entropic interpolation  $\varrho_t^\varepsilon$ . For

this, it is essential to control expressions such as

$$\psi_t^\varepsilon \rho_t^\varepsilon, \quad |\nabla \psi_t^\varepsilon|^2 \rho_t^\varepsilon, \quad \psi_t^\varepsilon \Delta \rho_t^\varepsilon,$$

and these controls are no longer immediate without compactness. Moreover, boundedness of the marginals is not inherited by the Schrödinger factors  $f^\varepsilon$  and  $g^\varepsilon$ : as already visible in explicit Gaussian examples, one may have exponential growth of  $g^\varepsilon$  at infinity even when the marginals themselves are smooth and rapidly decaying. This creates a genuine gap between the available abstract theory and the situations most relevant in practice.

The purpose of the present paper is to bridge this gap by identifying a sufficiently general class of marginals for which the Benamou–Brenier formula for the Schrödinger problem can still be proved rigorously. More precisely, we consider probability measures of the form

$$\mu_0 = e^{-V_0} \mathcal{L}^n, \quad \mu_1 = e^{-V_1} \mathcal{L}^n,$$

where  $V_0, V_1 \in C^2(\mathbb{R}^n)$  with uniform two-sided Hessian bounds. In particular, this framework includes Gaussian measures and, more generally, a broad class of strongly log-concave, sub-Gaussian distributions. The key feature of these assumptions is that they guarantee Gaussian-type decay at infinity while still allowing for unbounded support. Our strategy is to combine recent regularity estimates for Schrödinger potentials with careful moment and integrability estimates along the entropic interpolation. On the regularity side, we exploit the semiconvexity and semiconcavity bounds, obtained in [Con24] and [Chi+24], which imply at most quadratic growth for the potentials and linear growth for their gradients. On the analytic side, we prove that the entropic interpolation has sufficiently strong Gaussian moments, together with suitable bounds on its derivatives, to justify all the integrations by parts required in the proof. In this way, we obtain an almost self-contained derivation of the Benamou–Brenier formula in a non-compact setting that still covers the principal examples of interest. Practically speaking, the main contribution of this work can be summarized in finding sufficiently general assumptions on the marginals (the ones above) such that the following Theorem holds. Moreover, we also stress that we get a uniqueness result for the Benamou–Brenier formula under the Fokker–Planck constraint.

**Theorem 1.1** (Simplified version of Theorem 3.6). *Under suitable assumptions on  $\mu_0, \mu_1 \in \mathcal{P}(\mathbb{R}^n)$ , for every  $\varepsilon > 0$ , it holds*

$$\begin{aligned} \varepsilon \min_{\gamma \in \Gamma(\mu_0, \mu_1)} H(\gamma \mid \mathbf{R}_\varepsilon) &= \varepsilon H(\mu_0 \mid \mathcal{L}^n) \\ &+ \min \left\{ \int_0^1 \int \frac{|v_t|^2}{2} \eta_t d\mathcal{L}^n dt : \frac{\partial}{\partial t} \eta_t - \frac{\varepsilon}{2} \Delta \eta_t + \operatorname{div}(v_t \eta_t) = 0, \eta_0 \mathcal{L}^n = \mu_0, \eta_1 \mathcal{L}^n = \mu_1 \right\}, \end{aligned}$$

where the Fokker–Planck equation has to be understood in distributional sense.

**Outline of the contents** The paper is organized as follows: in Section 2, we collect a quick overview of the Schrödinger problem with the related theory and we establish the rigorous mathematical framework. In Section 3, we describe how we arrived at the choice of sub-Gaussian assumptions for the marginals and we state our main result with a sketch of the proof and a direct Corollary. Then, in Section 4, the first part is dedicated to prove all the technicalities and the second part to the rigorous proof of our main result, namely Theorem 3.6.

## 2 Preliminaries and setting

In this Section we aim at establishing the mathematical framework as well as introducing the Schrödinger problem and some of its properties which will be helpful throughout the paper.

**The Schrödinger problem** Consider, for a fixed  $\varepsilon > 0$ ,  $R_\varepsilon = r_\varepsilon \mathcal{L}^n \otimes \mathcal{L}^n$ , where

$$(1) \quad r_s(x, y) = \frac{1}{\sqrt{(2\pi s)^n}} e^{-\frac{|x-y|^2}{2s}}$$

is the heat kernel in  $(x, y)$  at time  $s$ . Given  $\mu_0 = \varrho_0 \mathcal{L}^n, \mu_1 = \varrho_1 \mathcal{L}^n \in \mathcal{P}_2(\mathbb{R}^n)$ , namely the set of probability measures over  $\mathbb{R}^n$  with finite second moment, the *Schrödinger problem* reads as the minimization problem defined as

$$(SP_\varepsilon) \quad \inf \{H(\gamma | R_\varepsilon) : \gamma \in \Gamma(\mu_0, \mu_1)\},$$

where

$$H(Q | P) = \begin{cases} \int \log \left( \frac{dQ}{dP} \right) dQ & \text{if } Q \ll P, \\ +\infty & \text{otherwise} \end{cases}$$

is the entropy functional and

$$\Gamma(\mu_0, \mu_1) = \left\{ \gamma \in \mathcal{P}(\mathbb{R}^n \times \mathbb{R}^n) : (p^1)_\# \gamma = \mu_0, (p^2)_\# \gamma = \mu_1 \right\},$$

where  $p^1 : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}^n$ ,  $p^1(x, y) := x$ , and  $p^2 : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}^n$ ,  $p^2(x, y) := y$ , are the canonical projections and where we use the standard notation  $f_\# \gamma$  for any measurable map  $f : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}^n$  to mean the push-forward measure defined as  $f_\# \gamma(B) := \gamma(f^{-1}(B))$  for all Borel sets  $B \subseteq \mathbb{R}^n$ .

It is known (see [GT21, Proposition 2.1] for details, who build on the previous contributions [Léo01; BLN94; RT93]) that if  $H(\mu_0 \otimes \mu_1 | R_\varepsilon) < +\infty$ , then there exists a unique solution  $\gamma^\varepsilon$  to the above problem, also called *Schrödinger plan*. Moreover, there exist two Borel functions  $f^\varepsilon, g^\varepsilon : \mathbb{R}^n \rightarrow [0, +\infty[$ , a.e. uniquely determined up to the trivial transformation  $(f, g) \mapsto (cf, \frac{g}{c})$ ,  $c > 0$ , such that

$$\gamma^\varepsilon = (f^\varepsilon \otimes g^\varepsilon) R_\varepsilon,$$

where the couple  $(f^\varepsilon, g^\varepsilon)$  is the unique solution (see [Rüs95, Section 2] or [Nut21, Theorem 2.1] ) to the so-called *Schrödinger system*

$$(2) \quad \begin{cases} f^\varepsilon(x) \int g^\varepsilon(y) r_\varepsilon(x, y) dy = \varrho_0(x) & \text{a.e. in } \mathbb{R}^n, \\ g^\varepsilon(y) \int f^\varepsilon(x) r_\varepsilon(x, y) dx = \varrho_1(y) & \text{a.e. in } \mathbb{R}^n. \end{cases}$$

Notice that the above system simply force the  $\gamma^\varepsilon$  to satisfy the marginal constraints, that is  $(f^\varepsilon \otimes g^\varepsilon)R_\varepsilon \in \Gamma(\mu_0, \mu_1)$ .

**The Benamou–Brenier formula** In Optimal Transport, the Benamou–Brenier formula is a classical result that translates the static variational definition of Kantorovich into a fluid dynamical version that permits to recover the true meaning of this interpolation problem.

**Theorem 2.1.** *For every  $\mu_0, \mu_1 \in \mathcal{P}_2(\mathbb{R}^n)$ , it holds*

$$\begin{aligned} & \min_{\gamma \in \Gamma(\mu_0, \mu_1)} \int |x - y|^2 d\gamma(x, y) \\ & = \min \left\{ \int_0^1 \int |v_t|^2 \eta_t d\mathcal{L}^n dt : \frac{\partial}{\partial t} \eta_t + \operatorname{div}(v_t \eta_t) = 0, \eta_0 = \rho_0, \eta_1 = \rho_1 \right\}, \end{aligned}$$

where the continuity equation has to be understood in distributional sense.

As said, the previous Theorem has been firstly proved by [BB00] and now it is one of the most important results of the theory (see for instance [ABS21; Vil08; San15; AGS05] etc.). Under some compactness assumptions on the marginals, the same results holds true also for the Schrödinger problem [GT20], but the continuity equation constraint in the RHS is replaced by the Fokker–Planck equation. Let us then fix some notation to better comprehend this result. For every  $x \in \mathbb{R}^n$ , let us denote

$$(3a) \quad f_t^\varepsilon(x) = \begin{cases} f^\varepsilon(x) & \text{if } t = 0, \\ \int f^\varepsilon(y) r_{\varepsilon t}(x, y) dy & \text{if } 0 < t \leq 1, \end{cases}$$

$$(3b) \quad g_t^\varepsilon(x) = \begin{cases} \int g^\varepsilon(y) r_{\varepsilon(1-t)}(x, y) dy & \text{if } 0 \leq t < 1, \\ g^\varepsilon(x) & \text{if } t = 1, \end{cases}$$

namely the solutions of the heat equations

$$\begin{cases} \frac{\partial}{\partial t} f_t^\varepsilon - \frac{\varepsilon}{2} \Delta f_t^\varepsilon = 0 & \text{in } ]0, 1[ \times \mathbb{R}^n, \\ f_0^\varepsilon = f^\varepsilon & \text{on } \mathbb{R}^n, \end{cases} \quad \begin{cases} \frac{\partial}{\partial t} g_t^\varepsilon + \frac{\varepsilon}{2} \Delta g_t^\varepsilon = 0 & \text{in } ]0, 1[ \times \mathbb{R}^n, \\ g_1^\varepsilon = g^\varepsilon & \text{on } \mathbb{R}^n. \end{cases}$$

Furthermore, for all  $t \in [0, 1]$  let us define

$$(4) \quad \varrho_t^\varepsilon := f_t^\varepsilon g_t^\varepsilon \quad \text{and} \quad \mu_t^\varepsilon := \varrho_t^\varepsilon \mathcal{L}^n,$$

observing that  $\varrho_i^\varepsilon = \varrho_i$  and  $\mu_i^\varepsilon = \mu_i$ , for  $i = 0, 1$ . Finally, let us denote for every  $t \in ]0, 1[$

$$(5) \quad \psi_t^\varepsilon = \varepsilon \log g_t^\varepsilon$$

and  $\psi_1^\varepsilon(x) = \varepsilon \log g^\varepsilon(x)$  for every  $x \in \text{supp}(\mu_1)$ , appropriately extended to  $\mathbb{R}^n$  thanks to (2).

**Remark 2.2.** Since  $f_t^\varepsilon, g_t^\varepsilon$  solve two heat equations,  $\varrho_t^\varepsilon$  and  $\psi_t^\varepsilon$  solve the following coupled system of Hamilton–Jacobi–Bellman/Fokker–Planck equations, that is

$$\begin{cases} \frac{\partial}{\partial t} \varrho_t^\varepsilon - \frac{\varepsilon}{2} \Delta \varrho_t^\varepsilon + \text{div}(\varrho_t^\varepsilon \nabla \psi_t^\varepsilon) = 0 & \text{in } ]0, 1[ \times \mathbb{R}^n, \\ \frac{\partial}{\partial t} \psi_t^\varepsilon + \frac{\varepsilon}{2} \Delta \psi_t^\varepsilon + \frac{1}{2} |\nabla \psi_t^\varepsilon|^2 = 0 & \text{in } ]0, 1[ \times \mathbb{R}^n, \\ \varrho_0^\varepsilon = \varrho_0 & \text{on } \mathbb{R}^n, \\ \varrho_1^\varepsilon = \varrho_1 & \text{on } \mathbb{R}^n, \\ \psi_1^\varepsilon = \log g^\varepsilon & \text{on } \mathbb{R}^n. \end{cases}$$

◆

Then, the main results in [GT20] can be summarized as follows.

**Theorem 2.3.** *For every  $\mu_0 = \varrho_0 \mathcal{L}^n, \mu_1 = \varrho_1 \mathcal{L}^n \in \mathcal{P}_2(\mathbb{R}^n)$  with bounded densities and supports, it holds*

$$\begin{aligned} & \varepsilon \min_{\gamma \in \Gamma(\mu_0, \mu_1)} H(\gamma \mid \mathbf{R}_\varepsilon) = \varepsilon H(\mu_0 \mid \mathcal{L}^n) \\ & + \min \left\{ \int_0^1 \int \frac{|v_t|^2}{2} \eta_t d\mathcal{L}^n dt : \frac{\partial}{\partial t} \eta_t - \frac{\varepsilon}{2} \Delta \eta_t + \text{div}(v_t \eta_t) = 0, \eta_0 \mathcal{L}^n = \mu_0, \eta_1 \mathcal{L}^n = \mu_1 \right\}, \end{aligned}$$

where the Fokker–Planck equation has to be understood in distributional sense.

In particular, on the right-hand side, a minimizer is  $(\mu_t^\varepsilon, \nabla \psi_t^\varepsilon)$ .

**Fokker–Planck pairs** In order to better understand in which sense the solution to Fokker–Planck equation has to be understood, we need to introduce some cut-off functions (note that we will use the same ones in the proof of the main result). Let  $\lambda : [1, 2] \rightarrow [0, 1]$  be defined as

$$\lambda(r) = \frac{e^{-\frac{1}{2-r}}}{e^{-\frac{1}{2-r}} + e^{-\frac{1}{r-1}}}, \quad r \in ]1, 2[$$

and extended by continuity at  $r = 1, 2$ . As  $\lambda$  is smooth, there exist  $\Lambda \in \mathbb{R}$  such that for every  $r \in [1, 2]$

$$\lambda'(r), \lambda''(r) \leq \Lambda.$$

Given  $R > 1$ , consider the smooth function  $\zeta_R : \mathbb{R}^n \rightarrow [0, 1]$  such that

$$(6) \quad \zeta_R(x) = \begin{cases} 1 & \text{if } |x| \leq R, \\ \lambda\left(\frac{|x|}{R}\right) & \text{if } R < |x| \leq 2R, \\ 0 & \text{if } |x| > 2R. \end{cases}$$

and observe that its gradient is

$$\nabla \zeta_R(x) = \begin{cases} 0 & \text{if } |x| \leq R, \\ \frac{1}{R} \lambda' \left( \frac{|x|}{R} \right) \frac{x}{|x|} & \text{if } R < |x| \leq 2R, \\ 0 & \text{if } |x| > 2R, \end{cases}$$

and its Laplacian is

$$\Delta \zeta_R(x) = \begin{cases} 0 & \text{if } |x| \leq R, \\ \frac{1}{R^2} \lambda'' \left( \frac{|x|}{R} \right) + \frac{n-1}{|x|R} \lambda' \left( \frac{|x|}{R} \right) & \text{if } R < |x| \leq 2R, \\ 0 & \text{if } |x| > 2R. \end{cases}$$

We need a convenient definition of distributional solution of the Fokker–Planck equation.

**Definition 2.4. (distributional solution of Fokker–Planck equation)** Let  $v : [0, 1] \times \mathbb{R}^n \rightarrow \mathbb{R}^n$  be a Borel map and  $c > 0$ . We say that a curve  $\nu \in C([0, 1]; \mathcal{P}(\mathbb{R}^n))$  is a *distributional solution* of the *Fokker–Planck equation* with velocity  $v$ , namely

$$\frac{\partial}{\partial t} \nu - c \Delta \nu + \operatorname{div}(v\nu) = 0,$$

provided the following facts:

- (i) for every  $t \in [0, 1]$ ,  $|v_t| \in L^1(\nu_t)$ ,
- (ii) the function  $\{t \mapsto \|v_t\|_{L^1(\nu_t)}\}$  belongs to  $L^1(0, 1)$ ,
- (iii) for every  $f \in C^1([0, 1]) \cap C_c^\infty(\mathbb{R}^n)$ , it holds

$$(7) \quad \int_0^1 \int \left( \frac{\partial}{\partial t} f_t + c \Delta f_t + v_t \cdot \nabla f_t \right) d\nu_t dt = \int f_1 d\nu_1 - \int f_0 d\nu_0.$$

In particular, the pair  $(\nu_t, v_t)$ , where  $\nu_t$  is a distributional solution of the forward Fokker–Planck equation defined by  $v_t$ , is called *Fokker–Planck pair*.

We observe that  $(\mu_t^\varepsilon, \nabla \psi_t^\varepsilon)$  is a Fokker–Planck pair.

**Remark 2.5.** It is worth noting that the class of test functions used in (7) can be extended from  $C^1([0, 1]) \cap C_c^\infty(\mathbb{R}^n)$  to  $C^1([0, 1]) \cap C_b^2(\mathbb{R}^n)$ , that is if  $(\nu_t, v_t)$  is a Fokker–Planck pair, then (7) holds true for all  $f \in C^1([0, 1]) \cap C_b^2(\mathbb{R}^n)$ . Passing from  $C_c^\infty(\mathbb{R}^n)$  to  $C_b^2(\mathbb{R}^n)$  is effortless, thanks to a mollification argument. Let us then explain how to pass from  $C_c^2(\mathbb{R}^n)$  to  $C_b^2(\mathbb{R}^n)$ . Take  $f \in C^1([0, 1]) \cap C_b^2(\mathbb{R}^n)$ , consider, for  $R > 1$ ,

$$f_R(t, x) = f(t, x)\zeta_R(x),$$

where  $\zeta_R$  is as in (6), and use it as a test function. This gives

$$\begin{aligned} & \int f_1 \zeta_R d\nu_1 - \int f_0 \zeta_R d\nu_0 - \int_0^1 \int \zeta_R \left( \frac{\partial}{\partial t} f_t + c \Delta f_t + \nabla f_t \cdot v_t \right) d\nu_t dt \\ &= \int_0^1 \int (c f_t \Delta \zeta_R + 2c \nabla \zeta_R \cdot \nabla f_t + f_t D \zeta_R \cdot v_t) d\nu_t dt. \end{aligned}$$

The left-hand side converges to a rearrangement of (7), since  $\zeta_R \rightarrow 1$  as  $R \rightarrow +\infty$  and the right-hand side vanishes in the limit, since  $\Delta \zeta_R, \nabla \zeta_R$  and  $f$  are bounded by a constant and  $\Delta \zeta_R, \nabla \zeta_R \rightarrow 0$  as  $R \rightarrow +\infty$ .  $\blacklozenge$

### 3 The choice of marginals and the main result

As mentioned in the Introduction, one of the key points of the proof strategy in [GT20] is the boundedness of  $g^\varepsilon$ . This easily follows if both marginals have bounded and compactly supported densities. However, if we drop the compactness assumption on the supports, then  $g^\varepsilon$  may fail to be bounded (even if so are both marginals' densities), as shown in the following example.

**Example 3.1.** Consider  $\mu_0, \mu_1 \in \mathcal{P}_2(\mathbb{R})$  defined as

$$\mu_0 := \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-1)^2}{2}} \mathcal{L}^1, \quad \mu_1 := \frac{1}{2\sqrt{\pi}} e^{-\frac{(y-2)^2}{4}} \mathcal{L}^1$$

and  $f, g : \mathbb{R} \rightarrow [0, +\infty[$  defined as

$$f(x) := \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad g(y) := e^{y-1}.$$

We claim that the unique solution to  $(\text{SP}_\varepsilon)$  with the above marginals and  $\varepsilon = 1$  is  $\gamma := (f \otimes g)R_1$ . Indeed

$$\begin{aligned} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \int_{-\infty}^{+\infty} e^{y-1} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-y)^2}{2}} dy &= \frac{1}{2\pi} e^{-x^2-1+\frac{(x+1)^2}{2}} \int_{-\infty}^{+\infty} e^{-\frac{(y-(x+1))^2}{2}} dy \\ &= \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-1)^2}{2}} \end{aligned}$$

and

$$e^{y-1} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-y)^2}{2}} dx = \frac{1}{2\pi} e^{-\frac{1}{4}y^2+y-1} \int_{-\infty}^{+\infty} e^{-(x-\frac{1}{2}y)^2} dx = \frac{1}{2\sqrt{\pi}} e^{-\frac{(y-2)^2}{4}}.$$

This shows that  $\gamma \in \Gamma(\mu_0, \mu_1)$  and by [Nut21, Theorem 2.1], this is sufficient to conclude that  $\gamma$  is the Schrödinger plan between  $\mu_0$  and  $\mu_1$ .  $\blacklozenge$

Nonetheless, the fact that  $g^\varepsilon$  may be unbounded is not an obstruction in itself. What is needed to replicate the proof of [GT20] is a suitable integrability property of the absolute value of the time derivative of  $\psi_t^\varepsilon \varrho_t^\varepsilon$ : therefore, even if  $\psi_t^\varepsilon$  is unbounded, it is sufficient that  $\varrho_t^\varepsilon$  compensates for its growth. This is indeed what happens in the next example, where we consider two marginals that are Gaussian mixtures (thus demonstrating that proper integrability also exists outside the Gaussian framework).

**Example 3.2.** Consider  $\mu_0, \mu_1 \in \mathcal{P}_2(\mathbb{R})$  defined as

$$\mu_0 := \frac{1}{2\sqrt{2\pi}} \left( e^{-\frac{(x-1)^2}{2}} + e^{-\frac{(x+1)^2}{2}} \right) \mathcal{L}^1, \quad \mu_1 := \frac{1}{4\sqrt{\pi}} \left( e^{-\frac{(y-2)^2}{4}} + e^{-\frac{(y+2)^2}{4}} \right) \mathcal{L}^1$$

and  $f, g : \mathbb{R} \rightarrow [0, +\infty[$  given by

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad g(y) = \frac{1}{2} (e^{y-1} + e^{-y-1}).$$

Then, arguing as in Example 3.1, the unique solution to  $(\text{SP}_\varepsilon)$  with the above marginals and  $\varepsilon = 1$  is  $\gamma = (f \otimes g)\mathbb{R}_1$ . Given that, a direct computation provides

$$f_t(x) = \frac{1}{\sqrt{2\pi(1+t)}} e^{-\frac{x^2}{2(1+t)}}, \quad g_t(x) = \frac{1}{2} \left( e^{x-\frac{1+t}{2}} + e^{-x-\frac{1+t}{2}} \right),$$

whence

$$\varrho_t(x) = \frac{1}{2\sqrt{2\pi(t+1)}} \left( e^{-\frac{(x-(1+t))^2}{2(1+t)}} + e^{-\frac{(x+(1+t))^2}{2(1+t)}} \right)$$

and

$$\psi_t(x) = -\log 2 + \log \left( e^{x-\frac{1+t}{2}} + e^{-x-\frac{1+t}{2}} \right).$$

Now, for every  $t \in [0, 1]$ , the function  $x \mapsto \psi_t(x)\varrho_t(x)$  is integrable and, for every  $x \in \mathbb{R}$ , the function  $t \mapsto \psi_t(x)\varrho_t(x)$  is differentiable with

$$\left| \frac{\partial}{\partial t} (\psi_t(x)\varrho_t(x)) \right| \leq 4e^{-\frac{x^2}{100}}, \quad \forall x \in \mathbb{R}.$$

We can thus differentiate under the integral sign and find, using the equations that  $\psi_t$  and  $\varrho_t$  solve and integration by parts,

$$\frac{d}{dt} \int_{-\infty}^{+\infty} \psi_t \varrho_t d\mathcal{L}^1 = \int_{-\infty}^{+\infty} \left( \rho_t \frac{\partial}{\partial t} \psi_t + \psi_t \frac{\partial}{\partial t} \varrho_t \right) d\mathcal{L}^1$$

$$\begin{aligned}
&= -\frac{1}{2} \int_{-\infty}^{+\infty} \Delta \psi_t \varrho_t d\mathcal{L}^1 - \frac{1}{2} \int_{-\infty}^{+\infty} |\nabla \psi_t|^2 \varrho_t d\mathcal{L}^1 \\
&\quad + \frac{1}{2} \int_{-\infty}^{+\infty} \psi_t \Delta \varrho_t d\mathcal{L}^1 - \int_{-\infty}^{+\infty} \psi_t \operatorname{div}(\varrho_t \nabla \psi_t) d\mathcal{L}^1 \\
&= \frac{1}{2} \int_{-\infty}^{+\infty} |\nabla \psi_t|^2 \varrho_t d\mathcal{L}^1.
\end{aligned}$$

If we further integrate both sides w.r.t.  $t \in [0, 1]$ , we find

$$\int \psi_1 d\mu_1 - \int \psi_0 d\mu_0 = \int_0^1 \int \frac{|\nabla \psi_t|^2}{2} \varrho_t d\mathcal{L}^1 dt.$$

Finally, noticing that  $\psi_0 = \log \varrho_0 - \log f$  and  $\psi_1 = \log g$ , and using also the marginal constraints on  $\gamma$ , we arrive at

$$H(\gamma | R_1) = H(\mu_0 | \mathcal{L}^1) + \int_0^1 \int \frac{|\nabla \psi_t|^2}{2} d\mu_t dt.$$

◆

With this example in mind we are now ready to get the right assumptions for our framework: the idea is to preserve the property of having very little mass at infinity instead of the strongest assumption of the compact support, in order to obtain the necessary integrability.

**Assumptions 3.3.** Suppose  $\mu_0, \mu_1 \in \mathcal{P}_2(\mathbb{R}^n)$  have the form

$$\mu_0 = e^{-V_0} \mathcal{L}^n, \quad \mu_1 = e^{-V_1} \mathcal{L}^n,$$

with  $V_0, V_1 \in C^2(\mathbb{R}^n)$  for which there exist  $C > 0$  such that, for every  $x, \xi \in \mathbb{R}^n$ ,

$$\begin{aligned}
C^{-1}|\xi|^2 &\leq \xi \cdot D^2 V_0(x) \xi \leq C|\xi|^2, \\
C^{-1}|\xi|^2 &\leq \xi \cdot D^2 V_1(x) \xi \leq C|\xi|^2.
\end{aligned}$$

**Remark 3.4.** It is possible to include more general cases in our framework if we consider  $\mu_0, \mu_1 \in \mathcal{P}_2(\mathbb{R}^n)$  having the form

$$\mu_0 = \begin{cases} e^{-V_0} \mathcal{L}^n & \text{if } V_0 \text{ is defined,} \\ 0 & \text{otherwise,} \end{cases} \quad \mu_1 = \begin{cases} e^{-V_1} \mathcal{L}^n & \text{if } V_1 \text{ is defined,} \\ 0 & \text{otherwise,} \end{cases}$$

with  $V_0 \in C^2(\operatorname{dom}(V_0)), V_1 \in C^2(\operatorname{dom}(V_1)), \operatorname{dom}(V_0), \operatorname{dom}(V_1) \subseteq \mathbb{R}^n$  non-empty, possibly unbounded and with a finite number of convex connected components, for which there exist  $C > 0$  such that

$$\begin{aligned}
C^{-1}|\xi|^2 &\leq \xi \cdot D^2 V_0(x) \xi \leq C|\xi|^2, & \text{for every } x \in \operatorname{dom}(V_0), \xi \in \mathbb{R}^n, \\
C^{-1}|\xi|^2 &\leq \xi \cdot D^2 V_1(x) \xi \leq C|\xi|^2, & \text{for every } x \in \operatorname{dom}(V_1), \xi \in \mathbb{R}^n.
\end{aligned}$$

With this slightly more general framework, it is possible to recover the case of compactly supported marginals, namely the result in [GT20], and also manage the case of marginals with non-connected supports. The difference in the proof is minor: while the underlying strategy remains unchanged, one should be more careful when applying Taylor expansions and on writing the domain of  $\psi_1^\varepsilon$ . It is merely a matter of expanding on technical details that are not essential to the true core of the proof. Therefore, for the sake of presentation we prefer to adopt Assumptions 3.3 throughout the whole manuscript.  $\blacklozenge$

First of all, let us verify that under these assumptions existence, uniqueness, and decomposition of the solution to  $(\text{SP}_\varepsilon)$  hold true.

**Proposition 3.5.** *Under Assumptions 3.3, for every  $\varepsilon > 0$  there exists a unique solution  $\gamma^\varepsilon$  to  $(\text{SP}_\varepsilon)$ . In particular, there exist two Borel functions  $f^\varepsilon, g^\varepsilon : \mathbb{R}^n \rightarrow [0, +\infty[$ , unique a.e. in  $\mathbb{R}^n$  up to a rescaling  $(f, g) \mapsto (cf, \frac{g}{c})$  for some  $c > 0$ , such that*

$$\gamma^\varepsilon = (f^\varepsilon \otimes g^\varepsilon) \mathbf{R}_\varepsilon.$$

*Proof.* According to [GT21, Proposition 2.1], if we exhibit a Borel function  $B : \mathbb{R}^n \times \mathbb{R}^n \rightarrow [0, +\infty[$  such that

$$(8) \quad \int e^{-B(x)-B(y)} d\mathbf{R}_\varepsilon(x, y) < +\infty, \quad \int B d\mu_0 < +\infty, \quad \int B d\mu_1 < +\infty$$

and show that  $H(\mu_0 \otimes \mu_1 | \mathbf{R}_\varepsilon) < +\infty$ , then all the claimed properties hold true.

As concerns (8), since the marginals have finite second moment, the choice  $B(x) = |x|^2$  satisfies the requirements. As for the feasibility of the marginal constraint, note that

$$\begin{aligned} H(\mu_0 \otimes \mu_1 | \mathbf{R}_\varepsilon) &= H(\mu_0 \otimes \mu_1 | \mathcal{L}^n \otimes \mathcal{L}^n) + \int \log \left( \sqrt{2\pi\varepsilon} e^{\frac{|x-y|^2}{2\varepsilon}} \right) d(\mu_0 \otimes \mu_1) \\ &= H(\mu_0 | \mathcal{L}^n) + H(\mu_1 | \mathcal{L}^n) + \int \log \left( \sqrt{2\pi\varepsilon} e^{\frac{|x-y|^2}{2\varepsilon}} \right) d(\mu_0 \otimes \mu_1) < +\infty. \end{aligned}$$

Thus [GT21, Proposition 2.1] grants existence and uniqueness of the solution, as well as the existence of the decomposition.  $\blacksquare$

We are now ready to state our main result. A complete proof is postponed after Proposition 4.6 because we need to deal at first with some technical Lemmas.

**Theorem 3.6. (Benamou–Brenier formula for the Schrödinger problem)** *Under Assumptions 3.3, for every  $\varepsilon > 0$  it holds*

$$\begin{aligned} \varepsilon \min_{\gamma \in \Gamma(\mu_0, \mu_1)} H(\gamma | \mathbf{R}_\varepsilon) &= \varepsilon H(\mu_0 | \mathcal{L}^n) \\ &+ \min \left\{ \int_0^1 \int \frac{|v_t|^2}{2} \eta_t d\mathcal{L}^n dt : (\eta_t \mathcal{L}^n, v_t) \text{ Fokker–Planck pair}, \eta_0 = \varrho_0, \eta_1 = \varrho_1 \right\}. \end{aligned}$$

*Moreover, on the right-hand side, the (unique) minimizer is  $(\mu_t^\varepsilon, \nabla \psi_t^\varepsilon)$ .*

*Proof. (Sketch)* The proof of the equality can be summarized in two main parts: the first is to prove

$$\varepsilon H(\gamma^\varepsilon \mid \mathbf{R}_\varepsilon) = \varepsilon H(\mu_0 \mid \mathcal{L}^n) + \int_0^1 \int \frac{|\nabla \psi_t^\varepsilon|^2}{2} d\mu_t^\varepsilon dt$$

that permits us to obtain the  $\geq$  inequality. The second part is to argue by duality to obtain the  $\leq$  inequality.

We postpone the proof of the uniqueness of the minimizer. ■

As a Corollary, we can also obtain a dual formulation for the Schrödinger problem.

**Corollary 3.7.** *Under Assumptions 3.3, for every  $\varepsilon > 0$  it holds*

$$\begin{aligned} \varepsilon \min_{\gamma \in \Gamma(\mu_0, \mu_1)} H(\gamma \mid \mathbf{R}_\varepsilon) &= \varepsilon H(\mu_0 \mid \mathcal{L}^n) \\ &+ \max \left\{ \int f_1 d\mu_1 - \int f_0 d\mu_0 : f \in \mathfrak{H}, \frac{\partial}{\partial t} f + \frac{\varepsilon}{2} \Delta f + \frac{1}{2} |\nabla f|^2 \leq 0 \right\}, \end{aligned}$$

where

$$\mathfrak{H} = \left\{ f \in C^1([0, 1]) \cap C^2(\mathbb{R}^n) : m_t^\varepsilon |\xi|^2 \leq \xi \cdot D^2 f(t, x) \xi \leq M_t^\varepsilon |\xi|^2 \right\},$$

where  $m_t^\varepsilon, M_t^\varepsilon \in \mathbb{R}$  and depend on the marginals as in Lemma 4.2 (b). Moreover, on the right-hand side, a maximizer is given by  $\psi_t^\varepsilon$ .

*Proof.* It is a direct consequence of Lemma 4.5, Proposition 4.6, and the previous Theorem. ■

## 4 Proofs

### 4.1 Technical results

In the following, we collect almost all the technical parts of the document, some of which we believe are interesting in their own right.

The first technical result is on the growth of  $\psi_1^\varepsilon = \varepsilon \log g^\varepsilon$ . We are able to prove, thanks to the results of [Chi+24], that it inherits Hessian bounds from the ones of the marginals.

**Lemma 4.1.** *Let  $\mu_0$  and  $\mu_1$  satisfy Assumptions 3.3 and  $\psi_1^\varepsilon = \varepsilon \log g^\varepsilon$ , with  $g^\varepsilon$  a solution to (2). Then for every  $\varepsilon > 0$ , the following facts hold true:*

(a)  $\psi_1^\varepsilon \in C^2(\mathbb{R}^n)$ ,

(b) for every  $x, \xi \in \mathbb{R}^n$ ,

$$m^\varepsilon |\xi|^2 \leq \xi \cdot D^2 \psi_1^\varepsilon(x) \xi \leq M^\varepsilon |\xi|^2,$$

where

$$m^\varepsilon = -1 - \frac{\varepsilon}{2C} + \frac{1}{C} \sqrt{\frac{\varepsilon^2}{4} + 1}, \quad M^\varepsilon = -1 - \frac{\varepsilon}{2} C + C \sqrt{\frac{\varepsilon^2}{4} + 1},$$

and  $C$  is as in Assumptions 3.3,

(c) for every  $x \in \mathbb{R}^n$ ,

$$|\psi_1^\varepsilon(x)| \leq c^\varepsilon + b^\varepsilon|x| + \frac{a^\varepsilon}{2}|x|^2,$$

where  $c^\varepsilon = |\psi_1^\varepsilon(0)|$ ,  $b^\varepsilon = |\nabla\psi_1^\varepsilon(0)|$ , and  $a^\varepsilon = \max\{|m^\varepsilon|, |M^\varepsilon|\}$ .

*Proof.*

(a) Applying log to both sides of the second equation of the Schrödinger system and recalling (3a), we can write  $\psi_1^\varepsilon = \varepsilon \log g^\varepsilon = \varepsilon \log \varrho_1 - \varepsilon \log f_1^\varepsilon$ , whence the desired regularity.

(b) Since, by (a),  $\psi_1^\varepsilon \in C^2(\mathbb{R}^n)$ , arguing as in [Chi+24, Corollary 2.3], we have, for every  $x, \xi \in \mathbb{R}^n$ ,

$$m^\varepsilon|\xi|^2 \leq \xi \cdot D^2\psi_1^\varepsilon(x)\xi \leq M^\varepsilon|\xi|^2.$$

(c) Using Taylor's formula with integral remainder, for every  $x \in \mathbb{R}^n$ ,

$$\psi_1^\varepsilon(x) = \psi_1^\varepsilon(0) + \nabla\psi_1^\varepsilon(0) \cdot x + \int_0^1 (1-t)x \cdot D^2\psi_1^\varepsilon(tx)xdt,$$

and the inequality from (b) we get

$$\psi_1^\varepsilon(0) + \nabla\psi_1^\varepsilon(0) \cdot x + \frac{1}{2}m^\varepsilon|x|^2 \leq \psi_1^\varepsilon(x) \leq \psi_1^\varepsilon(0) + \nabla\psi_1^\varepsilon(0) \cdot x + \frac{1}{2}M^\varepsilon|x|^2,$$

then, by Cauchy–Schwarz inequality and some elementary estimates with the absolute value and the maximum between real numbers,

$$-c^\varepsilon - b^\varepsilon|x| - a^\varepsilon|x|^2 \leq \psi_1^\varepsilon(x) \leq c^\varepsilon + b^\varepsilon|x| + a^\varepsilon|x|^2$$

and the result follows. ■

In this second Lemma, we apply the important result contained in [Con24] to propagate the Hessian bounds on  $\psi_1^\varepsilon$  to every time in  $[0, 1]$ .

**Lemma 4.2.** *Let  $\mu_0$  and  $\mu_1$  satisfy Assumptions 3.3 and  $\psi_t^\varepsilon = \varepsilon \log(g_t^\varepsilon)$ , with  $g_t^\varepsilon$  a solution to the backward heat equation with datum  $g^\varepsilon$ . Then, for every  $\varepsilon > 0$ , the following facts hold true:*

(a)  $\psi_t^\varepsilon \in C^1([0, 1]) \cap C^2(\mathbb{R}^n)$ ,

(b) for every  $t \in [0, 1]$ , for every  $x, \xi \in \mathbb{R}^n$ ,

$$\frac{m^\varepsilon}{1 + (1-t)m^\varepsilon}|\xi|^2 \leq \xi \cdot D^2\psi_t^\varepsilon(x)\xi \leq \frac{M^\varepsilon}{1 + (1-t)M^\varepsilon}|\xi|^2,$$

(c) for every  $t \in [0, 1]$ , for every  $x \in \mathbb{R}^n$ ,

$$|\nabla\psi_t^\varepsilon(x)| \leq b_t^\varepsilon + a_t^\varepsilon|x|,$$

where  $b_t^\varepsilon = |\nabla\psi_t^\varepsilon(0)|$  and

$$a_t^\varepsilon = \max \left\{ \left| \frac{m^\varepsilon}{1 + (1-t)m^\varepsilon} \right|, \left| \frac{M^\varepsilon}{1 + (1-t)M^\varepsilon} \right| \right\},$$

(d) for every  $t \in [0, 1]$ , for every  $x \in \mathbb{R}^n$ ,

$$|\psi_t^\varepsilon(x)| \leq c_t^\varepsilon + b_t^\varepsilon|x| + \frac{a_t^\varepsilon}{2}|x|^2,$$

where  $c_t^\varepsilon = |\psi_t^\varepsilon(0)|$ .

*Proof.*

(a) It comes from the very definition of  $\psi_t^\varepsilon$ . Indeed, the latter is obtained from a convolution against the heat kernel.

(b) It is a direct application of [Con24].

(c) Note that, for all  $x \in \mathbb{R}^n$ , it holds

$$|\nabla\psi_t^\varepsilon(x)| = \left| \nabla\psi_t^\varepsilon(0) + \int_0^1 D^2\psi_t^\varepsilon(sx)x \, ds \right| \leq |\nabla\psi_t^\varepsilon(0)| + \int_0^1 |D^2\psi_t^\varepsilon(sx)||x| \, ds,$$

so that the conclusion follows from (b).

(c) It is the same argument as in Lemma 4.1 (c). ■

In the following, we shall use the following result (for more details see for instance [BF81, Lemma 1] and [Kat72]).

**Lemma 4.3. (parabolic Kato's inequality)** *Let  $U$  be an open subset of  $\mathbb{R} \times \mathbb{R}^n$  and  $u \in L_{loc}^1(U)$  a weak solution of*

$$\frac{\partial}{\partial t}u - \Delta u = f,$$

with  $f \in L_{loc}^1(U)$ . Then, weakly,

$$\frac{\partial}{\partial t}|u| - \Delta|u| \leq \text{sgn}(u)f,$$

where

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

In the following technical result, which we believe to be interesting also in itself, we establish that, under some bounds on the velocity fields, a solution to the Fokker–Planck equation and its spatial derivatives have finite Gaussian and  $k$ -th moments, respectively.

**Proposition 4.4.** *Let  $\mu_0$  satisfying Assumptions 3.3 and  $(\rho_t^\varepsilon \mathcal{L}^n, \nabla\psi_t^\varepsilon)$  be a Fokker–Planck pair with initial datum  $\mu_0$  such that  $\psi_t^\varepsilon \in C^1([0, 1]) \cap C^2(\mathbb{R}^n)$  and, for every*

$t \in [0, 1]$ , for every  $x, \xi \in \mathbb{R}^n$ ,

$$m_t^\varepsilon |\xi|^2 \leq \xi \cdot D^2 \psi_t^\varepsilon(x) \xi \leq M_t^\varepsilon |\xi|^2,$$

where  $m_t^\varepsilon, M_t^\varepsilon$  are continuous functions on  $[0, 1]$ . Then, for every  $\varepsilon > 0$ , the following facts hold true:

(a)  $\varrho_t^\varepsilon \in C^1([0, 1]) \cap C^2(\mathbb{R}^n)$ ,

(b) there exists  $W : [0, 1] \rightarrow \mathbb{R}$  such that  $0 < W(1) \leq W < C_2$  and, for every  $R > 1$ ,

$$\sup_{t \in [0, 1]} \int_{\mathbb{R}^n \setminus B(0, R)} \varrho_t^\varepsilon d\mathcal{L}^n \leq I e^{-W(1)R^2},$$

where

$$I = \sup_{t \in [0, 1]} \int e^{W(t)|x|^2} \varrho_t^\varepsilon(x) dx < +\infty,$$

(c) for every  $k \in \mathbb{N}$ ,

$$\sup_{t \in [0, 1]} \int |x|^k \varrho_t^\varepsilon(x) dx < +\infty.$$

Moreover,

$$\begin{aligned} \sup_{t \in [0, 1]} \int (1 + |x|^2) \varrho_t^\varepsilon(x) dx &< +\infty, \\ \sup_{t \in [0, 1]} \int (1 + |x|^2) |\nabla \varrho_t^\varepsilon(x)| dx &< +\infty \\ \sup_{t \in [0, 1]} \int (1 + |x|^2) |\Delta \varrho_t^\varepsilon(x)| dx &< +\infty. \end{aligned}$$

*Proof.* We firstly notice that, arguing in the same way as in Lemma 4.2, for every  $t \in [0, 1]$ , for every  $x \in \mathbb{R}^n$ ,

$$|\nabla \psi_t^\varepsilon(x)| \leq b_t^\varepsilon + a_t^\varepsilon |x|,$$

where  $a_t^\varepsilon, b_t^\varepsilon$  are continuous functions on  $[0, 1]$ . So, in particular, for every  $t \in [0, 1]$ , for every  $x \in \mathbb{R}^n$ ,

$$|\psi_t^\varepsilon(x)| \leq c_t^\varepsilon + b_t^\varepsilon |x| + \frac{a_t^\varepsilon}{2} |x|^2,$$

where  $c_t^\varepsilon$  is another continuous function on  $[0, 1]$ .

(a) It follows by definition of  $\varrho_t^\varepsilon$ .

(b) First of all, given  $k \in ]0, C_2[$ , take

$$W(t) = k \frac{A e^{-At}}{A + 2k\varepsilon(1 - e^{-At})},$$

where

$$A = \frac{1}{2} \max_{t \in [0,1]} b_t^\varepsilon + 2 \max_{t \in [0,1]} a_t^\varepsilon,$$

and  $a_t^\varepsilon$  and  $b_t^\varepsilon$  are given by Lemmas 4.1 and 4.2. One can show that  $W(t)$  solves the following Cauchy-Lipschitz problem

$$\begin{cases} W' + 2\varepsilon W^2 + AW = 0, \\ W(0) = k \end{cases}$$

and  $0 < W(1) \leq W < C_2$ . Now, given  $N > 0$ , introducing, for ease of notation,

$$\Phi_R^N(x) = \zeta_R(x) \min \left\{ e^{W(t)|x|^2}, N \right\},$$

using the derivation under the integral sign, the equations solved by  $\varrho_t^\varepsilon$ , and integrating by parts we obtain

$$\begin{aligned} \frac{d}{dt} \int \Phi_R^N(x) \varrho_t^\varepsilon(x) dx &\leq \int W'(t)|x|^2 \zeta_R(x) e^{W(t)|x|^2} \varrho_t^\varepsilon(x) dx \\ &\quad + \int \Phi_R^N(x) \left( \frac{\varepsilon}{2} \Delta \varrho_t^\varepsilon - \operatorname{div}(\varrho_t^\varepsilon \nabla \psi_t^\varepsilon) \right) dx \\ &\leq \int W'(t)|x|^2 \zeta_R(x) e^{W(t)|x|^2} \varrho_t^\varepsilon(x) dx + \frac{\varepsilon}{2} \int \Delta (\Phi_R^N(x)) \varrho_t^\varepsilon dx \\ &\quad + \int \nabla (\Phi_R^N(x)) \cdot \nabla \psi_t^\varepsilon \varrho_t^\varepsilon dx. \end{aligned}$$

Using Leibniz's rule for gradient and Laplacian, the property of exponential, Cauchy-Schwarz inequality, since

$$\int_{\{e^{W(t)|x|^2}=N\}} \Delta \min \left\{ e^{W(t)|x|^2}, N \right\} \varrho_t^\varepsilon dx = - \int 2N \sqrt{\log N} \varrho_t^\varepsilon d\mathcal{H}^{n-1} \leq 0$$

we get

$$\begin{aligned} &\frac{\varepsilon}{2} \int \Delta (\Phi_R^N(x)) \varrho_t^\varepsilon dx \\ &\leq \varepsilon \int (n + 2W(t)|x|^2) W(t) \Phi_R^N(x) \varrho_t^\varepsilon dx + \frac{\varepsilon}{2} \frac{\Lambda}{R^2} n \int_{B(0,2R) \setminus B(0,R)} \min \left\{ e^{W(t)|x|^2}, N \right\} \varrho_t^\varepsilon dx \\ &\quad + 4\varepsilon \int_{B(0,2R) \setminus B(0,R)} \Lambda W(t) \min \left\{ e^{W(t)|x|^2}, N \right\} \varrho_t^\varepsilon dx \end{aligned}$$

and

$$\begin{aligned}
& \int \nabla (\Phi_R^N(x)) \cdot \nabla \psi_t^\varepsilon \varrho_t^\varepsilon dx \\
&= \int \left( \nabla \zeta_R(x) \cdot \nabla \psi_t^\varepsilon \min \left\{ e^{W(t)|x|^2}, N \right\} + \zeta_R(x) \nabla \min \left\{ e^{W(t)|x|^2}, N \right\} \cdot \nabla \psi_t^\varepsilon \right) \varrho_t^\varepsilon dx \\
&\leq \int_{\mathbb{B}(0,2R) \setminus \mathbb{B}(0,R)} \frac{\Lambda}{R} (b_t^\varepsilon + a_t^\varepsilon |x|) \min \left\{ e^{W(t)|x|^2}, N \right\} \varrho_t^\varepsilon dx \\
&\quad + 2 \int \Phi_R^N(x) W(t) |x| (b_t^\varepsilon + a_t^\varepsilon |x|) \varrho_t^\varepsilon dx,
\end{aligned}$$

then, putting all together,

$$\begin{aligned}
& \frac{d}{dt} \int \Phi_R^N(x) \varrho_t^\varepsilon(x) dx \\
&\leq \int \underbrace{(W'(t) + 2\varepsilon W(t)^2 + AW(t))}_{=0} |x|^2 \Phi_R^N(x) \varrho_t^\varepsilon dx \\
&\quad + \left( \varepsilon n + 2 \max_{t \in [0,1]} b_t^\varepsilon \right) C \int \Phi_R^N(x) \varrho_t^\varepsilon(x) dx + \hat{C} \int_{\mathbb{B}(0,2R) \setminus \mathbb{B}(0,R)} \min \left\{ e^{W(t)|x|^2}, N \right\} \varrho_t^\varepsilon dx \\
&\leq \left( \varepsilon n + 2 \max_{t \in [0,1]} b_t^\varepsilon \right) C \int \Phi_R^N(x) \varrho_t^\varepsilon(x) dx + \hat{C} N \left( 1 - \max_{t \in [0,1]} \mu_t^\varepsilon(\mathbb{B}(0,R)) \right).
\end{aligned}$$

By Grönwall's Lemma,

$$\begin{aligned}
& \int \Phi_R^N(x) \varrho_t^\varepsilon(x) dx \\
&\leq e^{\left( \varepsilon n + 2 \max_{t \in [0,1]} b_t^\varepsilon \right) C t} \int \zeta_R(x) \min \left\{ e^{W(0)|x|^2}, N \right\} \varrho_0(x) dx \\
&\quad + \frac{\hat{C} N \left( 1 - \max_{t \in [0,1]} \mu_t^\varepsilon(\mathbb{B}(0,R)) \right)}{\left( \varepsilon n + 2 \max_{t \in [0,1]} b_t^\varepsilon \right) C} \left( e^{\left( \varepsilon n + 2 \max_{t \in [0,1]} b_t^\varepsilon \right) C t} - 1 \right).
\end{aligned}$$

Passing to the limit as  $R \rightarrow +\infty$ , according to Fatou's Lemma and using the continuity of  $\max_{t \in [0,1]} \varrho_t^\varepsilon$ , we get

$$\int \min \left\{ e^{W(t)|x|^2}, N \right\} \varrho_t^\varepsilon(x) dx \leq e^{\left( \varepsilon n + 2 \max_{t \in [0,1]} b_t^\varepsilon \right) C_2 t} \int e^{W(0)|x|^2} \varrho_0(x) dx.$$

Passing now to the limit as  $N \rightarrow +\infty$ ,

$$\int e^{W(t)|x|^2} \varrho_t^\varepsilon(x) dx \leq e^{\left(\varepsilon n + 2 \max_{t \in [0,1]} b_t^\varepsilon\right) C_2 t} \int e^{W(0)|x|^2} \varrho_0(x) dx.$$

As a consequence,

$$\int e^{W(t)|x|^2} \varrho_t^\varepsilon(x) dx < +\infty,$$

then  $I < +\infty$  and, using monotonicity of exponential and Markov's inequality, for every  $R > 0$ ,

$$\begin{aligned} \int_{\mathbb{R}^n \setminus B(0,R)} \varrho_t^\varepsilon d\mathcal{L}^n &= \mu_t^\varepsilon(|x| \geq R) = \mu_t^\varepsilon\left(e^{W(t)|x|^2} \geq e^{W(t)R^2}\right) \\ &\leq \frac{\int e^{W(t)|x|^2} \varrho_t^\varepsilon(x) dx}{e^{W(t)R^2}} \leq I e^{-W(1)R^2}. \end{aligned}$$

The desired inequality simply follows passing to the supremum with respect to  $t$ .

(c) Consider  $k \in \mathbb{N}$ . If  $k = 0$ , the result is trivial. Let us, then, consider only the case  $k \geq 1$ . By Fubini–Tonelli's Theorem, combined with (b), we get

$$\int |x|^k \varrho_t^\varepsilon(x) dx \leq \tilde{C} + k \int_1^{+\infty} r^{k-1} e^{W(1)r^2} dr < +\infty$$

and, passing to the supremum with respect to  $t$ ,

$$\sup_{t \in [0,1]} \int |x|^k \varrho_t^\varepsilon(x) dx < +\infty.$$

Now, given  $\alpha_0 > 0$ , take

$$\alpha(t) = \alpha_0 e^{-Qt}$$

where

$$Q = \frac{9}{2} \varepsilon \Lambda + \left(\frac{1}{2} \Lambda + \frac{3}{2}\right) \max \left\{ \max_{t \in [0,1]} b_t^\varepsilon, \max_{t \in [0,1]} a_t^\varepsilon \right\}.$$

A direct computation provides

$$\begin{cases} \alpha'(t) + Q\alpha(t) = 0, \\ \alpha(0) = \alpha_0 \end{cases}$$

and  $0 < \alpha(1) \leq \alpha < \alpha_0$ . Now, being sufficient to study the case  $t \in ]0, 1[$ , given  $R > 1$  and  $i = 1, \dots, n$ , using the derivation under the integral sign and the parabolic Kato's inequality,

$$\frac{d}{dt} \int \zeta_R(x) \alpha(t) (1 + |x|^2) |\nabla_{x_i} \varrho_t^\varepsilon(x)| dx = \int \zeta_R(x) \alpha'(t) (1 + |x|^2) |\nabla_{x_i} \varrho_t^\varepsilon(x)| dx$$

$$\begin{aligned}
& + \int \zeta_R(x)\alpha(t)(1+|x|^2)\frac{\partial}{\partial t}|\nabla_{x_i}\varrho_t^\varepsilon(x)|dx \leq \int \zeta_R(x)\alpha'(t)(1+|x|^2)|\nabla_{x_i}\varrho_t^\varepsilon(x)|dx \\
& + \frac{\varepsilon}{2} \int \zeta_R(x)\alpha(t)(1+|x|^2)\Delta|\nabla_{x_i}\varrho_t^\varepsilon(x)|dx \\
& - \int \zeta_R(x)\alpha(t)(1+|x|^2)\operatorname{sgn}(\nabla_{x_i}\varrho_t^\varepsilon(x))\operatorname{div}\left(\nabla\psi_t^\varepsilon(x)\nabla_{x_i}\varrho_t^\varepsilon(x) + \nabla_{x_i}\nabla\psi_t^\varepsilon(x)\varrho_t^\varepsilon(x)\right)dx.
\end{aligned}$$

Now, using Gauss–Green’s formula,

$$\begin{aligned}
& \int \zeta_R(x)\alpha(t)(1+|x|^2)\Delta|\nabla_{x_i}\varrho_t^\varepsilon(x)|dx \\
& \leq \int 9\Lambda\alpha(t)|\nabla_{x_i}\varrho_t^\varepsilon(x)|dx + 2n \int \zeta_R(x)\alpha(t)|\nabla_{x_i}\varrho_t^\varepsilon(x)|dx
\end{aligned}$$

and

$$\begin{aligned}
& - \int \zeta_R(x)\alpha(t)(1+|x|^2)\operatorname{sgn}(\nabla_{x_i}\varrho_t^\varepsilon(x))\operatorname{div}\left(\nabla\psi_t^\varepsilon(x)\nabla_{x_i}\varrho_t^\varepsilon(x) + \nabla_{x_i}\nabla\psi_t^\varepsilon(x)\varrho_t^\varepsilon(x)\right)dx \\
& \leq \left(\frac{1}{2}\Lambda + \frac{3}{2}\right) \max\left\{\max_{t\in[0,1]}b_t^\varepsilon, \max_{t\in[0,1]}a_t^\varepsilon\right\} \int \alpha(t)(1+|x|^2)|\nabla_{x_i}\varrho_t^\varepsilon(x)|dx + \tilde{C}.
\end{aligned}$$

From this, it comes that

$$\frac{d}{dt} \int \zeta_R(x)\alpha(t)(1+|x|^2)|\nabla_{x_i}\varrho_t^\varepsilon(x)|dx \leq n\varepsilon \int \zeta_R(x)\alpha(t)|\nabla_{x_i}\varrho_t^\varepsilon(x)|dx + \tilde{C}$$

so, by Grönwall’s Lemma and Fatou’s Lemma,

$$\sup_{t\in[0,1]} \int (1+|x|^2)|\nabla_{x_i}\varrho_t^\varepsilon(x)|dx < +\infty.$$

It is now a direct consequence of the definition of gradient to obtain

$$\sup_{t\in[0,1]} \int (1+|x|^2)|\nabla\varrho_t^\varepsilon(x)|dx < +\infty.$$

The last part of the statement can be obtained in an analogous way. ■

The last technical Lemma is a density result for the class of test function used in the dual problem for the Benamou–Brenier formulation.

**Lemma 4.5.** *Let  $\mu_0$  and  $\mu_1$  satisfying Assumptions 3.3. Then, for every  $\varepsilon > 0$ ,*

$$\begin{aligned}
& \sup\left\{\int f_1d\mu_1 - \int f_0d\mu_0 : f \in C^1([0,1]) \cap C_b^2(\mathbb{R}^n), \frac{\partial}{\partial t}f + \frac{\varepsilon}{2}\Delta f + \frac{1}{2}|\nabla f|^2 \leq 0\right\} \\
& = \sup\left\{\int f_1d\mu_1 - \int f_0d\mu_0 : f \in \mathfrak{F}, \frac{\partial}{\partial t}f + \frac{\varepsilon}{2}\Delta f + \frac{1}{2}|\nabla f|^2 \leq 0\right\}
\end{aligned}$$

where

$$\mathfrak{H} = \{f \in C^1([0, 1]) \cap C^2(\mathbb{R}^n) : m_t^\varepsilon |\xi|^2 \leq \xi \cdot D^2 f(t, x) \xi \leq M_t^\varepsilon |\xi|^2, m_t^\varepsilon, M_t^\varepsilon \in \mathbb{R}\}$$

and  $m_t^\varepsilon, M_t^\varepsilon \in \mathbb{R}$  depend on the marginals as in Lemma 4.2 (b).

*Proof.* First of all, the inequality

$$\begin{aligned} & \sup \left\{ \int f_1 d\mu_1 - \int f_0 d\mu_0 : f \in C^1([0, 1]) \cap C_b^2(\mathbb{R}^n), \frac{\partial}{\partial t} f + \frac{\varepsilon}{2} \Delta f + \frac{1}{2} |\nabla f|^2 \leq 0 \right\} \\ & \leq \sup \left\{ \int f_1 d\mu_1 - \int f_0 d\mu_0 : f \in \mathfrak{H}, \frac{\partial}{\partial t} f + \frac{\varepsilon}{2} \Delta f + \frac{1}{2} |\nabla f|^2 \leq 0 \right\} \end{aligned}$$

is trivial. For the converse inequality, take  $f \in \mathfrak{H}$  such that

$$\frac{\partial}{\partial t} f + \frac{\varepsilon}{2} \Delta f + \frac{1}{2} |\nabla f|^2 \leq 0$$

and consider, given  $R > 1$ ,

$$f_R(t, x) = \varepsilon \log \left( \int \left( e^{\frac{f(1, y)}{\varepsilon}} \zeta_R(y) + \frac{1}{R} \right) r_{\varepsilon(1-t)}(x, y) dy \right).$$

By construction,  $f_R \in C^1([0, 1]) \cap C_b^2(\mathbb{R}^n)$  and

$$\frac{\partial}{\partial t} f_R + \frac{\varepsilon}{2} \Delta f_R + \frac{1}{2} |\nabla f_R|^2 = 0,$$

so the result follows by the dominated convergence Theorem. ■

## 4.2 Proof of the main Theorem

We are now able to prove the equivalence between the optimal relative entropy and the kinetic energy provided by the Fokker-Planck pair  $(\mu_t^\varepsilon, \nabla \psi_t^\varepsilon)$  defined in (4) and (5). The proof relies mostly on a derivation under the integral sign and on an application of Gauss–Green’s formula.

**Proposition 4.6.** *Let  $\mu_0$  and  $\mu_1$  satisfying Assumptions 3.3. Then, for every  $\varepsilon > 0$ ,*

$$\varepsilon H(\gamma^\varepsilon | \mathbb{R}_\varepsilon) = \varepsilon H(\mu_0 | \mathcal{L}^n) + \int_0^1 \int_0^1 \frac{|\nabla \psi_t^\varepsilon|^2}{2} d\mu_t^\varepsilon dt.$$

*Proof.* First of all, using the marginal constraints on  $\gamma^\varepsilon$ ,

$$H(\gamma^\varepsilon | \mathbb{R}_\varepsilon) = \int \log(f^\varepsilon \otimes g^\varepsilon) d\gamma^\varepsilon = \int \log f^\varepsilon d\mu_0 + \int \log g^\varepsilon d\mu_1.$$

Now, by Lemma 4.2 and Proposition 4.4, for every  $t \in [0, 1]$ , we can argue that the function  $x \mapsto \psi_t^\varepsilon(x) \varrho_t^\varepsilon(x)$  is integrable and that, for every  $x \in \mathbb{R}$ , the function  $t \mapsto$

$\psi_t^\varepsilon(x)\varrho_t^\varepsilon(x)$  is differentiable, with  $|\partial_t(\psi_t^\varepsilon(x)\varrho_t^\varepsilon(x))|$  dominated by an integrable function, uniformly in  $t$ . We can thus differentiate under the integral sign, finding

$$\frac{d}{dt} \int \psi_t^\varepsilon d\mu_t^\varepsilon = \int \frac{\partial}{\partial t}(\psi_t^\varepsilon \varrho_t^\varepsilon) d\mathcal{L}^n = \int \left( \frac{\partial}{\partial t} \psi_t^\varepsilon \varrho_t^\varepsilon + \psi_t^\varepsilon \frac{\partial}{\partial t} \varrho_t^\varepsilon \right) d\mathcal{L}^n,$$

but, using the equations that  $\psi_t$  and  $\varrho_t$  solve,

$$\begin{aligned} \frac{d}{dt} \int \psi_t^\varepsilon d\mu_t^\varepsilon &= -\frac{\varepsilon}{2} \int \Delta \psi_t^\varepsilon \varrho_t^\varepsilon d\mathcal{L}^n - \frac{1}{2} \int |\nabla \psi_t^\varepsilon|^2 \varrho_t^\varepsilon d\mathcal{L}^n \\ &\quad - \frac{\varepsilon}{2} \int \psi_t^\varepsilon \Delta \varrho_t^\varepsilon d\mathcal{L}^n - \int \psi_t^\varepsilon \operatorname{div}(\varrho_t^\varepsilon \nabla \psi_t^\varepsilon) d\mathcal{L}^n. \end{aligned}$$

Then, according to Gauss–Green’s Formula, thanks to Lemma 4.2 and Proposition 4.4,

$$\frac{d}{dt} \int \psi_t^\varepsilon d\mu_t^\varepsilon = \int \frac{|\nabla \psi_t^\varepsilon|^2}{2} \varrho_t^\varepsilon d\mathcal{L}^n.$$

Integrating both sides with respect to  $t$  we find,

$$\int \psi_1^\varepsilon d\mu_1 - \int \psi_0^\varepsilon d\mu_0 = \int_0^1 \int \frac{|\nabla \psi_t^\varepsilon|^2}{2} \varrho_t^\varepsilon d\mathcal{L}^n dt.$$

Noticing that  $\psi_0^\varepsilon = \varepsilon \log \varrho_0 - \varepsilon \log f^\varepsilon$  and  $\psi_1^\varepsilon = \varepsilon \log g^\varepsilon$ , the result follows. ■

Thanks to the previous Proposition, we get one inequality for our main Theorem. The other is obtained by duality.

*Proof of Theorem 3.6.* Consider  $(\eta_t \mathcal{L}^n, v_t)$  a Fokker–Planck pair such that  $\eta_0 = \varrho_0, \eta_1 = \varrho_1$  and  $f \in C^1([0, 1]) \cap C_b^2(\mathbb{R}^n)$  such that

$$\frac{\partial}{\partial t} f + \frac{\varepsilon}{2} \Delta f + \frac{1}{2} |\nabla f|^2 \leq 0$$

We have, thanks to Remark 2.5,

$$\begin{aligned} &\int_0^1 \int \frac{|v_t|^2}{2} \eta_t d\mathcal{L}^n dt = \int_0^1 \int \frac{|v_t|^2}{2} \eta_t d\mathcal{L}^n dt \\ &\quad - \int_0^1 \int \left( \frac{\partial}{\partial t} f + \frac{\varepsilon}{2} \Delta f + v_t \cdot \nabla f \right) \eta_t d\mathcal{L}^n dt + \int f_1 \eta_1 d\mathcal{L}^n - \int f_0 \eta_0 d\mathcal{L}^n \\ &\geq - \int_0^1 \int \left( \frac{\partial}{\partial t} f + \frac{\varepsilon}{2} \Delta f + \frac{1}{2} |\nabla f|^2 \right) \eta_t d\mathcal{L}^n dt + \int f_1 d\mu_1 - \int f_0 d\mu_0 \\ &\geq \int f_1 d\mu_1 - \int f_0 d\mu_0. \end{aligned}$$

Passing to the supremum with respect to  $f \in C^1([0, 1]) \cap C_b^2(\mathbb{R}^n)$  and to the infimum

with respect to  $(\eta_t \mathcal{L}^n, v_t)$ ,

$$\begin{aligned} & \inf \left\{ \int_0^1 \int \frac{|v_t|^2}{2} \eta_t d\mathcal{L}^n dt : (\eta_t \mathcal{L}^n, v_t) \text{ Fokker-Planck pair, } \eta_0 = \varrho_0, \eta_1 = \varrho_1 \right\} \\ & \geq \sup \left\{ \int f_1 d\mu_1 - \int f_0 d\mu_0 : f \in C^1([0, 1]) \cap C_b^2(\mathbb{R}^n), \frac{\partial}{\partial t} f + \frac{\varepsilon}{2} \Delta f + \frac{1}{2} |\nabla f|^2 \leq 0 \right\}. \end{aligned}$$

By Lemma 4.5,  $\psi_t^\varepsilon$  is an admissible competitor in the maximization problem, namely

$$\begin{aligned} & \inf \left\{ \int_0^1 \int \frac{|v_t|^2}{2} \eta_t d\mathcal{L}^n dt : (\eta_t \mathcal{L}^n, v_t) \text{ Fokker-Planck pair, } \eta_0 = \varrho_0, \eta_1 = \varrho_1 \right\} \\ & \geq \int \psi_1^\varepsilon d\mu_1 - \int \psi_0^\varepsilon d\mu_0. \end{aligned}$$

Observing that  $\psi_0^\varepsilon = \varepsilon \log \varrho_0 - \varepsilon \log f^\varepsilon$  and  $\psi_1^\varepsilon = \varepsilon \log g^\varepsilon$ , we get, by Propositions 3.5 and 4.6,

$$\begin{aligned} & \inf \left\{ \int_0^1 \int \frac{|v_t|^2}{2} \eta_t d\mathcal{L}^n dt : (\eta_t \mathcal{L}^n, v_t) \text{ Fokker-Planck pair, } \eta_0 = \varrho_0, \eta_1 = \varrho_1 \right\} \\ & \geq \int \varepsilon \log g^\varepsilon d\mu_1 + \int \varepsilon \log f^\varepsilon d\mu_0 - \varepsilon H(\mu_0 | \mathcal{L}^n) = \varepsilon H(\gamma^\varepsilon | \mathbb{R}_\varepsilon) - \varepsilon H(\mu_0 | \mathcal{L}^n) \\ & = \varepsilon \min_{\gamma \in \Gamma(\mu_0, \mu_1)} H(\gamma | \mathbb{R}_\varepsilon) - \varepsilon H(\mu_0 | \mathcal{L}^n) = \int_0^1 \int \frac{|\nabla \psi_t^\varepsilon|^2}{2} d\mu_t^\varepsilon dt \\ & = \inf \left\{ \int_0^1 \int \frac{|v_t|^2}{2} \eta_t d\mathcal{L}^n dt : (\eta_t \mathcal{L}^n, v_t) \text{ Fokker-Planck pair, } \eta_0 = \varrho_0, \eta_1 = \varrho_1 \right\}. \end{aligned}$$

Then,

$$\begin{aligned} & \varepsilon \min_{\gamma \in \Gamma(\mu_0, \mu_1)} H(\gamma | \mathbb{R}_\varepsilon) = \varepsilon H(\mu_0 | \mathcal{L}^n) \\ & + \min \left\{ \int_0^1 \int \frac{|v_t|^2}{2} \eta_t d\mathcal{L}^n dt : (\eta_t \mathcal{L}^n, v_t) \text{ Fokker-Planck pair, } \eta_0 = \varrho_0, \eta_1 = \varrho_1 \right\}. \end{aligned}$$

As regards the uniqueness, consider the change of variables  $q_t = \eta_t v_t$ , the set

$$\mathcal{C} = \{(\eta, q) : (\eta_t \mathcal{L}^n, v_t) \text{ is a Fokker-Planck pair, } \eta_0 \mathcal{L}^n = \mu_0, \eta_1 \mathcal{L}^n = \mu_1\},$$

the convex and lower semi-continuous function  $\Theta : [0, +\infty[ \times \mathbb{R} \rightarrow [0, +\infty]$  such that

$$\Theta(x, y) = \begin{cases} \frac{y^2}{x} & \text{if } x > 0, \\ 0 & \text{if } x = y = 0, \\ +\infty & \text{otherwise,} \end{cases}$$

and the convex functional  $\mathfrak{F} : \mathcal{C} \rightarrow [0, +\infty]$  such that

$$\mathfrak{F}(\eta, q) = \int_0^1 \int \frac{1}{2} \Theta(\eta_t, |q_t|) d\mathcal{L}^n dt = \int_0^1 \int \frac{|v_t|^2}{2} \eta_t d\mathcal{L}^n dt.$$

Take another minimizer  $(\bar{\eta}_t, \bar{v}_t)$  for kinetic energy minimization and set  $\bar{q} = \bar{\eta}\bar{v}$ . By minimality,  $\mathfrak{F}(\bar{\eta}_t, \bar{q}_t) = \mathfrak{F}(\varrho_t^\varepsilon, \varrho_t^\varepsilon \nabla \psi_t^\varepsilon)$  so, given  $\lambda \in ]0, 1[$  and setting

$$\eta_t^\lambda = (1 - \lambda)\eta_t + \lambda\varrho_t^\varepsilon, \quad q_t^\lambda = (1 - \lambda)q_t + \lambda\varrho_t^\varepsilon \nabla \psi_t^\varepsilon,$$

we get

$$\mathfrak{F}(\eta_t^\lambda, q_t^\lambda) \geq \mathfrak{F}(\varrho_t^\varepsilon, \varrho_t^\varepsilon \nabla \psi_t^\varepsilon) = (1 - \lambda)\mathfrak{F}(\bar{\eta}_t, \bar{q}_t) + \lambda\mathfrak{F}(\varrho_t^\varepsilon, \varrho_t^\varepsilon \nabla \psi_t^\varepsilon)$$

and, using also the convexity of  $\mathfrak{F}$  to obtain the converse inequality, we arrive to

$$\mathfrak{F}(\eta_t^\lambda, q_t^\lambda) = (1 - \lambda)\mathfrak{F}(\bar{\eta}_t, \bar{q}_t) + \lambda\mathfrak{F}(\varrho_t^\varepsilon, \varrho_t^\varepsilon \nabla \psi_t^\varepsilon),$$

so  $(\eta_t^\lambda, q_t^\lambda)$  also minimizes  $\mathfrak{F}$  for every  $\lambda \in [0, 1]$ . That means that  $\mathfrak{F}$  is constant on the line connecting  $(\varrho_t^\varepsilon, \varrho_t^\varepsilon \nabla \psi_t^\varepsilon)$  and  $(\bar{\eta}_t, \bar{q}_t)$  and (since we are in the case  $x > 0$ ) this permits us to say that for every  $\lambda \in [0, 1]$ ,

$$\Theta(\eta_t^\lambda, q_t^\lambda) = (1 - \lambda)\Theta(\bar{\eta}_t, \bar{q}_t) + \Theta(\varrho_t^\varepsilon, \varrho_t^\varepsilon \nabla \psi_t^\varepsilon).$$

But the previous implies that there exists  $\alpha(t, x) > 0$  such that

$$(\bar{\eta}_t, \bar{q}_t) = \alpha(t, x)(\varrho_t^\varepsilon, \varrho_t^\varepsilon \nabla \psi_t^\varepsilon),$$

so, using the fact that both  $(\bar{\eta}_t, \bar{v}_t)$  and  $(\varrho_t^\varepsilon, \varrho_t^\varepsilon \nabla \psi_t^\varepsilon)$  are Fokker–Planck pairs, we get, weakly,

$$\frac{\partial}{\partial t} \alpha_t + \nabla \alpha_t \cdot \nabla \psi_t^\varepsilon = 0.$$

By the fact that both  $\bar{\eta}_t$  and  $\varrho_t^\varepsilon$  are probability measures, we get that  $\alpha_0, \alpha_1 = 1$  a.e. in  $\mathbb{R}^n$ . In other words,  $\alpha$  is a weak solution of

$$\begin{cases} \frac{\partial}{\partial t} \alpha_t + \nabla \alpha_t \cdot \nabla \psi_t^\varepsilon = 0 & \text{in } ]0, 1[ \times \mathbb{R}^n, \\ \alpha_0 = 1 & \text{on } \mathbb{R}^n, \\ \alpha_1 = 1 & \text{on } \mathbb{R}^n, \end{cases}$$

and, since the characteristics method implies the uniqueness of the solution, we get  $\alpha = 1$ . It is now sufficient to use the strict convexity of the map  $v_t \mapsto \int_0^1 \int \frac{|v_t|^2}{2} \varrho_t^\varepsilon d\mathcal{L}^n dt$  to conclude. ■

**Acknowledgment** L.N. and M.G. benefited from the support of the FMJH Program PGMO and from the ANR project GOTA (ANR-23-CE46-0001). While this work was written, L.T. was associated to INdAM (Istituto Nazionale di Alta Matematica

“Francesco Severi”) and the group GNAMPA.

M. G. is funded by the European Union and, as part of the MSCA program, is also supported, as secondment, by Laboratoire Jean Alexandre Dieudonné, Université Côte d’Azur, CNRS. This project has received funding from the European Union’s MSCA–Horizon Europe, grant agreement No 101126554. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.



Co-funded by  
the European Union

This work is licensed under Creative Commons Attribution 4.0 International. To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>.

## References

- [ABS21] L. Ambrosio, E. Brué, and D. Semola. *Lectures on optimal transport*. Vol. 130. Springer, 2021.
- [Ada+13] S. Adams, N. Dirr, M. Peletier, and J. Zimmer. “Large deviations and gradient flows”. In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 371.2005 (2013), p. 20120341.
- [AGS05] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient flows: in metric spaces and in the space of probability measures*. Springer, 2005.
- [BB00] J.-D. Benamou and Y. Brenier. “A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem”. In: *Numerische Mathematik* 84.3 (2000), pp. 375–393.
- [Ben+15] J.-D. Benamou, G. Carlier, M. Cuturi, L. Nenna, and G. Peyré. “Iterative Bregman projections for regularized transportation problems”. In: *SIAM Journal on Scientific Computing* 37.2 (2015), A1111–A1138.
- [BF81] H. Brézis and A. Friedman. *Nonlinear Parabolic Equations Involving Measures as Initial Conditions*. Tech. rep. 1981.
- [BLN94] J. M. Borwein, A. S. Lewis, and R. D. Nussbaum. “Entropy minimization, DAD problems, and doubly stochastic kernels”. In: *Journal of Functional Analysis* 123.2 (1994), pp. 264–307.
- [Chi+20] L. Chizat, P. Roussillon, F. Léger, F.-X. Vialard, and G. Peyré. “Faster Wasserstein distance estimation with the Sinkhorn divergence”. In: *Advances in neural information processing systems* 33 (2020), pp. 2257–2269.

- [Chi+23] A. Chiarini, G. Conforti, G. Greco, and L. Tamanini. “Gradient estimates for the Schrödinger potentials: convergence to the Brenier map and quantitative stability”. In: *Communications in Partial Differential Equations* 48.6 (2023), pp. 895–943.
- [Chi+24] A. Chiarini, G. Conforti, G. Greco, and L. Tamanini. “A semiconcavity approach to stability of entropic plans and exponential convergence of Sinkhorn’s algorithm”. In: *preprint arXiv: 2412.09235* (2024).
- [CL95] P. Cattiaux and C. Léonard. “Large deviations and Nelson processes”. In: *Forum mathematicum*. Vol. 7. 1. 1995, pp. 95–116.
- [Con24] G. Conforti. “Weak semiconvexity estimates for Schrödinger potentials and logarithmic Sobolev inequality for Schrödinger bridges”. In: *Probability Theory and Related Fields* 189.3 (2024), pp. 1045–1071.
- [CPT23] G. Carlier, P. Pegon, and L. Tamanini. “Convergence rate of general entropic optimal transport costs”. In: *Calculus of Variations and Partial Differential Equations* 62.4 (2023), p. 116.
- [CT21] G. Conforti and L. Tamanini. “A formula for the time derivative of the entropic cost and applications”. In: *Journal of Functional Analysis* 280.11 (2021), p. 108964.
- [Cut13] M. Cuturi. “Sinkhorn distances: Lightspeed computation of optimal transport”. In: *Advances in neural information processing systems* 26 (2013).
- [EMR15] M. Erbar, J. Maas, and D. R. M. Renger. “From large deviations to Wasserstein gradient flows in multiple dimensions”. In: *Electron. Commun. Probab.* 20 (2015), no. 89, 12. ISSN: 1083-589X. DOI: [10.1214/ECP.v20-4315](https://doi.org/10.1214/ECP.v20-4315).
- [EN24] S. Eckstein and M. Nutz. “Convergence rates for regularized optimal transport via quantization”. In: *Mathematics of Operations Research* 49.2 (2024), pp. 1223–1240.
- [FG97] H. Föllmer and N. Gantert. “Entropy minimization and Schrödinger processes in infinite dimensions”. In: *The annals of probability* 25.2 (1997), pp. 901–926.
- [GLR17] I. Gentil, C. Léonard, and L. Ripani. “About the analogy between optimal transport and minimal entropy”. In: *Annales de la Faculté des sciences de Toulouse: Mathématiques*. Vol. 26. 3. 2017, pp. 569–600.
- [GT20] N. Gigli and L. Tamanini. “Benamou–Brenier and duality formulas for the entropic cost on  $RCD^*(K, N)$  spaces”. In: *Probability Theory and Related Fields* 176.1 (2020), pp. 1–34.
- [GT21] N. Gigli and L. Tamanini. “Second order differentiation formula on  $RCD^*(K, N)$  spaces”. In: *Journal of the European Mathematical Society* 23.5 (2021), pp. 1727–1795.

- [GT25] G. Greco and L. Tamanini. “Hessian stability and convergence rates for entropic and Sinkhorn potentials via semiconcavity”. In: *preprint arXiv:2504.11133* (2025).
- [Kat72] T. Kato. “Schrödinger operators with singular potentials”. In: *Israel Journal of Mathematics* 13.1 (1972), pp. 135–148.
- [Léo01] C. Léonard. “Minimizers of energy functionals”. In: *Acta Mathematica Hungarica* 93.4 (2001), pp. 281–325.
- [Léo12] C. Léonard. “From the Schrödinger problem to the Monge–Kantorovich problem”. In: *Journal of Functional Analysis* 262.4 (2012), pp. 1879–1920.
- [Léo14] C. Léonard. “A survey of the Schrödinger problem and some of its connections with optimal transport”. In: *Discrete Contin. Dyn. Syst.* 34.4 (2014), pp. 1533–1574. ISSN: 1078-0947,1553-5231. DOI: 10.3934/dcds.2014.34.1533.
- [MGM22] A. Mallasto, A. Gerolin, and H. Q. Minh. “Entropy-regularized 2-Wasserstein distance between Gaussian measures”. In: *Information Geometry* 5.1 (2022), pp. 289–323.
- [Mik04] T. Mikami. “Monge’s problem with a quadratic cost by the zero-noise limit of h-path processes”. In: *Probability theory and related fields* 129.2 (2004), pp. 245–260.
- [MT08] T. Mikami and M. Thieullen. “Optimal transportation problem by stochastic optimal control”. In: *SIAM Journal on Control and Optimization* 47.3 (2008), pp. 1127–1139.
- [NP25] L. Nenna and P. Pegon. “Convergence rate of entropy-regularized multi-marginal optimal transport costs”. In: *Canadian Journal of Mathematics* 77.3 (2025), pp. 1072–1092.
- [NPT25] L. Nenna, P. Pegon, and L. Tocquec. “Convergence rates for regularized unbalanced optimal transport: the discrete case”. In: *arXiv preprint arXiv:2507.07917* (2025).
- [Nut21] M. Nutz. “Introduction to entropic optimal transport”. In: *Lecture notes, Columbia University* (2021).
- [Pal24] S. Pal. “On the difference between entropic cost and the optimal transport cost”. In: *The Annals of Applied Probability* 34.1B (2024), pp. 1003–1028.
- [RT93] L. Rüschendorf and W. Thomsen. “Note on the Schrödinger equation and I-projections”. In: *Statistics & probability letters* 17.5 (1993), pp. 369–375.
- [Rüs95] L. Rüschendorf. “Convergence of the iterative proportional fitting procedure”. In: *The Annals of Statistics* (1995), pp. 1160–1174.
- [San15] F. Santambrogio. *Optimal Transport for Applied Mathematicians: Calculus of Variations, PDEs, and Modeling*. Vol. 87. Progress in Nonlinear Differential Equations and Their Applications. Birkhäuser/Springer, Cham, 2015.

- [Sch31] E. Schrödinger. *Über die umkehrung der naturgesetze*. Verlag der Akademie der Wissenschaften in Kommission bei Walter De Gruyter u. Company, 1931.
- [Sch32] E. Schrödinger. “Sur la théorie relativiste de l’électron et l’interprétation de la mécanique quantique”. In: *Annales de l’institut Henri Poincaré*. Vol. 2. 4. 1932, pp. 269–310.
- [Sin64] R. Sinkhorn. “A relationship between arbitrary positive matrices and doubly stochastic matrices”. In: *The annals of mathematical statistics* 35.2 (1964), pp. 876–879.
- [Sin67] R. Sinkhorn. “Diagonal equivalence to matrices with prescribed row and column sums”. In: *The American Mathematical Monthly* 74.4 (1967), pp. 402–405.
- [SK67] R. Sinkhorn and P. Knopp. “Concerning nonnegative matrices and doubly stochastic matrices”. In: *Pacific Journal of Mathematics* 21.2 (1967), pp. 343–348.
- [Vil08] C. Villani. *Optimal transport: old and new*. Vol. 338. Springer, 2008.