The Schrödinger problem

The Schrödinger problem

where analysis meets stochastics

Giacomo Greco

Printed by ADC Nederland

Cover generated with the AI Image Generator *deepai.org* by the author's prompt "black cat on top of an iron bridge, painted by Van Gogh".

A catalogue record is available from the Eindhoven University of Technology Library

ISBN: 978-90-386-6031-8

The Schrödinger problem

where analysis meets stochastics

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de
Technische Universiteit Eindhoven, op gezag
van de rector magnificus prof.dr. S. K. Lenaerts, voor een
commissie aangewezen door het College voor Promoties, in het
openbaar te verdedigen op donderdag 23 Mei 2024 om 16:00 uur

door

Giacomo Greco

geboren te Rome, Italië

Dit proefschrift is goedgekeurd door de promotoren en de samenstelling van de promotiecommissie is als volgt:

voorzitter:	prof.dr.ir. S. C. Borst
promotor:	prof.dr. M. A. Peletier
copromotor:	prof.dr. A. Chiarini (Università degli Studi di Padova)
copromotor:	prof.dr. G. Conforti (Università degli Studi di Padova - École Polytechnique IP Paris)
leden:	prof.dr. G. Carlier (Université Paris Dauphine)
	prof.dr. M. Erbar (Universität Bielefeld)
	prof.dr.ir. B. Koren
	dr. R. C. Kraaij (Technische Universiteit Delft)
	prof.dr. O. Mula Hernández

Het onderzoek dat in dit proefschrift wordt beschreven is uitgevoerd in overeenstemming met de TU/e Gedragscode Wetenschapsbeoefening.

Laugh hard
Run fast
Be kind

Twelfth Doctor

We all change, when you think about it.
We're all different people all through our lives.
And that's OK, that's good,
you gotta keep moving,
so long as you remember
all the people that you used to be.
I will not forget one line of this, not one day, I swear.

Eleventh Doctor

Siamo come nani
sulle spalle di giganti

Bernardo di Chartres

Ik maak steeds wat ik nog niet kan
om het te leeren kunnen

Vincent van Gogh

This page was intentionally left blank.

Abstract

The Schrödinger problem is a statistical mechanics problem consisting in finding the most likely evolution of a cloud of independent Brownian particles conditionally on the observation of their initial and final configurations. Although this problem had been already introduced back in 1932 by E. Schrödinger, in the past decade it has become increasingly popular thanks to its connections with (computational) optimal transport, statistical mechanics and stochastic optimal control theory.

In this thesis, with the help of stochastic control theory, we show how solutions to the Schrödinger problem provide good proxies for the optimal transport map. Furthermore, we prove the exponential convergence of Sinkhorn's algorithm, which provides an efficient and fast way of explicitly computing such solutions.

Lastly, we introduce an instance of the problem in a hypocoercive setting, where the Brownian particles are replaced by independent particles following the kinetic Fokker-Planck equation, and analyse the long-time behaviour of the most likely evolution of the particle system.

Keywords. Schrödinger problem, entropic optimal transport, stochastic optimal control, large deviations, Sinkhorn's algorithm, hypocoercivity.

MSC2020: 49Q22, 60E15, 34K20, 93E20, 90C25, 47D07, 53C21, 49N05.

This page was intentionally left blank.

Contents

This page was intentionally left blank.

Chapter 1

Introduction

Imagine finding yourself in front of a cloud of particles that evolves over time, you may think of a cloud in the sky or smoke in an empty room. Maybe its strange position and shape got your attention and you decide to take a photo. Intrigued by what you have seen before, after some time, you decide to come back and take another photo. You immediately notice that shape and position do not meet your expectations at all, *i.e.*, gas uniformly spread all over the sky/room, based on your *Statistical Mechanics 101* knowledge.

Then what you might have wondered, at least if you are a mathematician, is "what happened to the cloud and how has it evolved over time so that it ended up in this unexpected configuration?" or maybe you might also wonder "what has lead to this unexpected final configuration?". Since I am not a physicist, I will not seek the cause of this unexpected behaviour. On the contrary, in this thesis we are going to be interested in answering the first question, that is, inferring the most likely behaviour of the random cloud system, conditionally on its initial and final configuration.

This problem is not new at all and it dates back at least to the two seminal papers *Über die Umkehrung der Naturgesetze* [Sch31] and *Sur la théorie relativiste de l'électron et l'interprétation de la mécanique quantique* [Sch32] where Erwin Schrödinger writes

> *"Imaginez que vous observez un système de particules en diffusion, qui soient en équilibre thermodynamique. Admettons qu'à un instant donné t_0 vous les ayez trouvées en répartition à peu près uniforme et qu'à $t_1 > t_0$ vous ayez trouvé un écart spontané et considérable par rapport à cette uniformité. On vous demande de quelle manière cet écart s'est produit. Quelle en est la manière la plus probable?"*[1]

In view of this description, this problem is commonly referred to as the Schrödinger problem (hereafter SP). In the next section we will see that Large

[1] Differently to Schrödinger's own words, in this thesis we will assume that the particles' initial distribution might be different from their thermodynamic equilibrium.

Deviations Theory provides the proper mathematical framework to address Schrödinger's original question. The connection between Large Deviations Theory and SP dates back to [Föl88], where Föllmer rigorously proved that solving SP boils down to solving an entropic minimisation problem on the path space, subject to two marginal constraints.

Despite the problem itself being quite old, it has gained much more attention in the past two decades because of its deep connection with Optimal Transport (OT) theory. This link, first established by Mikami in [Mik04], shows that solutions to SP provide good proxies for optimal plans in the Monge-Kantorovich optimal transport problem. We are going to further investigate this connection in Chapter 3 and prove that the gradient of SP's solutions can be used as proxy for the optimal transport map. Moreover, in the past decade OT has proven to be a useful tool in Machine Learning applications [PC19] and consequently computing solutions to SP in a rapid way has become of primal interest. We are going to show in Chapters 5 and 6 that indeed they can be efficiently computed via an algorithm, known as Sinkhorn's algorithm, which stems at the core of many OT applications in the Machine Learning realm. Lastly, in the last couple of years, SP has found a tremendous use in *generative modelling* [DBTHD21, SDBCD23], namely *creating data from noise*, which has led to an increasing interest in Schrödinger bridges (*i.e.*, SP's solutions) themselves seen as an alternative to *diffusion generative models* and *score matching* [Hyv05].

Aside from its relation with OT theory and ML applications, SP admits a stochastic optimal control formulation [CGP16b], which has led to different practical uses of SP in control engineering [CGP16c, CGP16d]. For these reasons, SP is nowadays a very popular problem at the crossroad of different disciplines. The objective of this thesis can then be summarised as finding new results for SP and its applications to OT, while keeping in mind the stochastic interpretation of the problem. In what follows firstly we are going to derive SP mathematical formulation (cf. (1.1.4)) via Large Deviations Theory, then we are going to show its connection with a regularised version of the Optimal Transport problem and finally we will show that SP admits also a stochastic optimal control formulation. In view of these equivalent formulations, the proof strategies we are going to employ in this thesis will lie at the crossroad between Optimal Transport and Stochastic Optimal Control theories, *where analysis meets stochastics*. Lastly, we conclude the Introduction with a thorough overview of our main contributions.

1.1 From large deviations to the Schrödinger problem

Let us start our discussion by fixing a sequence of independent and identically distributed (i.i.d.) Brownian motions $(B^i)_{i\in\mathbb{N}}$ in the Euclidean space $\mathbb{R}^d$. In this thesis these random variables will represent the particles in our cloud, since in practice we are not able to distinguish each particle from the others. Then, for any fixed $t > 0$ the random variable B^i_t is the position at time t of the i^{th} Brownian particle. Let us further fix a final time $T > 0$ and denote the space of continuous trajectories as $\Omega := \mathcal{C}([0,T],\mathbb{R}^d)$, so that B^i is a random variable taking values in the path space Ω. For later reference, we will denote with $(X_t)_{t\in[0,T]}$ the canonical process on Ω, that is $X_t(\omega) = \omega(t)$ for any $\omega \in \Omega$.

Since we are not able to distinguish particles from each other, any information we can deduce by observing the behaviour of the cloud will be a macroscopic observation. Mathematically speaking this means that any information we can deduce from observations is encoded via the empirical density measure

$$\mathbf{P}^N := \frac{1}{N}\sum_{i=1}^{N} \delta_{B^i} \tag{1.1.1}$$

which is a random variable taking values in the space of probability measures over the path space, *i.e.*, taking values in $\mathcal{P}(\Omega)$. In the subsequent discussion, let Prob denote the law of this random variable, or equivalently that $\mathbf{P}^N \sim \mathrm{Prob}$.

Then, if the two probability distributions μ, $\nu \in \mathcal{P}(\mathbb{R}^d)$ encode the unlikely behaviour we have observed at the initial and final time, from a mathematical point of view we would say that for N large (*i.e.*, in the *many-particles limit*)

$$\mathbf{P}^N_0 = \frac{1}{N}\sum_{i=1}^{N} \delta_{B^i_0} \approx \mu \quad \text{and} \quad \mathbf{P}^N_T = \frac{1}{N}\sum_{i=1}^{N} \delta_{B^i_T} \approx \nu. \tag{1.1.2}$$

This constraint can formally be translated as saying that the empirical density measure $\mathbf{P}^N$, in the many-particles limit, does not take values on the whole probability space $\mathcal{P}(\Omega)$, but rather in its subset of constrained path-space measures

$$\left\{\mathrm{P} \in \mathcal{P}(\Omega) \colon \mathrm{P}_0 = \mu \text{ and } \mathrm{P}_T = \nu\right\}, \tag{1.1.3}$$

where P_0, P_T are respectively the pushforward measures $(X_0)_\#\mathrm{P}$ and $(X_T)_\#\mathrm{P}$.

Given the above premises, let us try now to infer the most likely behaviour of the cloud of Brownian motions. Without imposing any marginal constraint, the Law of Large Numbers guarantees the convergence of the sequence of random variables $\mathbf{P}^N$ to the random variable constantly equal to $\mathrm{R} \in \mathcal{M}(\Omega)$, that is the Wiener measure on Ω (or equivalently the law of a stationary Brownian motion B^i on Ω). This means that without marginal constraints $\mathrm{Prob} \approx \delta_{\mathrm{R}}$ for N large enough. Then, in order to take into account the fact that $\mathbf{P}^N$ takes values in the constrained set (1.1.3), we might rely on Sanov's Theorem [DZ10, Theorem 6.2.10], whose core message is that the likelihood of a given evolution P is measured through the relative entropy with respect to the Wiener measure R (*i.e.*, the unconstrained limit, according to the Law of Large Numbers). More precisely, for any nice subset $\mathsf{A} \subseteq \mathcal{P}(\Omega)$, for N large enough we have

$$\mathrm{Prob}\left[\mathbf{P}^N \in \mathsf{A}\right] \approx \exp\left(-N \inf_{\mathrm{P}\in\mathsf{A}} \mathscr{H}(\mathrm{P}|\mathrm{R})\right),$$

where the $\mathscr{H}(\cdot|\mathrm{R})$ denotes the relative entropy functional with respect to R. We refer the reader to Section 1.A for its formal definition and its properties. In the Large Deviations Theory own words, the above identity states that $\mathbf{P}^N$ satisfies a Large Deviation Principle with rate function $\mathscr{H}(\cdot|\mathrm{R})$.

Then, the *most likely evolution* of our cloud particles system can be found by minimising the rate function on the constrained set (1.1.3), that is solving the minimisation problem

$$\inf_{\{\mathrm{P}\in\mathcal{P}(\Omega)\colon\ \mathrm{P}_0=\mu,\,\mathrm{P}_T=\nu\}} \mathscr{H}(\mathrm{P}|\mathrm{R})\ . \tag{1.1.4}$$

We will refer to the above problem as to the *dynamical* Schrödinger problem, since the minimisation takes place on the set of probability measures on the path space. Under mild assumptions on the marginals μ, ν, in Chapter 2 we will see that the above problem admits a unique solution, which we will denote as $\mathbf{P}^T$ and we will refer to it as to the *Schrödinger bridge* from μ to ν (in time T). In the rest of this introduction we tacitly assume the existence and uniqueness of the Schrödinger bridge $\mathbf{P}^T$.

1.2 From the Schrödinger problem to Optimal Transport

Remarkably, the previous dynamical problem (c.f. (1.1.4)), whose motivation comes from statistical mechanics, is equivalent to a regularised version of the classic optimal transportation problem. In this section we will explicit this equivalence, by firstly considering a static formulation of SP.

Static Schrödinger problem. Consider the measurable map $(X_0, X_T)\colon \Omega \to \mathbb{R}^d \times \mathbb{R}^d$ which associates to any trajectory $\omega \in \Omega$ the joint vector of starting and terminal positions, *i.e.*, $(X_0, X_T)(\omega) = (\omega_0, \omega_T)$. Then the additive property of the relative entropy (1.A.4) reads as

$$\mathscr{H}(\mathrm{P}|\mathrm{R}) = \mathscr{H}(\mathrm{P}_{0,T}|\mathrm{R}_{0,T}) + \int \mathscr{H}(\mathrm{P}^{xy}|\mathrm{R}^{xy})\,\mathrm{dP}_{0,T}(x,y)\ , \tag{1.2.1}$$

where $\mathrm{P}_{0,T} := (X_0, X_T)_\#\mathrm{P}$ is the joint law of the random vector (X_0, X_T) under the law P (and similarly for $\mathrm{R}_{0,T}$), whereas $\mathrm{P}^{xy} := \mathrm{P}(\cdot|(X_0, X_T) = (x,y)) \in \mathcal{P}(\Omega)$ is the conditional law of the random variable X conditioned to its initial and final positions, also known as P-bridge from x to y, and similarly for R^{xy} which is also known as the Brownian bridge from x to y.

Notice that the last integral term in the above identity does not depend on the initial and final marginal law of P. Therefore, minimising the above left hand side over the subset $\{\mathrm{P} \in \mathcal{P}(\Omega)\colon \mathrm{P}_0 = \mu,\, \mathrm{P}_T = \nu\}$ is equivalent to minimising the integral term over all possible bridges P^{xy} (which is an unconstrained problem) while independently solving the finite dimensional minimisation problem

$$\inf_{\pi\in\Pi(\mu,\nu)} \mathscr{H}(\pi|\mathrm{R}_{0,T}) \ . \tag{1.2.2}$$

Here $\Pi(\mu,\nu)$ is the set of couplings between μ and ν, *i.e.*, the set of probability measures on $\mathbb{R}^{2d}$ satisfying $\pi(\mathsf{A}\times\mathbb{R}^d) = \mu(\mathsf{A})$ and $\pi(\mathbb{R}^d\times\mathsf{A}) = \nu(\mathsf{A})$, for any measurable subset $\mathsf{A}\subseteq\mathbb{R}^d$. This problem is known as the *static Schrödinger problem* and for now let us assume that this static problem admits a unique solution, to which we will refer to as the *Schrödinger plan* and denote it as $\boldsymbol{\pi}^T\in\mathcal{P}(\mathbb{R}^d\times\mathbb{R}^d)$.

We observe that the minimisation problem concerning the integral term in (1.2.1) is unconstrained and the relative entropies in the integral are all nonnegative and vanish if and only if $\mathrm{P}^{xy} = \mathrm{R}^{xy}$ (cf. Lemma 1.A.1, since $\mathrm{R}^{xy}\in\mathcal{P}(\Omega)$). Therefore, the integral term achieves its minimum (*i.e.*, it is null) when $\mathrm{P}^{xy} = \mathrm{R}^{xy}$ for $\boldsymbol{\pi}^T$-a.e. $x,y\in\mathbb{R}^d$; henceforth we have shown that

$$\inf_{\{\mathrm{P}\in\mathcal{P}(\Omega):\ \mathrm{P}_0=\mu,\,\mathrm{P}_T=\nu\}} \mathscr{H}(\mathrm{P}|\mathrm{R}) = \inf_{\pi\in\Pi(\mu,\nu)} \mathscr{H}(\pi|\mathrm{R}_{0,T}) \ . \tag{1.2.3}$$

Moreover, the solutions to the minimisation problems in (1.2.3) satisfy

$$(X_0,X_T)_{\#}\mathbf{P}^T = \boldsymbol{\pi}^T \quad \text{and} \quad \mathbf{P}^T(\cdot|X_0=x,\,X_T=y) = \mathrm{R}^{xy} \quad \boldsymbol{\pi}^T\text{-a.s.}\,,$$

or equivalently, that for any Borel set $\mathsf{A}\subseteq\Omega$ it holds

$$\mathbf{P}^T(\mathsf{A}) = \int_{\mathbb{R}^{2d}} \mathrm{R}^{xy}(\mathsf{A})\,\mathrm{d}\boldsymbol{\pi}^T(x,y) \ .$$

The above relation can be interpreted as follows: sampling the most likely evolution of the Brownian cloud (*i.e.*, sampling from the Schrödinger bridge $\mathbf{P}^T$) is equivalent to sampling a *travel plan*, *i.e.*, jointly sampling the couple (x,y) (initial and terminal point) according to the Schrödinger plan $\boldsymbol{\pi}^T$ and then connecting the two extremes x and y via the standard Brownian bridge R^{xy} between initial and final position.

Entropic Optimal Transport. We will now explain a connection of SP with a certain regularisation of OT, known as Entropic Optimal Transport (EOT). This is immediate once we recall that the density (w.r.t. Lebesgue measure) for $\mathrm{R}_{0,T}$ (*i.e.*, the joint law at time 0 and T of a stationary Brownian motion) reads as

$$\mathrm{R}_{0,T}(\mathrm{d}x,\mathrm{d}y) = (2\pi T)^{-d/2}\,\exp(-|x-y|^2/_{2T})\,\mathrm{d}x\mathrm{d}y \ . \tag{1.2.4}$$

Indeed, after some algebraic manipulations one can easily deduce that for any $\pi\in\Pi(\mu,\nu)$ we have

$$\mathscr{H}(\pi|\mathrm{R}_{0,T}) - \mathrm{Ent}(\mu) - \mathrm{Ent}(\nu) - \frac{d}{2}\log(2\pi T) = \int \frac{|x-y|^2}{2T}\mathrm{d}\pi + \mathscr{H}(\pi|\mu\otimes\nu)\,, \tag{1.2.5}$$

with $\mathrm{Ent}(\cdot)$ being the relative entropy w.r.t. the Lebesgue measure. Therefore, solving (1.2.2) is equivalent to solving

$$\inf_{\pi\in\Pi(\mu,\nu)} \int \frac{|x-y|^2}{2}\mathrm{d}\pi + T\,\mathscr{H}(\pi|\mu\otimes\nu)\,, \tag{1.2.6}$$

which is known as *Entropic Optimal Transport problem*, since it is related to the Kantorovich formulation of the optimal transportation problem (OT), namely

$$\inf_{\pi\in\Pi(\mu,\nu)} \int \frac{|x-y|^2}{2}\mathrm{d}\pi\,, \tag{1.2.7}$$

to which we have added the relative entropy term $\mathscr{H}(\pi|\mu\otimes\nu)$ as a regularising term with the time-horizon $T>0$ seen as a regularising parameter. Moreover, the EOT formulation (1.2.6) clearly suggests its convergence as $T\downarrow 0$ towards the Kantorovich-OT problem (1.2.7) and hence that (up to a rescaling factor $T>0$) SP converges to OT. In light of this and in analogy with the notation adopted for the SP plan $\boldsymbol{\pi}^T$, we will use the symbol $\boldsymbol{\pi}^0$ to denote a OT plan (*i.e.*, a optimiser in (1.2.7)).

Optimal Transport. The first formulation of OT dates back to the seminal work of Gaspard Monge [Mon81], where a worker faces the problem of moving a pile of sand to a prescribed location, shaping it into a different configuration while at the same time trying to minimise the cost of his efforts. This cost could be either the fuel consumed by the construction trucks, the time spent on the job, or the distance between the two configurations. In this thesis we will be mostly interested in the (squared) distance cost. Mathematically speaking Monge's problem can be stated as minimising

$$\inf_{\mathcal{T}} \int \frac{|x-\mathcal{T}(x)|^2}{2}\,\mathrm{d}\mu\,, \tag{1.2.8}$$

where the infimum runs over the set of all measurable functions $\mathcal{T}$ that transport the distribution μ into the target distribution ν, *i.e.*, such that $\mathcal{T}_\#\mu=\nu$. If the above infimum is attained at $\boldsymbol{\mathcal{T}}$ we will refer to it as to the optimal transport map, or as the *Brenier map* (for the specific case of the squared distance cost).

Clearly the Kantorovich formulation introduced in (1.2.7) is a relaxation of Monge's original problem (1.2.8) since for any transport map $\mathcal{T}$ one could consider the induced coupling $\pi_{\mathcal{T}} := (\mathrm{Id}\times\mathcal{T})_\#\mu \in \Pi(\mu,\nu)$. Furthermore, a key property of the Kantorovich problem (1.2.7) is that it admits a dual formulation, often referred to as to *Kantorovich duality*, and it states that

$$\inf_{\pi\in\Pi(\mu,\nu)} \int \frac{|x-y|^2}{2}\mathrm{d}\pi = \sup_{\varphi\in\mathrm{L}^1(\mu),\,\psi\in\mathrm{L}^1(\nu)\,:\,\varphi\oplus\psi\le\frac{1}{2}\mathsf{d}^2(\cdot,\cdot)} \left(\int\varphi\,\mathrm{d}\mu+\int\psi\,\mathrm{d}\nu\right)\,. \tag{1.2.9}$$

Under mild assumptions, the optimisation problem on the right hand side of (1.2.9) admits optimal solutions (φ^0, ψ^0). These two functions are known as *Kantorovich potentials*. We will later see that also SP admits a similar dual formulation with a corresponding pair of (Schrödinger) potentials.

Let us conclude this section by mentioning a cornerstone result in OT theory due to Brenier [Bre91] which gives sufficient conditions for the equivalence between (1.2.8) and its relaxation (1.2.7).

Theorem 1.2.1 (Informal). *Under mild regularity assumptions on the marginals,* (1.2.7) *and* (1.2.8) *are equivalent. Formulation* (1.2.8) *admits a unique solution, the Brenier map* $\boldsymbol{\mathcal{T}}$*. Moreover, the OT plan* $\boldsymbol{\pi}^0$ *is supported on the graph of the Brenier map* $\boldsymbol{\mathcal{T}}$*, i.e.,* $\boldsymbol{\pi}^0 = (\mathrm{Id}, \boldsymbol{\mathcal{T}})_{\#}\mu$*. Finally* $\boldsymbol{\mathcal{T}} = \mathrm{Id} - \nabla\varphi^0$ *where* φ^0 *is a Kantorovich potential, i.e., a optimiser in* (1.2.9).

As the SP plans $\boldsymbol{\pi}^T$ provide good proxies for the OT plan $\boldsymbol{\pi}^0$, we will introduce at (2.2.14) measurable maps (the Schrödinger potentials) associated to SP such that their gradients would eventually provide good proxies for the gradients of Kantorovich potentials and hence for the Brenier map $\boldsymbol{\mathcal{T}}$ (cf. Theorem 3.2.3). We will further elaborate on and prove these claims in Chapter 3.

1.3 The Schrödinger problem and stochastic control

The dynamical formulation of SP given at (1.1.4) can be further translated into a stochastic optimal control problem. More precisely, SP can be seen as the control problem which aims at modifying the law of a diffusion process so as to match specifications on marginal distributions at given times. This follows from Girsanov's Theorem [Léo12b, Theorems 2.1 and 2.3], which guarantees that any probability measure $\mathrm{P} \in \mathcal{P}(\Omega)$ with finite relative entropy w.r.t. R (*i.e.*, such that $\mathscr{H}(\mathrm{P}|\mathrm{R}) < +\infty$) is the law of a semimartingale. Particularly, if $B^{\mathrm{R}}_{\cdot} \sim \mathrm{R}$ is an R-Brownian motion and we consider its associated canonical filtration $(\mathcal{F}_t)_{t\in[0,T]}$, then there exists an $\mathbb{R}^d$-valued adapted process $u_{\cdot}$ such that P is a weak solution of the controlled SDE

$$\mathrm{d}X_t = u_t\,\mathrm{d}t + \mathrm{d}B^{\mathrm{P}}_t\ ,$$

where $B^{\mathrm{P}}_{\cdot}$ is a P-Brownian motion. Moreover, the Radon-Nikodym derivative equals

$$\frac{\mathrm{dP}}{\mathrm{dR}} = \mathbb{1}_{\{\frac{\mathrm{dP}_0}{\mathrm{dR}_0}>0\}}\frac{\mathrm{dP}_0}{\mathrm{dR}_0}(\mathrm{X}_0)\ \exp\left(\int_0^T u_t\,\mathrm{d}B^{\mathrm{R}}_t - \frac{1}{2}\int_0^T |u_t|^2\mathrm{d}t\right) \quad \text{R-a.e.}$$

and it holds

$$\mathscr{H}(\mathrm{P}|\mathrm{R}) = \mathscr{H}(\mathrm{P}_0|\mathrm{R}_0) + \frac{1}{2}\mathbb{E}_{\mathrm{P}}\left[\int_0^T |u_t|^2\mathrm{d}t\right].$$

As a direct consequence of the above stochastic representation, solving the dynamic SP (1.1.4) is then equivalent to minimising

$$\min_{u_\cdot \in \mathcal{A}} \mathbb{E}_{\mathrm{P}}\left[\int_0^T |u_t|^2 \mathrm{d}t\right] \tag{1.3.1}$$

where $\mathcal{A}$ denotes the set of admissible controls, *i.e.*, the set of $\mathbb{R}^d$-valued adapted control processes $u_\cdot$ such that if P^u denotes the law of the controlled process

$$\begin{cases} \mathrm{d}X_t^u = u_t\, \mathrm{d}t + \mathrm{d}B_t\,, \\ X_0^u \sim \mu\,, \end{cases} \tag{1.3.2}$$

then it holds $\mathrm{P}_T^u = \nu$. Equivalently, we can say that a control process $u_\cdot$ is admissible if it steers the controlled diffusion process (1.3.2) from $X_0^u \sim \mu$ into the target final measure $X_T^u \sim \nu$.

For exposition's sake here let us assume that (1.3.1) admits a unique solution $\boldsymbol{u}_\cdot^T$, then the Schrödinger bridge coincides with the diffusion process $X^{\boldsymbol{u}^T}$ controlled via $\boldsymbol{u}_\cdot^T$, that is $\mathbf{P}^T = \mathrm{P}^{\boldsymbol{u}^T}$. Moreover we have

$$\mathscr{H}(\mathbf{P}^T|\mathrm{R}) = \mathrm{Ent}(\mu) + \frac{1}{2}\mathbb{E}_{\mathbf{P}^T}\left[\int_0^T |\boldsymbol{u}_t^T|^2 \mathrm{d}t\right]. \tag{1.3.3}$$

In Chapter 2 we will further characterise the optimal control process $\boldsymbol{u}_\cdot^T$ as a feedback control and provide the reader with its explicit expression in (2.3.2) (see also Lemma 2.3.2).

Throughout this thesis, we will often employ this stochastic control interpretation either in our proof strategies or as a starting motivation for our discussion.

1.4 Overview of the main contributions

In this thesis we study a wide class of Schrödinger problems, where the ambient space considered is a (possibly unbounded) Riemannian manifold and the reference particles' dynamics R is a Langevin diffusion process. Under some regularity assumptions on the state space and on the reference measure, namely a *curvature-dimension condition* (cf. Section 2.1.1) SP still admits unique dynamic and static solutions $\mathbf{P}^T$ and $\boldsymbol{\pi}^T$ respectively and moreover there exists a couple of potentials known as Schrödinger potentials φ^T and ψ^T such that

$$\frac{\mathrm{d}\mathbf{P}^T}{\mathrm{d}\mathrm{R}}(X_\cdot) = \frac{\mathrm{d}\boldsymbol{\pi}^T}{\mathrm{d}\mathrm{R}_{0,T}}(X_0, X_T) = \exp\Big(-\varphi^T(X_0) - \psi^T(X_T)\Big)\,. \tag{1.4.1}$$

These two potentials play in SP a role analogous to the one of the Kantorovich potentials for the Kantorovich-OT, since also the former can be seen as optimiser of a dual formulation (cf. Proposition 2.2.2).

Convergence to the Brenier map. Our first contribution is showing in Chapter 3 that the gradients of the Schrödinger potentials can be used when approximating the Brenier optimal transport map $\boldsymbol{\mathcal{T}}$. More precisely, we show in Theorem 3.2.3 that in the small-time limit these gradients converge to the gradients of Kantorovich potentials:

$$-T\nabla\varphi^T \xrightarrow{\mathrm{L}^2(\mu)} \nabla\varphi^0 \quad \text{and} \quad -T\nabla\psi^T \xrightarrow{\mathrm{L}^2(\nu)} \nabla\psi^0\,. \tag{1.4.2}$$

Our proof strategy relies in proving gradient estimates for the Schrödinger potentials and their (backward-in-time) evolution along Hamilton-Jacobi-Bellman equations[2], by showing in Proposition 3.1.2 that

$$\|T\,\nabla u_0\|^2_{\mathrm{L}^2(\mu)} \le T\,\mathscr{H}(\mathbf{P}^T|\mathrm{R}) \quad \text{where} \quad \begin{cases}\partial_t u_t + \frac{1}{2}\Delta u_t + \frac{1}{2}|\nabla u_t|^2 = 0\\ u_T = \psi^T\,.\end{cases}$$

Such estimates are referred to as *corrector estimates* because they provide sharp bounds for the L^2-norm of the optimal control process $\boldsymbol{u}^T_\cdot$ in the stochastic optimal control formulation. These estimates were already known in literature in the Euclidean setting and, under stronger regularity assumptions, were used for studying the long-time behaviour of Schrödinger bridges [Con19]. Here we firstly generalise their validity to Riemannian manifolds and most importantly we show their sharpness in the small-time asymptotics as well. These qualitative convergence results are proven in [CCGT23]. In Chapter 3 we are further able to provide unpublished quantitative convergence rates for (1.4.2) in the Euclidean setting, based on convexity and functional inequalities ideas.

The gradient estimates employed in the convergence towards the Brenier map find a different application in Chapter 4. There we provide novel quantitative stability estimates, published in [CCGT23], which allow to measure how sensible Schrödinger bridges and plans are to perturbations of the two marginal constraints. The bounds we provide are expressed in terms of (symmetric) relative entropies and negative order Sobolev norms $\|\cdot\|_{\dot{\mathrm{H}}^{-1}}$. Particularly, if $\boldsymbol{\pi}^{\mu\to\nu,T}$ and $\boldsymbol{\pi}^{\mu\to\bar{\nu},T}$ denote the Schrödinger plans between the couple of marginals μ, ν and μ, $\bar{\nu}$ respectively, then we show in Theorem 4.2.2 that

$$\mathscr{H}^{\mathrm{sym}}(\boldsymbol{\pi}^{\mu\to\nu,T},\boldsymbol{\pi}^{\mu\to\bar{\nu},T}) \lesssim \mathscr{H}^{\mathrm{sym}}(\nu,\bar{\nu}) + T^{-1/2}\left(\|\nu-\bar{\nu}\|_{\dot{\mathrm{H}}^{-1}(\nu)} + \|\bar{\nu}-\nu\|_{\dot{\mathrm{H}}^{-1}(\bar{\nu})}\right)$$

where $\mathscr{H}^{\mathrm{sym}}$ is the symmetrised relative entropy (cf. (4.0.1)) and where $a \lesssim b$ whenever there exists a positive constant $C > 0$ such that $a \le C\,b$.

[2] Here we have considered the Hamilton-Jacobi-Bellman equation associated to the classical SP; see (5.1.2) for the Hamilton-Jacobi-Bellman equation for the general SP associated to a general Langevin dynamics reference.

Exponential convergence of Sinkhorn's algorithm. In our second main contribution we have analysed how Schrödinger bridges and potentials can be efficiently (and rapidly) computed. The starting point is noting that, since $\boldsymbol{\pi}^T \in \Pi(\mu,\nu)$, by integrating out either the first or the second variable in (1.4.1), we may deduce that the Schrödinger potentials satisfy

$$\begin{cases} \varphi^T = U_\mu + \log P_T \exp(-\psi^T) \\ \psi^T = U_\nu + \log P_T \exp(-\varphi^T)\,, \end{cases}$$

where U_μ, U_ν are the negative log-densities of the two marginals μ, ν, while P_T is the heat semigroup at time $t = T$. This suggests that the above system, also known as the *Schrödinger system*, can be solved via a fixed point iterative algorithm, namely by considering the iterates

$$\begin{cases} \varphi^{n+1} = U_\mu + \log P_T \exp(-\psi^n) \\ \psi^{n+1} = U_\nu + \log P_T \exp(-\varphi^{n+1})\,, \end{cases}$$

which are commonly referred to as *Sinkhorn's iterates*. The exponential convergence of Sinkhorn's algorithm has been widely studied in discrete settings [Sin64, SK67, PC19] and for compactly supported marginals [CGP16a, DdBD24, Ber20]. When considering unbounded settings, *i.e.*, non-compactly supported marginals with unbounded densities, much less is known. Particularly, prior to the results presented in this thesis, the most recent contribution [GN22] has solely shown convergence rates that are polynomial in the number of iterates.[3] In this thesis we provide the very first exponential convergence estimates in the Euclidean unbounded setting. Moreover, the approach we present here is quite novel since we mainly focus on the convergence of the gradients along Sinkhorn's algorithm and then deduce the convergence of the iterates as a corollary. The main reason behind adopting this strategy stems from the fact that in practice one would like to compute the gradients of the Schrödinger potentials, since we have shown that those provide good proxies for the optimal transport map, which is eventually the mathematical object one is interested in finding in optimal transport applications.

More precisely, we provide two different approaches.

The first is a *perturbative* approach and we employ it in Chapter 5 where we have adapted to the unbounded setting results coming from [GNCD23]. In Chapter 5, we consider as underlying particle system an ergodic Langevin dynamics

$$\begin{cases} \mathrm{d}X_t = -\nabla U(X_t)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}B_t \\ X_0 \sim \mathfrak{m}(\mathrm{d}x) \propto \exp(-U(x))\mathrm{d}x\,, \end{cases}$$

[3] We refer the reader to the Bibliographical Remarks section to Chapter 6 for a far more exhaustive literature review on the convergence of Sinkhorn's algorithm, where we also take into account the most recent developments.

with U strongly convex, and we ask the marginals' densities to be log-Lipschitz (*i.e.*, the marginals are log-Lipschitz perturbations of the underlying equilibrium measure $\mathfrak{m}$). Then, if $(P_t)_{t\in[0,T]}$ denotes the semigroup associated to the Langevin dynamics, by considering for any Lipschitz function h the function

$$\mathcal{U}_t^{T,h} := -\log P_{T-t}\exp(-h)\,,$$

we immediately notice that Sinkhorn's algorithm equivalently reads as

$$\begin{cases}\varphi^{n+1} = U_\mu - \mathcal{U}_0^{T,\psi^n}\\ \psi^{n+1} = U_\nu - \mathcal{U}_0^{T,\varphi^{n+1}}\,.\end{cases} \tag{1.4.3}$$

This gives a dynamic interpretation of Sinkhorn's algorithm since $\mathcal{U}_t^{T,h}$ solves the Hamilton-Jacobi-Bellman equation

$$\begin{cases}\partial_t u_t + \Delta u_t - \nabla U\cdot\nabla u_t - |\nabla u_t|^2 = 0\\ u_T = h\,,\end{cases} \tag{1.4.4}$$

and therefore one step of Sinkhorn's algorithm can be interpreted as considering the backward-in-time evolution of (1.4.4) combined with one of the two marginal constraints. Then, by exploiting the link between Hamilton-Jacobi-Bellman equations and value functions for stochastic optimal control problems (cf. (5.1.3)), we show in Lemma 5.1.1 that Lipschitzianity backward propagates along (1.4.4) and hence, in view of (1.4.3), that Lipschitzianity propagates along Sinkhorn's algorithm. By reasoning in a similar way, we are then able to show that

$$\mathrm{Lip}(\varphi^n - \varphi^T),\ \mathrm{Lip}(\psi^n - \psi^T) \lesssim \gamma^{2n}\,,$$

where $\gamma < 1$ is the convergence rate, explicitly given in Theorem 5.2.6. From the above Lip-convergence we are then able to deduce the exponential convergence of the algorithm in Theorems 5.2.6 and 5.2.7.

The second approach is *non-perturbative* and of a more geometric nature and it works for the classical SP setting (when considering Brownian motions in $\mathbb{R}^d$). We employ this approach in Chapter 6 where we present results coming from [CDG23]. There, we consider marginals with *asymptotically log-concave*[4] densities and show how convexity and concavity propagate from the marginals towards the limit potentials, along Sinkhorn's algorithm. In view of the link between Sinkhorn's algorithm and Hamilton-Jacobi-Bellman equations portrayed in (1.4.3), this is accomplished by finding a good set of (almost) convex functions which is invariant under the backward-in-time evolution along Hamilton-Jacobi-Bellman equations. This result is proven in Theorem 6.1.4, which extends the existing [Con24, Theorem 2.1].

[4]See the definition at (6.0.11), it includes densities with log-densities that are strongly concave outside a compact set or Lipschitz perturbations of strongly concave functions.

Then, by relying on the link between the gradient of Sinkhorn's iterates and conditional probability measures (as portrayed in (6.0.3)), we notice that

$$\int |\nabla\varphi^{n+1} - \nabla\varphi^{\star}|(x)\,\mu(\mathrm{d}x) \leq T^{-1} \int \mathbf{W}_1(\pi_T^{x,\psi^n}, \pi_T^{x,\psi^\star})\,\mu(\mathrm{d}x)\,,$$

where $\pi_T^{x,\psi^\star}$, π_T^{x,ψ^n} are respectively the invariant law of the SDEs

$$\begin{cases} \mathrm{d}X_t = -\Big(\frac{X_t - x}{2T} + \frac{1}{2}\nabla\psi^\star(X_t)\Big)\mathrm{d}t + \mathrm{d}B_t \\ \mathrm{d}Y_t = -\Big(\frac{Y_t - x}{2T} + \frac{1}{2}\nabla\psi^n(Y_t)\Big)\mathrm{d}t + \mathrm{d}B_t\,. \end{cases} \tag{1.4.5}$$

We then employ synchronous and reflection coupling techniques for the above diffusion processes in order to bound the Wasserstein distance $\mathbf{W}_1(\pi_T^{x,\psi^n}, \pi_T^{x,\psi^\star})$ with the integrated difference between the drift appearing in (1.4.5) and deduce contraction estimates such as

$$\begin{aligned} \int |\nabla\varphi^{n+1} - \nabla\varphi^T|(x)\,\mu(\mathrm{d}x) &\lesssim T^{-1}\,\gamma^\nu \int |\nabla\psi^n - \nabla\psi^T|(y)\nu(\mathrm{d}y)\,, \\ \int |\nabla\psi^n - \nabla\psi^T|(y)\,\nu(\mathrm{d}y) &\lesssim T^{-1}\,\gamma^\mu \int |\nabla\varphi^n - \nabla\varphi^T|(x)\mu(\mathrm{d}x)\,. \end{aligned} \tag{1.4.6}$$

Our coupling technique differs from its common use since we are not interested here in showing ergodicity of (1.4.5), but rather interested in obtaining in a finite-time window sharp bounds for $\mathbf{W}_1(\pi_T^{x,\psi^n}, \pi_T^{x,\psi^\star})$ in terms of the difference between the two drifts. Since this novel approach is of independent interest, we show its validity in wide generality in Section 6.2.

Finally, from (1.4.6) we are able to bootstrap contraction estimates along Sinkhorn's algorithm and deduce in Theorem 6.3.2 the exponential integrated convergence of the gradient of Sinkhorn's iterates as well as the exponential convergence in Wasserstein ($\mathbf{W}_1$) distance for the corresponding couplings (defined at (2.2.19)). By slightly modifying our proof strategy we deduce also pointwise exponential convergence for iterates and gradients (respectively in Theorems 6.3.5 and 6.4.1), exponential convergence in symmetric relative entropy (in Theorem 6.4.4) and convergence of the Hessian of Sinkhorn's iterates (in Theorem 6.5.2).

The kinetic Schrödinger problem. Our last contribution focuses on a different type of the Schrödinger problem and is based on [CCGR22]. Namely, in Chapter 7 we introduce the *kinetic Schrödinger problem* as an instance of SP where the underlying process is replaced with an *underdamped Langevin dynamics*, that is

$$\begin{cases} \mathrm{d}X_t = V_t\mathrm{d}t, \\ \mathrm{d}V_t = -\nabla U(X_t)\mathrm{d}t - \gamma V_t\mathrm{d}t + \sqrt{2\gamma}\,\mathrm{d}B_t\,. \end{cases}$$

Here the particles are described both by their position and their momenta, which satisfy the kinetic Fokker-Planck equation, and we are provided just snapshots of the positions at initial and final time (therefore the marginals' information is only partial). We fully characterise this problem and, inspired by its stochastic control formulation, we study its long-time behaviour. To do so, our main tools are corrector estimates similar to the ones we have obtained in SP classical setting. Even though the problem is just a different instance of SP, the hypoelliptic nature of the kinetic Fokker-Planck equation requires a more careful analysis when studying the corrector estimates. Particularly, these estimates are strictly related to the long-time behaviour of the kinetic Fokker-Planck equation, which is strongly affected by its hypocoercivity [Vil09].

We then employ these estimates in establishing the validity of the *turnpike property*, which can be loosely described as saying that in the long-time regime Schrödinger bridges spend most of the time exponentially close to the equilibrium. More precisely, in the long-time regime, Schrödinger bridges converge exponentially fast towards the equilibrium, spend most of the time exponentially close to the equilibrium and then, just at the very end, leave the equilibrium and reach the final prescribed target distribution. Namely we show in Theorem 7.5.1 that for any $t \in (0, T)$ it holds

$$\mathscr{H}(\mathbf{P}_t^T|\mathfrak{m}) \lesssim \exp\Big(-2\kappa\Big[t \wedge (T-t)\Big]\Big)\Big[\mathscr{H}(\mu|\mathfrak{m}_X) + \mathscr{H}(\nu|\mathfrak{m}_X)\Big],$$

with $\mathfrak{m}_X \propto \exp(-U(x))$, while $\mathfrak{m} \propto \exp(-U(x) - |v|^2/2)$ being the equilibrium measure associated to the (underdamped) Langevin dynamics and with $\kappa > 0$ being its ergodicity rate.

The turnpike property is a general principle, widely investigated in deterministic control problems [TZ15, TZZ18, Zas05, Zas19]. In the field of stochastic control the understanding of this phenomenon is much more limited. Moreover, as we have already mentioned the hypocoercive nature of the kinetic Fokker-Planck equation makes the analysis even more delicate.

Organisation of the thesis Finally, we now discuss the organisation of the thesis. The thesis is divided into seven chapters, each one followed by an appendix in which we have proved some technical and ancillary results useful to the rest of the exposition. At the end of Chapters 3 to 7 we further provide a *Bibliographical Remarks* section where we point out the main references those chapters are based on and where we provide a literature review on the subject analysed in each chapter.

We conclude the thesis with the chapter *Looking forward*, devoted to future research directions that could possibly follow from the results presented here.

This page was intentionally left blank.

Appendix 1

1.A On the notion of relative entropy

In this appendix we collect some useful properties of the relative entropy functional $\mathscr{H}$ that will be employed later in the thesis and we extend the notion of relative entropy to the case when the reference measure is not a probability measure.

Let $\mathcal{X}$ be a Polish measurable space and consider $\mathrm{P}, \mathrm{Q} \in \mathcal{P}(\mathcal{X})$ be two probability measures. Then the relative entropy of P with respect to Q is defined as

$$\mathscr{H}(\mathrm{P}|\mathrm{Q}) := \begin{cases} \int \log \frac{\mathrm{dP}}{\mathrm{dQ}} \, \mathrm{dP} & \text{if } \mathrm{P} \ll \mathrm{Q}, \\ +\infty & \text{otherwise.} \end{cases} \tag{1.A.1}$$

Equivalently it can be defined for any $\mathrm{P}, \mathrm{Q} \in \mathcal{P}(\mathcal{X})$ as

$$\mathscr{H}(\mathrm{P}|\mathrm{Q}) = \int h\left(\frac{\mathrm{dP}}{\mathrm{dQ}}\right) \mathrm{dQ} \in [0, +\infty],$$

where $h(a) := a \log a - a + 1 \geq 0$ for all $a \geq 0$ (with $h(0) = 0$). Since h is non-negative and $\mathrm{Q} \in \mathcal{P}(\mathcal{X})$, it is immediate to notice that as soon as $\mathscr{H}(\mathrm{P}|\mathrm{Q}) < +\infty$ it holds $\log \mathrm{dP}/\mathrm{dQ} \in \mathrm{L}^1(\mathrm{P})$.

Lemma 1.A.1 (Lemma 1.4.3 in [DE97]). *Assume* $\mathrm{Q} \in \mathcal{P}(\mathcal{X})$. *Then the relative entropy functional* $\mathscr{H}(\cdot|\mathrm{Q})$

- *is convex and lower-semicontinuous in* $\mathcal{P}(\mathcal{X})$ *in the weak topology, and strictly convex on the subset* $\{\mathrm{P}\colon \mathscr{H}(\mathrm{P}|\mathrm{Q}) < +\infty\}$*;*
- *has weakly-compact level-sets, i.e., for any finite* L *the subset* $\{\mathrm{P}\colon \mathscr{H}(\mathrm{P}|\mathrm{Q}) \leq L\}$ *is compact in the weak topology of* $\mathcal{P}(\mathcal{X})$*;*
- *for any* $\mathrm{P} \in \mathcal{P}(\mathcal{X})$ *we have* $\mathscr{H}(\mathrm{P}|\mathrm{Q}) = 0$ *if and only if* $\mathrm{P} \equiv \mathrm{Q}$*;*
- *for any* $\mathrm{P} \in \mathcal{P}(\mathcal{X})$ *the* Donsker-Varadhan variational formula *holds, that is*

$$\begin{aligned} \mathscr{H}(\mathrm{P}|\mathrm{Q}) &= \sup_{h \in \mathcal{C}_b(\mathcal{X})} \left\{ \int h \, \mathrm{dP} - \log \int e^h \mathrm{dQ} \right\} \\ &= \sup_{h \in \mathcal{M}_b(\mathcal{X})} \left\{ \int h \, \mathrm{dP} - \log \int e^h \mathrm{dQ} \right\}. \end{aligned} \tag{1.A.2}$$

In order to define the relative entropy when the reference measure Q is not a probability measure but it is just σ-finite (as it is when R is the Wiener measure associated to a Brownian motion), we follow [CT21, Section 2] and [Léo14, Appendix A]. As Q is σ-finite, there exists a measurable function $W\colon \mathcal{X} \to [0,+\infty)$ such that

$$z_W := \int e^{-W}\,\mathrm{dQ} < +\infty\,.$$

Then, for any $\mathrm{P} \in \mathcal{P}(\mathcal{X})$ such that $W \in \mathrm{L}^1(\mathrm{P})$ we can define its relative entropy with respect to Q as

$$\mathscr{H}(\mathrm{P}|\mathrm{Q}) := \mathscr{H}(\mathrm{P}|\mathrm{Q}_W) - \int W\,\mathrm{dP} - \log z_W \in (-\infty,+\infty]\,, \tag{1.A.3}$$

where $\mathrm{Q}_W := z_W^{-1}\,e^{-W}\,\mathrm{Q} \in \mathcal{P}(\mathcal{X})$ and $\mathscr{H}(\mathrm{P}|\mathrm{Q}_W)$ is defined via the standard definition (1.A.1) of relative entropy between probabilities. Moreover, it is immediate to see that the above definition is independent of the choice of the measurable function $W\colon M \to [0,+\infty)$ satisfying $z_W < +\infty$.

In the next result we collect some nice integrability properties linked with the above definition of relative entropy with respect to σ-finite measures.

Lemma 1.A.2 (Lemma 4.1.1 in [Tam17])**.** *Assume that* Q *is a σ-finite measure on our Polish space $\mathcal{X}$ and fix a probability measure* $\mathrm{P} \in \mathcal{P}(\mathcal{X})$.

1. *If there exists a non-negative measurable function* $W \in \mathrm{L}^1(\mathrm{P})$ *such that* $z_W < +\infty$, *then either* $\mathrm{P} \not\ll \mathrm{Q}$ *or* $(\log \mathrm{dP}/\mathrm{Q})^- \in \mathrm{L}^1(\mathrm{P})$.

2. *In addition to that, if* $\mathscr{H}(\mathrm{P}|\mathrm{Q})$ *(defined via* (1.A.3)*) is finite, then* $\mathrm{P} \ll \mathrm{Q}$ *and* $\log \mathrm{dP}/\mathrm{dQ} \in \mathrm{L}^1(\mathrm{P})$.

Proof. Firstly, assume that $\mathrm{P} \ll \mathrm{Q}$. Then we can bound

$$\begin{aligned}\|(\log \mathrm{dP}/\mathrm{dQ})^-\|_{\mathrm{L}^1(\mathrm{P})} &= \|(\log \mathrm{dP}/\mathrm{dQ}_W - W - \log z_W)^-\|_{\mathrm{L}^1(\mathrm{P})}\\ &\leq \int \frac{\mathrm{dP}}{\mathrm{dQ}_W}\left(\log\frac{\mathrm{dP}}{\mathrm{dQ}_W}\right)^- \mathrm{dQ}_W + \|W\|_{\mathrm{L}^1(\mathrm{P})} + |\log z_W|\end{aligned}$$

which is finite since for any $a > 0$ it holds $a(\log a)^- \leq e^{-1}$ and since Q_W is a probability measure. This proves the first item of the lemma.

Next, notice that the same argument gives also that

$$\|(\log \mathrm{dP}/\mathrm{dQ})^+\|_{\mathrm{L}^1(\mathrm{P})} \leq \|(\log \mathrm{dP}/\mathrm{dQ}_W)^+\|_{\mathrm{L}^1(\mathrm{P})} + \|W\|_{\mathrm{L}^1(\mathrm{P})} + |\log z_W|\,.$$

and the above right hand side is finite since $\mathscr{H}(\mathrm{P}|\mathrm{Q})$ is finite and from the definition (1.A.3) we may deduce that also $\mathscr{H}(\mathrm{P}|\mathrm{Q}_W)$ is finite as well or equivalently that $\log \mathrm{dP}/\mathrm{dQ}_W \in \mathrm{L}^1(\mathrm{P})$ (since $\mathrm{Q}_W \in \mathcal{P}(\mathcal{X})$). By combining the above with the first item of our lemma, we conclude our proof. □

Additive property.

Finally, let us conclude by stating the additive property of relative entropy which guarantees the connection between the (dynamical) Schrödinger problem, its static formulation and therefore (entropic) optimal transport.

Let $\mathcal{Y}$ be a Polish space and let $\phi\colon \mathcal{X} \to \mathcal{Y}$ be a measurable map. Then for any non-negative measure Q on $\mathcal{X}$ we have the disintegration formula (cf. [DM78, Chapter III - paragraph 70])

$$\mathsf{Q}(\mathsf{A}) = \int_{\mathcal{Y}} \mathsf{Q}(\mathsf{A}|\phi = y)\, \phi_{\#}\mathsf{Q}(\mathrm{d}y)$$

which holds for any Borel set $\mathsf{A} \subseteq \mathcal{X}$ and where the map $y \mapsto \mathrm{Q}(\cdot|\phi = y) \in \mathcal{P}(\mathcal{X})$ is measurable. Then, for any $\mathrm{P} \in \mathcal{P}(\mathcal{X})$ and any non-negative σ-finite measure Q we have

$$\mathscr{H}(\mathrm{P}|\mathrm{Q}) = \mathscr{H}(\phi_{\#}\mathrm{P}|\phi_{\#}\mathrm{Q}) + \int \mathscr{H}(\mathrm{P}(\cdot|\phi = y)|\mathrm{Q}(\cdot|\phi = y))\, \phi_{\#}\mathrm{P}(\mathrm{d}y)\,. \tag{1.A.4}$$

Moreover, since $\mathrm{Q}(\cdot|\phi = y)$ is a probability measure for any given $y \in \mathcal{Y}$, the relative entropy w.r.t. such probability is always non negative and hence we deduce the *data-processing inequality*

$$\mathscr{H}(\mathrm{P}|\mathrm{Q}) \geq \mathscr{H}(\phi_{\#}\mathrm{P}|\phi_{\#}\mathrm{Q})\,. \tag{1.A.5}$$

We refer the reader to [Léo14, Appendix A] and [DE97, Lemma 1.4.3-(f)] for further discussion and for a proof of the above statements.

This page was intentionally left blank.

Chapter 2

Preliminaries on the Schrödinger problem

This chapter serves as a more in depth introduction to the Schrödinger problem (SP) in a more abstract setting, namely when considering as underlying reference process a Langevin diffusion on a (possibly unbounded) Riemannian manifold. This is accomplished in Section 2.1. In order to generalise SP to this setting we will briefly introduce in Section 2.1.1 the *curvature-dimension condition*, a crucial tool that guarantees existence and uniqueness of solutions to SP and at the same time guarantees gradient estimates on top of which we will deduce the corrector estimates in Chapter 3. In Section 2.2 we introduce the assumptions we impose on the marginals and completely characterise the solution of SP in Theorem 2.2.1. In Section 2.2.1 we then introduce Sinkhorn's algorithm, which provides an iterative method that computes Schrödinger bridges and whose convergence will be studied in depth in Chapters 5 and 6. Lastly, in Section 2.3 we link the Schrödinger potentials (as introduced in Theorem 2.2.1) with the optimal control of the stochastic optimal control formulation.

2.1 Underlying setting

Unless otherwise stated, in this thesis we consider as ambient space a weighted Riemannian manifold, *i.e.*, a triplet $(M, \mathsf{d}, \mathfrak{m})$ where M is a smooth, connected, complete (possibly non-compact) Riemannian manifold without boundary and with metric tensor g and corresponding geodesic distance d, whereas $\mathfrak{m}$ denotes the σ-finite measure $\mathrm{d}\mathfrak{m}(x) = e^{-U(x)}\mathrm{vol}(\mathrm{d}x)$, where vol is the volume measure of M and U is a $\mathcal{C}^2$ potential.

We denote by $\mathcal{P}_p(M)$, $p \geq 1$, the set of probability measures on M with finite p-moment, that is such that for a fixed $z_0 \in M$ it holds

$$M_p(\mathfrak{p}) := \int \mathsf{d}^p(y, z_0)\mathrm{d}\mathfrak{p}(y) < +\infty$$

Let us also notice that $\mathcal{P}_2(M) \subseteq \mathcal{P}_1(M)$ and in particular $M_1^2(\mathfrak{p}) \leq M_2(\mathfrak{p})$ for any $\mathfrak{p} \in \mathcal{P}_2(M)$. When M is a vector space (e.g., $M = \mathbb{R}^d$) in the above definition we pick the origin $z_0 = 0 \in M$.

Moreover, instead of considering just Brownian motions in Schrödinger's thought experiment, we are going to assume that the particles of the cloud follow the Langevin SDE

$$\begin{cases} \mathrm{d}X_t = -\nabla U(X_t)\,\mathrm{d}t + \sqrt{2}\mathrm{d}B_t \\ X_0 \sim \mathfrak{m}\,, \end{cases} \tag{2.1.1}$$

i.e., the diffusion process associated to the generator $L = \Delta_g - \nabla U \cdot \nabla$[1]. Notice that $\mathfrak{m}$ is the invariant measure of the above dynamics. Then, given any two probability measures $\mu, \nu \in \mathcal{P}(M)$, we may again deduce from Sanov's Theorem (as we already did in Section 1.1) that the most likely evolution of our system can be computed by solving the dynamical minimisation problem

$$\inf_{\{\mathrm{P}\in\mathcal{P}(\Omega):\ \mathrm{P}_0=\mu,\, \mathrm{P}_T=\nu\}} \mathscr{H}(\mathrm{P}|\mathrm{R}) \tag{2.1.2}$$

this time with R being the law of the Langevin dynamics (2.1.1). As we already did in the Introduction, we denote with $\mathbf{P}^T$ the optimiser of the above problem, to which we refer again as to the *Schrödinger bridge*. By arguing as in Section 1.2 we may again deduce a static formulation for SP (as in (1.2.2)) which reads again as

$$\inf_{\pi\in\Pi(\mu,\nu)} \mathscr{H}(\pi|\mathrm{R}_{0,T})\,, \quad \text{with } \mathrm{R}_{0,T} = \mathcal{L}(X_0, X_T)\,. \tag{2.1.3}$$

We denote the optimiser of the above problem with $\boldsymbol{\pi}^T$ and refer to it as to the *Schrödinger (or entropic) plan*. Clearly, as we have already explained in Section 1.2, it is possible to recover $\mathbf{P}^T$ from $\boldsymbol{\pi}^T$ by connecting initial and final position via Langevin-bridges, *i.e.*, it holds

$$\mathbf{P}^T(\cdot) = \int_{M\times M} \mathrm{R}^{xy}(\cdot)\,\mathrm{d}\boldsymbol{\pi}^T(x,y)\,.$$

For later reference, let us denote the density of $\mathrm{R}_{0,T}$ as $\mathsf{p}_T = \frac{\mathrm{dR}_{0,T}}{\mathrm{d}(\mathfrak{m}\otimes\mathfrak{m})}$. We will refer to the optimal value in (2.1.3) as to the *Schrödinger cost* and we will denote it by $\mathcal{C}_T(\mu,\nu)$.

In the flat Euclidean setting when the Riemannian manifold $(M, \mathsf{d}) = \mathbb{R}^d$, the volume measure vol corresponds to the the Lebesgue measure Leb and we recover the standard Brownian SP considered in the Introduction by choosing $U \equiv 0$ and hence $\mathfrak{m} = \mathrm{vol} = \mathrm{Leb}$ whereas p_T corresponds to the heat kernel (up to the rescaling factor 2 in time, due to the presence of $\sqrt{2}$ in the SDE).

[1]We refer the reader to [BGL13, Appendix B] for the connection between the generator L and diffusion processes on Riemannian manifolds. In this thesis we will rely on the SDE (2.1.1) representation solely in the Euclidean standard setting.

2.1.1 Curvature-dimension condition

We are going to assume $(M, \mathsf{d}, \mathfrak{m})$ to satisfy a *curvature-dimension* condition, which will bring enough structure and regularity to the ambient space such that existence and uniqueness of solutions to SP hold. From a stochastic point of view one could think of the curvature-dimension condition as a more abstract generalisation of the Bakry-Émery criterion for the ergodicity of the Langevin SDE (2.1.1).

Bakry-Émery and Γ-calculus

Bakry-Émery approach to hypercontractivity relies on establishing whether there exists a positive $\kappa \in \mathbb{R}$ such that it holds

$$\Gamma_2(f,f) \geq \kappa\,\Gamma(f,f) \tag{2.1.4}$$

for any (suitably regular) function $f \in \mathrm{L}^2(\mathfrak{m})$, where Γ and Γ_2 are respectively the *carré du champ* and the *iterated carré du champ* operators associated to the generator $L = \Delta_g - \nabla U \cdot \nabla$, *i.e.*, they correspond to the operators defined for any f, g (in a suitable subalgebra of $\mathrm{L}^2(\mathfrak{m})$) as

$$\Gamma(f,g) := \frac{1}{2}\Big[L(fg) - fLg - gLf\Big]\,,$$
$$\Gamma_2(f,g) := \frac{1}{2}\Big[L\Gamma(f,g) - \Gamma(f,Lg) - \Gamma(Lf,g)\Big]\,.$$

Inequality (2.1.4) is referred to as the Bakry-Émery criterion after the seminal work [BÉ85] and it is well known that this local differential condition guarantees exponential ergodicity in $\mathbf{W}_2$-distance for the diffusion process (2.1.1), with $\kappa > 0$ as exponential convergence rate [BGL13, Theorem 9.7.2].

A first example one could think of is the flat Euclidean space with the measure $\mathrm{d}\mathfrak{m} = e^{-U}\mathrm{dLeb}$, where the two operators equal

$$\Gamma(f,f) = |\nabla f|^2 \quad \text{and} \quad \Gamma_2(f,f) = |\nabla^2 f|^2 + \nabla^2 U(\nabla f, \nabla f)$$

and therefore a sufficient condition for (2.1.4) to hold is the κ-convexity of the potential U, *i.e.*, that uniformly it holds

$$\nabla^2 U \geq \kappa\,.$$

Lastly, let us further mention that (2.1.4) implies useful gradient estimates and *local log-Sobolev inequalities* [BGL13, Theorem 5.5.2], as we will detail later.

Nevertheless, the class of log-concave measure is not rich enough for the purposes of this thesis. For instance the very first example considered in the Introduction (*i.e.*, rescaled Brownian motion on $\mathbb{R}^d$, where $L = \Delta$) fails to satisfy the Bakry-Émery condition (2.1.4) though it is immediate to notice that Cauchy-Schwartz inequality gives

$$\Gamma_2(f,f) = |\nabla^2 f|^2 \geq \frac{1}{d}(\Delta f)^2 = \frac{1}{d}(Lf)^2\,, \tag{2.1.5}$$

which lower bounds the Γ_2 operator with the generator L itself. Despite being different from (2.1.4), the above inequality carries other geometric and analytical properties such as the validity of *local dimensional log-Sobolev inequalities* [BGL13, Theorem 6.7.3] and the Li-Yau and Harnack inequalities [BGL13, Corollaries 6.7.5 and 6.7.6]. Lastly, it provides a bound with explicit dependence from the dimension of the ambient space (which does not appear on the contrary in (2.1.4)).

Therefore, the above discussion suggests to consider a condition on the operator Γ_2 which combines both a lower bound on the curvature as in (2.1.4) and an upper bound on the dimension as in (2.1.5), namely

$$\Gamma_2(f,f) \geq \kappa\,\Gamma(f,f) + \frac{1}{N}(Lf)^2\,. \tag{2.1.6}$$

If the above inequality holds for $\kappa \in \mathbb{R}$ and $N \in \mathbb{N}$ (or possibly $N = +\infty$), we will say that the curvature-dimension condition $\mathsf{CD}(\kappa, N)$ is satisfied.

CD in Riemannian formalism

Condition (2.1.6) can be immediately generalised to weighted Riemannian manifolds. Indeed if $L = \Delta_g - \nabla U \cdot \nabla$, then for any (regular enough) $f \in \mathrm{L}^2(\mathfrak{m})$ it holds

$$\Gamma(f,f) = |\nabla f|^2 \quad \text{and} \quad \Gamma_2(f,f) = |\nabla^2 f|^2 + \mathrm{Ric}_U(\nabla f, \nabla f),$$

where the *Bakry-Émery Ricci tensor* is defined as

$$\mathrm{Ric}_U := \mathrm{Ric}_g + \mathrm{Hess}(U)\,.$$

We refer the reader to [BGL13, Section C.5] where the connection between the Γ_2 operator and the tensor Ric_U is established as a consequence of the Bochner-Lichnerowicz formula [BGL13, Theorem C.3.3]. Particularly, the splitting decomposition of Γ_2 allows to immediately deduce that inequality (2.1.6) holds true for $N \geq \dim(M)$ if and only if it holds

$$\mathrm{Ric}_U(\nabla f, \nabla f) \geq \kappa |\nabla f|^2 + \frac{1}{N}(Lf)^2\,. \tag{$\mathsf{CD}(\kappa,N)$}$$

For the readers more interested in the geometric aspects of $\mathsf{CD}(\kappa, N)$, let us mention that the above condition is also equivalent [Vil08, Theorem 14.8] to the more geometric condition

$$\mathrm{Ric}_U - \frac{\nabla U \otimes \nabla U}{N-d} \geq \kappa g, \qquad d := \dim(M).$$

Therefore one could consider κ as a lower bound on the Bakry-Émery Ricci curvature tensor $\mathrm{Ric}_U := \mathrm{Ric}_g + \mathrm{Hess}(U)$, whereas N has to be seen as an upper bound on the *effective* dimension of the generator induced by (2.1.1). The term *effective* employed here is to remark that this dimension N does not necessarily coincide with the topological dimension of the underlying manifold. For instance, $\mathbb{R}^d$ weighted with the standard Gaussian measure $\mathfrak{m} \propto \exp(-|x|^2/2)$ satisfies $\mathsf{CD}(1,\infty)$ but no $\mathsf{CD}(\kappa, N)$ for $\kappa \in \mathbb{R}$ and finite N can be satisfied (because $\mathsf{CD}(\kappa, N)$ implies that $\mathfrak{m}$ is locally doubling, which is not the case for the standard Gaussian).

Consequences of CD

Thorough this thesis we are going to assume that our ambient space $(M, \mathsf{d}, \mathfrak{m})$ satisfies a curvature-dimension condition, *i.e.*, that either one of the following holds

- $(M, \mathsf{d}, \mathfrak{m})$ satisfies $\mathsf{CD}(\kappa, N)$ for some $\kappa \in \mathbb{R}$ and $N < +\infty$, or
- $(M, \mathsf{d}, \mathfrak{m})$ satisfies $\mathsf{CD}(\kappa, \infty)$ for some $\kappa \in \mathbb{R}$ and $\mathfrak{m}(M) = 1$.

(CD)

Clearly, as we have already pointed out at (2.1.5), this condition (namely, the condition $\mathsf{CD}(0, d)$) is met in our landmark classical SP example where we considered Brownian particles in the Euclidean space $\mathbb{R}^d$.

The above curvature assumption yields many important consequences.

Firstly, it implies Gaussian lower bounds for $\mathsf{p}_t = \frac{\mathrm{d}\mathrm{R}_{0,t}}{\mathrm{d}(\mathfrak{m}\otimes\mathfrak{m})}$. More precisely in [JLZ16, Theorem 1.1 and Theorem 1.2] it is proven that $\mathsf{CD}(\kappa, N)$ with $N < \infty$, implies that for every $\delta > 0$ there exist constants C_1, C_2 depending only on κ, N, δ such that it holds

$$\mathsf{p}_t(x,y) \geq \frac{1}{C_1 \mathfrak{m}(B_{\sqrt{t}}(x))} \exp\Big(- \frac{\mathsf{d}^2(x,y)}{(4-\delta)t} - C_2 t \Big) \tag{2.1.7}$$

for all $x, y \in M$ and all $t > 0$. On the other hand, if $(M, \mathsf{d}, \mathfrak{m})$ satisfies $\mathsf{CD}(\kappa, \infty)$ with $\mathfrak{m}(M) = 1$, then from [Wan11, Corollary 1.3] one can deduce for all positive $t > 0$ the Gaussian lower bound

$$\mathsf{p}_t(x,y) \geq \exp\Big(- \frac{\kappa \mathsf{d}^2(x,y)}{2(1-e^{-\kappa t})} \Big) \tag{2.1.8}$$

for all $x, y \in M$.

Moreover, the curvature-dimension condition guarantees Lipschitz and gradient estimates along the time evolution of the semigroup $(P_t)_{t\geq 0}$ associated to the generator L. Namely, $\mathsf{CD}(\kappa, \infty)$ implies the L^∞-Lipschitz regularisation [AGS14b, Theorem 6.8]

$$\sqrt{2E_{2\kappa}(t)}\, \mathrm{Lip}(P_t u) \leq \|u\|_{L^\infty(\mathfrak{m})}, \qquad \forall t > 0,\ u \in \mathrm{L}^\infty(\mathfrak{m}), \tag{2.1.9}$$

where the Lip-norm is defined for any continuous function $h \in \mathcal{C}(M)$ as

$$\mathrm{Lip}(h) := \sup_{x \neq y} \frac{|h(x) - h(y)|}{\mathsf{d}(x, y)}$$

whereas the factor $E_{2\kappa}(t)$ is defined for any $t \geq 0$ as

$$E_{2\kappa}(t) := \int_0^t e^{2\kappa s} \mathrm{d}s = \begin{cases} \frac{e^{2\kappa t} - 1}{2\kappa} & \text{for } \kappa \neq 0\,, \\ t & \text{for } \kappa = 0\,. \end{cases} \tag{2.1.10}$$

Under $\mathsf{CD}(\kappa, \infty)$, Hamilton's gradient estimate [Ham93, Kot07] (see also [JZ16]) holds true as well, *i.e.*, for any positive $u \in \mathrm{L}^p \cap \mathrm{L}^\infty(\mathfrak{m})$ for some $p \in [1, \infty)$ it holds

$$t|\nabla \log P_t u|^2 \leq (1 + 2\kappa^- t) \log\Big(\frac{\|u\|_{L^\infty(\mathfrak{m})}}{P_t u}\Big). \tag{2.1.11}$$

On the other hand, when the effective dimension N is finite, the condition $\mathsf{CD}(\kappa, N)$ is equivalent to the validity of the following gradient commutation estimate

$$\begin{aligned} |\nabla P_t u|^2 &\leq e^{-2\kappa t} P_t |\nabla u|^2 - \frac{2}{N} \int_0^t e^{-2\kappa s} P_s (P_{t-s} L u)^2 \mathrm{d}s \\ |\nabla P_t u|^2 &\leq e^{-2\kappa t} P_t |\nabla u|^2 - \frac{1 - e^{-2\kappa t}}{\kappa N} (P_t L u)^2 \end{aligned} \tag{2.1.12}$$

for any compactly supported smooth function u and for any $t \geq 0$, see [Wan11, Theorem 1.1].

Finally, let us mention a more geometric consequence: $\mathsf{CD}(\kappa, N)$ condition with $N < \infty$ entails the Bishop-Gromov inequality (see for instance [Stu06b, Theorem 2.3]): for any $x \in \mathrm{supp}(\mathfrak{m}) = M$ and $0 < r \leq R \leq \pi\sqrt{(N-1)/\kappa^+}$, where $(N-1)/\kappa^+ = +\infty$ if $\kappa \leq 0$, it holds

$$\frac{\mathfrak{m}(B_r(x))}{\mathfrak{m}(B_R(x))} \geq \begin{cases} \dfrac{\int_0^r \sin(t\sqrt{\kappa/(N-1)})^{N-1} \mathrm{d}t}{\int_0^R \sin(t\sqrt{\kappa/(N-1)})^{N-1} \mathrm{d}t}\,, & \text{if } \kappa > 0\,, \\ \Big(\dfrac{r}{R}\Big)^N\,, & \text{if } \kappa = 0\,, \\ \dfrac{\int_0^r \sinh(t\sqrt{-\kappa/(N-1)})^{N-1} \mathrm{d}t}{\int_0^R \sinh(t\sqrt{-\kappa/(N-1)})^{N-1} \mathrm{d}t}\,, & \text{if } \kappa < 0\,. \end{cases} \tag{2.1.13}$$

The Bishop-Gromov inequality implies a first log-integrability result, which will later be useful for the existence and uniqueness of solutions to SP.

Lemma 2.1.1. *Let* $(M, \mathsf{d}, \mathfrak{m})$ *satisfies* $\mathsf{CD}(\kappa, N)$ *with* $N < +\infty$. *Then for any* $r > 0$ *and for any measure* $\mathfrak{p} \in \mathcal{P}_1(M)$ *with finite first moment, it holds*

$$(\log \mathfrak{m}\,(B_r(\cdot)))^+ \in \mathrm{L}^1(\mathfrak{p})\,.$$

Proof. Assume $\kappa < 0$. If we fix $\bar{x} \in M$ and observe that for any $x \in M$ we have $B_r(x) \subset B_{\mathsf{d}(x,\bar{x})+r}(\bar{x})$. Then, (2.1.13) implies

$$\mathfrak{m}(B_r(x)) \leq C_1 \mathfrak{m}(B_r(\bar{x})) \int_0^{\mathsf{d}(x,\bar{x})+r} \sinh(t\sqrt{-\kappa/(N-1)})^{N-1} \mathrm{d}t \leq C_2\, e^{C_3 \mathsf{d}(x,\bar{x})},$$

where C_1, C_3 only depend on κ, N, r and C_2 also depends on $\bar{x}$. As a consequence,

$$\log \mathfrak{m}(B_r(x)) \leq \log C_2 + C_3 \mathsf{d}(x, \bar{x}) \tag{2.1.14}$$

and therefore $(\log \mathfrak{m}\,(B_r(\cdot)))^+ \in \mathrm{L}^1(\mathfrak{p})$ for any $\mathfrak{p} \in \mathcal{P}_1(M)$. In the case $\kappa \geq 0$, the argument works verbatim, the only difference being in the application of the Bishop-Gromov inequality. □

Finally, let us conclude by stating an important consequence of CD condition.

Lemma 2.1.2. *Assume that* $(M, \mathsf{d}, \mathfrak{m})$ *satisfies* (CD). *The relative entropy functional* $\mathscr{H}(\cdot|\mathfrak{m})$ *is lower semicontinuous with respect to* $\mathbf{W}_2$*-convergence.*

Proof. This follows from [Stu06a, Theorem 4.24], which allows to choose W as $C\mathsf{d}^2(\cdot, x_0)$ for any $x_0 \in M$ and $C > 0$ sufficiently large, in the definition of relative entropy (1.A.3). □

2.2 Structural properties of Schrödinger problem

In order to guarantee existence and uniqueness of solution to SP we are going to further assume the validity of

A1. *The marginals* $\mu, \nu \in \mathcal{P}_2(M)$ *have finite second moments and finite relative entropies with respect to the reference equilibrium measure* $\mathfrak{m}$, *i.e.,*

$$\mathscr{H}(\mu|\mathfrak{m}),\ \mathscr{H}(\nu|\mathfrak{m}) < +\infty\,.$$

Notice that as soon as $\mathsf{CD}(\kappa, \infty)$ holds for some positive $\kappa > 0$ and $\mathfrak{m}(M) = 1$ and we further assume $M_2(\mathfrak{m}) < +\infty$, then the finite second moments' assumption can be dropped since Talagrand's transportation inequality holds [BGL13, Corollary 9.3.2] and hence for any probability measure $\mathfrak{p} \in \mathcal{P}(M)$ with finite relative entropy $\mathscr{H}(\mathfrak{p}|\mathfrak{m})$ it holds

$$M_2(\mathfrak{p}) \leq 2(M_2(\mathfrak{m}) + \mathbf{W}_2^2(\mathfrak{p}, \mathfrak{m})) \leq 2M_2(\mathfrak{m}) + \frac{4}{\kappa}\mathscr{H}(\mathfrak{p}|\mathfrak{m})\,.$$

Let us further point out that in the standard Euclidean setting $\mathsf{CD}(\kappa,\infty)$ with $\kappa > 0$ guarantees $\mathfrak{m}$ to satisfy the Poincarè inequality with rate $1/\kappa$, which implies $M_2(\mathfrak{m}) < +\infty$. Henceforth Assumption **A**1 boils down to asking for the marginals to have solely finite relative entropies, which is a strictly necessary condition for SP to be well-defined. Indeed the data-processing inequality (1.A.5) implies that for any coupling $\pi \in \Pi(\mu,\nu)$

$$\mathscr{H}(\mu|\mathfrak{m}),\, \mathscr{H}(\nu|\mathfrak{m}) \leq \mathscr{H}(\pi|\mathrm{R}_{0,T})\,,$$

where the right hand side ought to be finite for at least one coupling in order to make SP a well-defined problem.

We are now ready to state the main result of this section. We postpone its proof after a few remarks and important first consequences.

Theorem 2.2.1. *Let $(M,\mathsf{d},\mathfrak{m})$ satisfy* (CD) *and suppose μ, ν satisfy **A1**. Then SP admits a unique minimiser $\boldsymbol{\pi}^T \in \Pi(\mu,\nu)$ and there are two non-negative measurable functions f, g satisfying*

$$\frac{\mathrm{d}\boldsymbol{\pi}^T}{\mathrm{dR}_{0,T}}(x,y) = f(x)\,g(y) \quad \mathrm{R}_{0,T}\text{-}a.e.\,. \tag{2.2.1}$$

They are $\mathfrak{m}$-a.e. unique up to the trivial transformation $(f,g) \mapsto (cf, g/c)$ for some $c > 0$ and moreover $\varphi = -\log f \in \mathrm{L}^1(\mu)$, $\psi = -\log g \in \mathrm{L}^1(\nu)$. Finally, f and g satisfy the Schrödinger system

$$\begin{cases} \mu = f\,P_T g\,\mathfrak{m}\,, \\ \nu = P_T f\,g\,\mathfrak{m}\,. \end{cases} \tag{2.2.2}$$

The splitting decomposition (2.2.1) is often referred to as the *fg-decomposition*. For later reference, we will refer to the functions $\varphi \in \mathrm{L}^1(\mu)$, $\psi \in \mathrm{L}^1(\nu)$ as to the *Schrödinger potentials* and from (2.2.1) we immediately have

$$\frac{\mathrm{d}\boldsymbol{\pi}^T}{\mathrm{dR}_{0,T}} = e^{-\varphi\oplus\psi} \quad \mathrm{R}_{0,T}\text{-a.e.}\,. \tag{2.2.3}$$

Furthermore, let us notice that the Schrödinger potentials (actually their opposites $(-\varphi)$, $(-\psi)$) can be seen as the optimiser of a dual characterisation of SP, similarly to what happens in the OT setting with the Kantorovich potentials in (1.2.9). Namely, it holds

Proposition 2.2.2. *Assume $(M,\mathsf{d},\mathfrak{m})$ satisfies* (CD) *and suppose the marginals μ, ν satisfy **A1**. Then $\mathcal{C}_T(\mu,\nu) < \infty$ and*

$$\mathcal{C}_T(\mu,\nu) = \sup_{\alpha,\beta\in\mathcal{M}_b(M)} \left\{ \int_M \alpha\,\mathrm{d}\mu + \int_M \beta\,\mathrm{d}\nu - \log\int_{M\times M} e^{\alpha\oplus\beta}\,\mathrm{dR}_{0,T} \right\}.$$

Finally, the supremum is attained at the couple $(-\varphi,-\psi)$.

Proof. We have already seen in Theorem 2.2.1 that $\mathcal{C}_T(\mu,\nu)$ is finite. Now, from [Léo01, Proposition 6.1] it follows

$$\begin{aligned}
\mathcal{C}_T(\mu,\nu) &= \sup_{\alpha,\beta\in\mathcal{M}_b(M)}\left\{\int_M \alpha\,\mathrm{d}\mu + \int_M \beta\,\mathrm{d}\nu - \int_{M\times M}\left(e^{\alpha\oplus\beta}-1\right)\mathrm{dR}_{0,T}\right\}\\
&\le \sup_{\alpha,\beta\in\mathcal{M}_b(M)}\left\{\int_M \alpha\,\mathrm{d}\mu + \int_M \beta\,\mathrm{d}\nu - \log\int_{M\times M} e^{\alpha\oplus\beta}\,\mathrm{dR}_{0,T}\right\}\\
&= \sup_{\alpha,\beta\in\mathcal{M}_b(M)}\left\{\int_{M\times M}(\alpha\oplus\beta)\,\mathrm{d}\boldsymbol{\pi}^T - \log\int_{M\times M} e^{\alpha\oplus\beta}\,\mathrm{dR}_{0,T}\right\}\\
&\le \sup_{h\in\mathcal{M}_b(M\times M)}\left\{\int_{M\times M} h\,\mathrm{d}\boldsymbol{\pi}^T - \log\int_{M\times M} e^{h}\,\mathrm{dR}_{0,T}\right\} \overset{(1.A.2)}{=} \mathscr{H}\left(\boldsymbol{\pi}^T|\mathrm{R}_{0,T}\right).
\end{aligned}$$

Since $\mathcal{C}_T(\mu,\nu) = \mathscr{H}\left(\boldsymbol{\pi}^T|\mathrm{R}_{0,T}\right)$ the above chain of inequalities is a chain of equivalences and this, combined with (2.2.3), concludes our proof. □

Let us also immediately provide here a remarkable consequence of the decomposition proven in Theorem 2.2.1, which is usually referred to as *Pythagoras Theorem for entropic projections*:

Corollary 2.2.3. *Let* $(M,\mathsf{d},\mathfrak{m})$ *satisfy* (CD)*, suppose the marginals* μ*,* ν *satisfy **A1** and let* $\boldsymbol{\pi}^T\in\Pi(\mu,\nu)$ *be the unique optimiser in SP. Then for any coupling* $\pi\in\Pi(\mu,\nu)$ *with* $\mathscr{H}(\pi|\mathrm{R}_{0,T})<+\infty$ *it holds*

$$\mathscr{H}(\pi|\boldsymbol{\pi}^T) = \mathscr{H}(\pi|\mathrm{R}_{0,T}) - \mathscr{H}(\boldsymbol{\pi}^T|\mathrm{R}_{0,T})$$

Proof. It is enough noticing that Theorem 2.2.1 implies

$$\begin{aligned}
\mathscr{H}(\pi|\boldsymbol{\pi}^T) &= \int \log\frac{\mathrm{d}\pi}{\mathrm{d}\boldsymbol{\pi}^T}\,\mathrm{d}\pi = \int\left(\log\frac{\mathrm{d}\pi}{\mathrm{dR}_{0,T}} - \log\frac{\mathrm{d}\boldsymbol{\pi}^T}{\mathrm{dR}_{0,T}}\right)\mathrm{d}\pi\\
&= \int\left(\log\frac{\mathrm{d}\pi}{\mathrm{dR}_{0,T}}(x,y) + \varphi(x) + \psi(y)\right)\mathrm{d}\pi(x,y),
\end{aligned}$$

where the last step holds since (2.2.3) holds $\mathrm{R}_{0,T}$-a.e. and hence also π-a.e. (since $\mathscr{H}(\pi|\mathrm{R}_{0,T})<+\infty$). Moreover, $\mathscr{H}(\pi|\mathrm{R}_{0,T})<+\infty$ and Theorem 2.2.1 implies that the above summands are all in $\mathrm{L}^1(\pi)$ (since $\varphi\in\mathrm{L}^1(\mu)$ and $\psi\in\mathrm{L}^1(\nu)$) and therefore we have

$$\begin{aligned}
\mathscr{H}(\pi|\boldsymbol{\pi}^T) &= \int\log\frac{\mathrm{d}\pi}{\mathrm{dR}_{0,T}}\,\mathrm{d}\pi + \int\varphi\mathrm{d}\mu + \int\psi\mathrm{d}\nu\\
&= \mathscr{H}(\pi|\mathrm{R}_{0,T}) - \int\log\frac{\mathrm{d}\boldsymbol{\pi}^T}{\mathrm{R}_{0,T}}\,\mathrm{d}\boldsymbol{\pi}^T = \mathscr{H}(\pi|\mathrm{R}_{0,T}) - \mathscr{H}(\boldsymbol{\pi}^T|\mathrm{R}_{0,T}).
\end{aligned}$$

□

For sake of clarity we will split the proof of Theorem 2.2.1 in two parts: firstly we are going to show the existence and uniqueness of the optimal coupling π^T and then we will characterise it with the fg-decomposition.

Proof of existence and uniqueness in Theorem 2.2.1. Firstly, let us note that we may equivalently consider the minimisation problem

$$\inf_{\Pi(\mu,\nu)} \mathscr{H}(\cdot|\mathrm{R}^{\mu}_{0,T}) \tag{2.2.4}$$

where $\mathrm{R}^{\mu}_{0,T} \in \mathcal{P}(\mathbb{R}^{2d})$ is defined as $\mathrm{R}^{\mu}_{0,T}(\mathrm{d}x,\mathrm{d}y) := \mathsf{p}_T(x,y)\mu(\mathrm{d}x)\mathfrak{m}(\mathrm{d}y)$. Indeed it is enough to note that on $\Pi(\mu,\nu)$ it holds

$$\mathscr{H}(\cdot|\mathrm{R}_{0,T}) = \mathscr{H}(\mu|\mathfrak{m}) + \mathscr{H}(\cdot|\mathrm{R}^{\mu}_{0,T}),$$

identity which follows from the definition of relative entropy (consider $W = -\log \mathrm{d}\mu/\mathrm{d}\mathfrak{m}$ in the unbounded σ-finite case). This shows that (2.2.4) and SP are equivalent and share the same solutions.

Given the above premises, we should prove that (2.2.4) (and hence SP) is a well-posed problem, *i.e.*, that there exists at least one coupling $\pi \in \Pi(\mu,\nu)$ such that $\mathscr{H}(\pi|\mathrm{R}^{\mu}_{0,T})$ is finite. Particularly we are going to show that this is the case for the independent coupling $\mu \otimes \nu$.

In order to prove that, let us derive a lower bound for $\log \mathsf{p}_T$. First, assume that $(M, \mathsf{d}, \mathfrak{m})$ satisfies $\mathsf{CD}(\kappa,\infty)$ for some $\kappa \in \mathbb{R}$ and $\mathfrak{m}(M) = 1$. From (2.1.8) we get the lower bound

$$\log \mathsf{p}_T(x,y) \geq -\frac{\kappa\, \mathsf{d}^2(x,y)}{2(e^{\kappa T}-1)}, \tag{2.2.5}$$

which, combined with the trivial inequalities $\mathsf{d}(x,y)^2 \leq 2(\mathsf{d}^2(x,z) + \mathsf{d}^2(y,z))$ valid for any $z \in M$ and $e^{\kappa T} - 1 \geq \kappa T$, implies

$$\log \mathsf{p}_T(x,y) \geq -\frac{\mathsf{d}^2(x,z)}{T} - \frac{\mathsf{d}^2(y,z)}{T}. \tag{2.2.6}$$

On the other hand, if $\mathsf{CD}(\kappa,N)$ holds with $N < +\infty$, then from the heat kernel lower bound (2.1.7) we obtain

$$\log \mathsf{p}_T(x,y) \geq -\log\left[C_1\, \mathfrak{m}\left(B_{\sqrt{T}}(x)\right)\right] - C_2\, T - \frac{\mathsf{d}^2(x,z)}{T} - \frac{\mathsf{d}^2(y,z)}{T}. \tag{2.2.7}$$

In both cases, thanks to Assumption **A**1 and Lemma 2.1.1 with $r = \sqrt{T}$ we conclude that $-\int \log \mathsf{p}_T\, \mathrm{d}(\mu\otimes\nu) \in [-\infty,+\infty)$. Therefore we may consider the well-defined summation

$$\mathscr{H}(\mu\otimes\nu|\mathrm{R}^{\mu}_{0,T}) = \mathscr{H}(\nu|\mathfrak{m}) - \int \log \mathsf{p}_T\, \mathrm{d}(\mu\otimes\nu) < +\infty. \tag{2.2.8}$$

Now, since $\Pi(\mu,\nu)$ is a weakly-closed subset of $\mathcal{P}(M\times M)$ and $\mathscr{H}(\cdot|\mathrm{R}^{\mu}_{0,T})$ has weakly-compact level-sets (cf. Lemma 1.A.1), the subset

$$\{\pi\in\Pi(\mu,\nu)\colon \mathscr{H}(\pi|\mathrm{R}^{\mu}_{0,T})\leq \mathscr{H}(\mu\otimes\nu|\mathrm{R}^{\mu}_{0,T})\}$$

is weakly-compact. Therefore, from the lower-semicontinuity of $\mathscr{H}(\cdot|\mathrm{R}^{\mu}_{0,T})$ and its strict convexity (cf. Lemma 1.A.1), we may finally deduce that there exists a unique minimiser $\boldsymbol{\pi}^T\in\Pi(\mu,\nu)$ such that

$$\mathscr{H}(\boldsymbol{\pi}^T|\mathrm{R}^{\mu}_{0,T})=\inf_{\Pi(\mu,\nu)}\mathscr{H}(\cdot|\mathrm{R}^{\mu}_{0,T})\,.$$

By recalling that SP and (2.2.4) are equivalent, we may finally conclude that $\boldsymbol{\pi}^T$ is the unique minimiser for SP. □

Before proceeding with the remaining part of the proof, let us spend some lines on the marginals' constraint $\pi\in\Pi(\mu,\nu)$ and how we could already infer the validity of the fg-decomposition from it. Let us start by noticing that since (M,d) is separable we know that there exists a countable dense family of bounded measurable functions $\{\phi_i\}_{i\in\mathbb{N}}$ and $\{\psi_i\}_{i\in\mathbb{N}}$ such that

$$(\mathrm{proj}_{x_1})_{\#}\pi=\mu\quad\Leftrightarrow\quad\int\phi_i(x)\mathrm{d}\pi(x,y)=\int\phi_i\mathrm{d}\mu\qquad\forall\, i\in\mathbb{N}$$

and similarly for $(\mathrm{proj}_{x_2})_{\#}\pi=\nu$ and $\{\psi_i\}_{i\in\mathbb{N}}$. Therefore one might approximate SP (or equivalently (2.2.4)), for any finite $K\in\mathbb{N}$, with the following minimisation problem

$$\inf_{\mathcal{Q}_K}\mathscr{H}(\cdot|\mathrm{R}^{\mu}_{0,T})\,, \tag{2.2.9}$$

where

$$\mathcal{Q}_K:=\left\{\pi\in\mathcal{P}(M^2)\colon\int\phi_i\,\mathrm{d}\pi=\int\phi_i\,\mathrm{d}\mu\,,\int\psi_i\,\mathrm{d}\pi=\int\psi_i\,\mathrm{d}\nu\ \text{ for all } i\leq K\right\}.$$

Particularly, notice that $\mathcal{Q}_K$ is a convex set and it is defined via a finite number of linear constraints, in contrast to what happens for $\Pi(\mu,\nu)$ where the marginals' constraint is an infinite-dimensional one. This allows to deduce, via Lagrange multipliers, that indeed the density of the unique solution $\boldsymbol{\pi}^K$ of (2.2.9) can be decomposed as

$$\frac{\mathrm{d}\boldsymbol{\pi}^K}{\mathrm{d}\mathrm{R}^{\mu}_{0,T}}\propto\exp\left(\sum_{i=1}^K a_i\,\phi_i\oplus\sum_{i=1}^K b_i\,\psi_i\right)$$

for some weights $a_i,b_i\in\mathbb{R}$. If we introduce the measurable functions $\varphi^K:=-\sum_{i=1}^K a_i\,\phi_i\in\mathcal{C}_b(M)$ and $\psi^K:=-\sum_{i=1}^K b_i\,\psi_i\in\mathcal{C}_b(M)$, then clearly the fg-decomposition for $\boldsymbol{\pi}^K$ holds with $f_K:=\exp(-\varphi^K)$ and $g_K:=\exp(-\psi^K)$. We refer the reader to [Nut21, Example 1.18] (which is based on [FS11, Section 3]) where a more general situation is discussed and where it is proven such decomposition via Lagrange multipliers. Therefore one possible approach in the proof of the fg-decomposition in Theorem 2.2.1 would be showing the convergence of the approximated problem (2.2.9) towards SP as $K \uparrow +\infty$ and showing that $\varphi^K \oplus \psi^K$ converges in $\mathrm{L}^1(\boldsymbol{\pi}^T)$ to a limit element which belongs to the sum space $\mathrm{L}^1(\mu) \oplus \mathrm{L}^1(\nu)$. Clearly $\varphi^K \oplus \psi^K \in \mathrm{L}^1(\mu) \oplus \mathrm{L}^1(\nu)$ since it is a continuous bounded function, therefore it would be enough showing that $\mathrm{L}^1(\mu) \oplus \mathrm{L}^1(\nu)$ is a closed subspace of $\mathrm{L}^1(\boldsymbol{\pi}^T)$. This can be done for instance by arguing as in [RT93, Proposition 1], under some additional regularity assumptions.

The above sketched approach is the one adopted in [Nut21, Section 3]. In what follows, we are going to present a different approach, namely the one presented in [Tam17, Proposition 4.1.5] (see also [GT21, Proposition 2.1]), combined with some integrability results coming from [Nut21], which so far is the most general approach presented in the literature. Let us just mention here that the closure of $\mathrm{L}^1(\mu) \oplus \mathrm{L}^1(\nu)$ plays a crucial role for the existence of the fg-decomposition and indeed also our approach requires an appropriate direct sum space to be close in $\mathrm{L}^1(\boldsymbol{\pi}^T)$ (cf. **Step 1.** in the proof of Lemma 2.B.1).

We may finally proceed with the proof of the fg-decomposition.

Proof of the fg-decomposition in Theorem 2.2.1. Let $\boldsymbol{\pi}^T$ be the unique solution of SP and let $A_\mu := \{\mathrm{d}\mu/\mathrm{d}\mathfrak{m} > 0\}$ and $A_\nu := \{\mathrm{d}\nu/\mathrm{d}\mathfrak{m} > 0\}$. Consider the function spaces

$$\begin{aligned} V_+ &:= \Big(\mathrm{L}^0(M, \mathfrak{m}_{|_{A_\mu}}) \oplus \mathrm{L}^0(M, \mathfrak{m}_{|_{A_\nu}})\Big) \cap \mathrm{L}^1(M^2, \boldsymbol{\pi}^T)\,, \\ V_0 &:= \{\ell \in \mathrm{L}^\infty(M^2, \boldsymbol{\pi}^T) \colon (\mathrm{proj}_{x_1})_\#(\ell \boldsymbol{\pi}^T) = (\mathrm{proj}_{x_2})_\#(\ell \boldsymbol{\pi}^T) = 0\}\,, \\ V_+^\perp &:= \{\ell \in \mathrm{L}^\infty(M^2, \boldsymbol{\pi}^T) \colon \int u\ell \,\mathrm{d}\boldsymbol{\pi}^T = 0 \quad \forall\, u \in V_+\}\,, \\ {}^\perp V_0 &:= \{u \in \mathrm{L}^1(M^2, \boldsymbol{\pi}^T) \colon \int u\ell \,\mathrm{d}\boldsymbol{\pi}^T = 0 \quad \forall\, \ell \in V_0\}\,. \end{aligned} \tag{2.2.10}$$

Claim 1. We claim that

$$\frac{\mathrm{d}\boldsymbol{\pi}^T}{\mathrm{d}\mathrm{R}^\mu_{0,T}} > 0 \quad \mathfrak{m} \otimes \mathfrak{m}\text{-}a.s. \text{ on the set } A_\mu \times A_\nu\,. \tag{2.2.11}$$

Assume by contradiction that the above is false and hence, since $\mu \otimes \nu \sim \mathfrak{m} \otimes \mathfrak{m}$ on $A_\mu \times A_\nu$, it holds $\mu \otimes \nu(Z) > 0$ where $Z := (A_\mu \times A_\nu) \cap \{\mathrm{d}\boldsymbol{\pi}^T/\mathrm{d}\mathrm{R}^\mu_{0,T} = 0\}$. Now, for any $\delta \in (0,1)$ consider the probability measure $\boldsymbol{\pi}^{T,\delta} \in \Pi(\mu,\nu)$ defined as the convex combination between $\boldsymbol{\pi}^T$ and the independent coupling $\mu \otimes \nu$, *i.e.*, the probability measure

$$\boldsymbol{\pi}^{T,\delta} := (1-\delta)\,\boldsymbol{\pi}^T + \delta\,\mu \otimes \nu\,.$$

Then, if we consider the convex function $h(a) := a\log a$, we may deduce that

$$h\left(\frac{\mathrm{d}\boldsymbol{\pi}^{T,\delta}}{\mathrm{dR}^\mu_{0,T}}\right) \leq \begin{cases} h\left(\delta\frac{\mathrm{d}(\mu\otimes\nu)}{\mathrm{dR}^\mu_{0,T}}\right) = h(\delta)\frac{\mathrm{d}(\mu\otimes\nu)}{\mathrm{dR}^\mu_{0,T}} + \delta\, h\left(\frac{\mathrm{d}(\mu\otimes\nu)}{\mathrm{dR}^\mu_{0,T}}\right) & \text{on } Z \\ \delta\, h\left(\frac{\mathrm{d}(\mu\otimes\nu)}{\mathrm{dR}^\mu_{0,T}}\right) + (1-\delta)\, h\left(\frac{\mathrm{d}\boldsymbol{\pi}^T}{\mathrm{dR}^\mu_{0,T}}\right) & \text{on } M^2\setminus Z \end{cases}$$

which, combined with the optimality of $\boldsymbol{\pi}^T$ as unique solution of SP, implies that

$$\begin{aligned} 0 &\leq \lim_{\delta\downarrow 0} \frac{\mathscr{H}(\boldsymbol{\pi}^{T,\delta}|\mathrm{R}^\mu_{0,T}) - \mathscr{H}(\boldsymbol{\pi}^T|\mathrm{R}^\mu_{0,T})}{\delta} \\ &= \lim_{\delta\downarrow 0}\delta^{-1}\int h(\mathrm{d}\boldsymbol{\pi}^{T,\delta}/\mathrm{dR}^\mu_{0,T}) - h(\mathrm{d}\boldsymbol{\pi}^T/\mathrm{dR}^\mu_{0,T})\,\mathrm{dR}^\mu_{0,T} \\ &\leq \mathscr{H}(\mu\otimes\nu|\mathrm{R}^\mu_{0,T}) - \mathscr{H}(\boldsymbol{\pi}^T|\mathrm{R}^\mu_{0,T}) + \lim_{\delta\downarrow 0}\log(\delta)\,\mu\otimes\nu(Z) = -\infty \end{aligned}$$

which is clearly a contradiction.

Claim 2. We claim that ${}^\perp V_0 \subseteq V_+$. For exposition's sake we postpone the proof of this claim to Lemma 2.B.1 in the Appendix.

Claim 3. Next, we claim that $\log \mathrm{d}\boldsymbol{\pi}^T/\mathrm{dR}_{0,T} \in {}^\perp V_0$, *i.e.*, that for any $\ell \in V_0$ it holds

$$\int \ell\, h\left(\frac{\mathrm{d}\boldsymbol{\pi}^T}{\mathrm{dR}_{0,T}}\right)\mathrm{dR}^\mu_{0,T} = \int \ell \log\frac{\mathrm{d}\boldsymbol{\pi}^T}{\mathrm{dR}_{0,T}}\,\mathrm{d}\boldsymbol{\pi}^T = 0\,. \tag{2.2.12}$$

Hence, pick $\ell \in \mathrm{L}^\infty(M^2,\boldsymbol{\pi}^T)$ with $(\mathrm{proj}_{x_1})_\#(\ell\boldsymbol{\pi}^T) = (\mathrm{proj}_{x_2})_\#(\ell\boldsymbol{\pi}^T) = 0$ and for any $\delta \in (0, \|\ell\|^{-1}_{\mathrm{L}^\infty(M^2,\boldsymbol{\pi}^T)})$ consider the variation along ℓ of $\boldsymbol{\pi}^T$, that is the coupling probability measure $(1+\delta\,\ell)\boldsymbol{\pi}^T \in \Pi(\mu,\nu)$. Then from the optimality of $\boldsymbol{\pi}^T$ we may once again deduce that

$$\begin{aligned} 0 &\leq \frac{\mathscr{H}((1+\delta\,\ell)\boldsymbol{\pi}^T|\mathrm{R}^\mu_{0,T}) - \mathscr{H}(\boldsymbol{\pi}^T|\mathrm{R}^\mu_{0,T})}{\delta} \\ &= \frac{1}{\delta}\int h\left((1+\delta\,\ell)\frac{\mathrm{d}\boldsymbol{\pi}^T}{\mathrm{dR}_{0,T}}\right) - h\left(\frac{\mathrm{d}\boldsymbol{\pi}^T}{\mathrm{dR}_{0,T}}\right)\mathrm{dR}^\mu_{0,T} \\ &= \int\left(\ell h\left(\frac{\mathrm{d}\boldsymbol{\pi}^T}{\mathrm{dR}_{0,T}}\right) + \delta^{-1}\, h(1+\delta\,\ell)\frac{\mathrm{d}\boldsymbol{\pi}^T}{\mathrm{dR}_{0,T}}\right)\mathrm{dR}^\mu_{0,T} \\ &\leq \int\left(\ell h\left(\frac{\mathrm{d}\boldsymbol{\pi}^T}{\mathrm{dR}_{0,T}}\right) + \ell\,(1+\delta\,\ell)\frac{\mathrm{d}\boldsymbol{\pi}^T}{\mathrm{dR}_{0,T}}\right)\mathrm{dR}^\mu_{0,T} \end{aligned}$$

where the last step follows from $1+\delta\ell > 0$ and the standard inequality $\log(1+a) \leq a$ for any $a \in (-1,+\infty)$. Since $\ell \in \mathrm{L}^\infty(M^2,\boldsymbol{\pi}^T)$ and $\mathscr{H}(\boldsymbol{\pi}^T|\mathrm{R}^\mu_{0,T}) < +\infty$ the Dominated Convergence Theorem implies that as soon as $\delta\downarrow 0$ it holds

$$0 \leq \int \ell h\left(\frac{\mathrm{d}\boldsymbol{\pi}^T}{\mathrm{dR}_{0,T}}\right)\mathrm{dR}^\mu_{0,T} + \int \ell\,\mathrm{d}\boldsymbol{\pi}^T = \int \ell h\left(\frac{\mathrm{d}\boldsymbol{\pi}^T}{\mathrm{dR}_{0,T}}\right)\mathrm{dR}^\mu_{0,T}\,,$$

where the last step follows from the definition of V_0 and Fubini's Theorem.

By considering $-\ell \in V_0$ in the above discussion, the last inequality reads as an equality, which implies that (2.2.12) holds true for any $\ell \in V_0$. This proves our third claim.

Conclusion. In order to conclude it is enough combining the second and third claim which imply that $\log \mathrm{d}\boldsymbol{\pi}^T/\mathrm{dR}^{\mu}_{0,T} \in V_+$. Particularly, this guarantees the existence of two measurable functions $\bar{\varphi} \in \mathrm{L}^0(M, \mathfrak{m}_{|A_\mu})$ and $\psi \in \mathrm{L}^0(M, \mathfrak{m}_{|A_\nu})$ such that $\log \mathrm{d}\boldsymbol{\pi}^T/\mathrm{dR}^{\mu}_{0,T} = -\bar{\varphi} \oplus \psi$ and hence that $\mathrm{R}_{0,T}$-a.e. on $A_\mu \times A_\nu$

$$\frac{\mathrm{d}\boldsymbol{\pi}^T}{\mathrm{dR}_{0,T}}(x,y) = \frac{\mathrm{d}\boldsymbol{\pi}^T}{\mathrm{dR}^{\mu}_{0,T}}(x,y)\,\frac{\mathrm{d}\mu}{\mathrm{d}\mathfrak{m}}(x) = f(x)\,g(y)\,,$$

where $g := \exp(-\psi)$ and $f := \exp(-\bar{\varphi} + \log \mathrm{d}\mu/\mathrm{d}\mathfrak{m})$. Outside A_μ and A_ν is enough extending f and respectively g trivially equal to zero. The uniqueness part of the claim is also trivial.

Lastly, the (disjoint) integrability of φ and ψ can be discussed as follows. Indeed the bound $\mathscr{H}(\mu \otimes \nu | \mathrm{R}^{\mu}_{0,T}) < +\infty$, [Nut21, Corollary 1.13] guarantees that

$$\bar{\varphi} \oplus \psi = -\log \frac{\mathrm{d}\boldsymbol{\pi}^T}{\mathrm{dR}^{\mu}_{0,T}} \in \mathrm{L}^1(\mu \otimes \nu)$$

which combined with Fubini's Theorem implies $\bar{\varphi} \in \mathrm{L}^1(\mu)$ and $\psi \in \mathrm{L}^1(\nu)$ (see also [Nut21, Remark 2.22] for a comprehensive discussion on this last step). Since $\mathscr{H}(\mu|\mathfrak{m}) < +\infty$ implies that $\log \mathrm{d}\mu/\mathrm{d}\mathfrak{m} \in \mathrm{L}^1(\mu)$ (cf. Lemma 1.A.2) we can conclude that also $\varphi = \bar{\varphi} - \log \mathrm{d}\mu/\mathrm{d}\mathfrak{m} \in \mathrm{L}^1(\mu)$. □

Notice that the above fg-decomposition is unique up to a scalar multiplicative factor, or equivalently that the couple of potentials (φ, ψ) is unique up to additive shift, *i.e.*, $(\varphi, \psi) \mapsto (\varphi + a, \psi - a)$ for any $a \in \mathbb{R}$. Therefore, unless differently specified, in this thesis we are going to assume the following symmetric normalisation

$$\int \varphi \,\mathrm{d}\mu + \mathscr{H}(\mu|\mathfrak{m}) = \int \psi \,\mathrm{d}\nu + \mathscr{H}(\nu|\mathfrak{m}) = \frac{1}{2}\left[\mathscr{H}(\mu|\mathfrak{m}) + \mathscr{H}(\nu|\mathfrak{m}) - \mathcal{C}_T(\mu,\nu)\right]. \tag{2.2.13}$$

In most of this thesis we will always deal with marginals that are defined on unbounded supports. However, we will sometimes perform approximating arguments in our proofs in order to consider bounded and compactly supported marginals, since this property will be inherited in the fg-decomposition.

Lemma 2.2.4. *Let* $(M, \mathsf{d}, \mathfrak{m})$ *satisfy* (CD)*, suppose that the two marginals* μ*,* ν *satisfy* ***A1*** *and that their densities (w.r.t.* $\mathfrak{m}$*) are bounded and compactly supported. Then* f, g *considered in Theorem 2.2.1 are in* $\mathrm{L}^1(\mathfrak{m}) \cap \mathrm{L}^\infty(\mathfrak{m})$*.*

Moreover, if we further assume that $\mathrm{d}\mu/\mathrm{d}\mathfrak{m} \in \mathcal{C}^k(M)$ *for some* $k \in \mathbb{N} \cup \{\infty\}$ *(resp.* $\mathrm{d}\nu/\mathrm{d}\mathfrak{m}$*) then the measurable function* f *(resp.* g*) also belongs to* $\mathcal{C}^k(M)$*.*

Proof. The boundedness result follows straightforwardly from [GT21, Proposition 2.1-(ii)] since the Gaussian lower bounds (2.2.6) and (2.2.7), combined with the bounded densities and bounded supports assumption, guarantee the validity of the assumptions considered there.

The proof of the $\mathcal{C}^k(M)$-regularity inheritance is given in [GLRT20, Proposition 2.8-b-(iii)]. □

In the next chapter we are going to see how the Schrödinger potentials and their gradients provide good proxies for the optimal transport problem. More precisely we will show that the Schrödinger map[2], which is defined as the measurable map

$$\boldsymbol{\mathcal{T}}^T := \mathrm{Id} + 2\nabla\varphi^T \tag{2.2.14}$$

will converge to the optimal transport map $\boldsymbol{\mathcal{T}}$ in $\mathrm{L}^2(\mu)$. Therefore it would be interesting understanding what properties the gradients of the Schrödinger potentials inherit from the marginals (e.g., Lipschitzianity of potentials). The next theorem gives a first partial result in this direction in the classical Euclidean setting, *i.e.*, when considering the Brownian motion SP (as introduced in (1.2.2)) with the Gaussian measure as reference measure $\mathrm{R}_{0,T} \propto \exp(-|x-y|^2/_{2T})$. We will show that the Schrödinger potentials inherit some convexity from the log-concavity/convexity of the two marginals. Therefore, let us consider the following assumption:

A2. *Let U_μ, U_ν denote the (negative) log-densities of the marginals, i.e., such that*

$$\mu(\mathrm{d}x) = \exp(-U_\mu(x))\mathrm{d}x\,, \qquad \nu(\mathrm{d}x) = \exp(-U_\nu(x))\mathrm{d}x\,.$$

Assume that there exist $\alpha_\nu \in (0,+\infty)$ and $\beta_\mu \in (0,+\infty]$ such that

$$\nabla^2 U_\nu \geq \alpha_\nu \quad \textit{and} \quad \nabla^2 U_\mu \leq \beta_\mu\,.$$

The result we are going to prove reads as follows

Theorem 2.2.5. *Consider the classical SP in $\mathbb{R}^d$ introduced in* (1.2.2)*, with Gaussian reference* $\mathrm{R}_{0,T} \propto \exp(-|x-y|^2/_{2T})$. *Assume the validity of **A1** and **A2**. Then it holds*

$$\nabla^2\psi \geq \alpha_\psi \quad \textit{with } \alpha_\psi := \begin{cases} \frac{1}{2}\left(\alpha_\nu + \sqrt{\alpha_\nu^2 + 4\alpha_\nu/(\beta_\mu T^2)}\right) - T^{-1} & \textit{if } \beta_\mu < +\infty\,, \\ \alpha_\nu - T^{-1} & \textit{for } \beta_\mu = +\infty\,. \end{cases} \tag{2.2.15}$$

Moreover, if we consider for any $t \in [0,T]$ the function $\mathcal{U}_t^{T,\psi} := -\log P_{T-t}\exp(-\psi)$ then it further holds

$$\nabla^2\mathcal{U}_t^{T,\psi} \geq \frac{\alpha_\psi}{1+(T-t)\alpha_\psi}\,.$$

[2]We should point out that $\boldsymbol{\mathcal{T}}^T$ does not define a transport map between μ and ν since in general $\boldsymbol{\mathcal{T}}^T_\# \mu \neq \nu$

We will not prove here this result now since in Chapter 6 we will prove a more general result that implies this one as a direct consequence (cf. Theorems 6.1.1 and 6.6.1). A direct proof of (2.2.15) can otherwise be found in [CP23], where the authors provide an entropic proof of Caffarelli's Theorem (*i.e.*, that under **A**2 the optimal transport map $\mathcal{T}$ from μ to ν is $\sqrt{\beta_\mu/\alpha_\nu}$-Lipschitz). The approach we are going to employ later in the proof of Theorem 2.2.5 is based on the study of convexity propagation along Sinkhorn's algorithm, that is an iterative algorithm that computes Schrödinger potentials.

2.2.1 Sinkhorn's algorithm

If we suppose that the marginals admit densities of the form

$$\mu(\mathrm{d}x) = \exp(-U_\mu(x))\mathfrak{m}(\mathrm{d}x)\,, \qquad \nu(\mathrm{d}x) = \exp(-U_\nu(x))\mathfrak{m}(\mathrm{d}x)\,, \tag{2.2.16}$$

then, the Schrödinger system (2.2.2) equivalently reads as

$$\begin{cases} \varphi = U_\mu + \log P_T \exp(-\psi) \\ \psi = U_\nu + \log P_T \exp(-\varphi)\,, \end{cases} \tag{2.2.17}$$

where $(P_t)_{t\geq 0}$ is the Markov semigroup generated by the SDE (2.1.1).

Then, starting from a given initialisation $\psi^0, \varphi^0 \colon \mathbb{R}^d \to \mathbb{R}$ (usually $\varphi^0 = 0$ and $\psi^0 := U_\nu$), one may consider an iterative algorithm that solves (2.2.17) as a fixed point problem by generating two sequences of potentials $\{\varphi^n, \psi^n\}_{n\in\mathbb{N}}$, called Sinkhorn potentials, according to the following recursive scheme:

$$\begin{cases} \varphi^{n+1} = U_\mu + \log P_T \exp(-\psi^n) \\ \psi^{n+1} = U_\nu + \log P_T \exp(-\varphi^{n+1})\,. \end{cases} \tag{2.2.18}$$

The above algorithm is known as Sinkhorn's algorithm or as Iterative Proportional Fitting Procedure (hereafter IPFP). Its convergence has been extensively studied, specifically for compact spaces or for compactly supported marginals. We postpone the literature review on the convergence of Sinkhorn's algorithm to the bibliographical remarks section in Chapter 6 where we prove its exponential convergence for unbounded costs and unbounded marginals.

Below we are just going to interpret Sinkhorn's algorithm from the primal point of view, *i.e.*, when considering the coupling probability measures defined (by mimicking (2.2.3)) via

$$\begin{aligned} \mathrm{d}\pi^{n+1,n} &\propto \exp(-\varphi^{n+1} \oplus \psi^n)\mathrm{d}\mathrm{R}_{0,T}\,, \\ \mathrm{d}\pi^{n+1,n+1} &\propto \exp(-\varphi^{n+1} \oplus \psi^{n+1})\mathrm{d}\mathrm{R}_{0,T}\,. \end{aligned} \tag{2.2.19}$$

In the sequel, we will refer to the couplings $(\pi^{n,n}, \pi^{n+1,n})_{n\in\mathbb{N}^*}$ as to Sinkhorn's plans. It has been pointed out in [BCC+15] (see also [Nut21, Section 6]) that

Sinkhorn's algorithm is a special case of the Bregman's iterated projection algorithm for the relative entropy functional. Indeed, the sequence of Sinkhorn's plans $\{\pi^{n,n}, \pi^{n+1,n}\}_{n\in\mathbb{N}^*}$ satisfies the following recursion:

$$\begin{cases} \pi^{n+1,n} := \arg\min_{\Pi(\mu,\star)} \mathscr{H}(\cdot|\pi^{n,n}) \,, \\ \pi^{n+1,n+1} := \arg\min_{\Pi(\star,\nu)} \mathscr{H}(\cdot|\pi^{n+1,n}) \,, \end{cases} \tag{2.2.20}$$

where $\Pi(\mu,\star)$ (resp. $\Pi(\star,\nu)$) is the set of probability measures π on $M\times M$ such that the first marginal is μ, *i.e.*, $(\mathrm{proj}_x)_\sharp\pi=\mu$ (resp. the second marginal is ν, *i.e.*, $(\mathrm{proj}_y)_\sharp\pi=\nu$). The primal formulation of Sinkhorn's algorithm (2.2.20) justifies the name *Iterative Fitting Procedure*, indeed each iterate is chosen by fitting one marginal constraint in the best possible way *i.e.*, by considering the entropic projection on the subset $\mathcal{P}(M\times M)$ that fits one of the two marginals (cf. Figure 2.1).

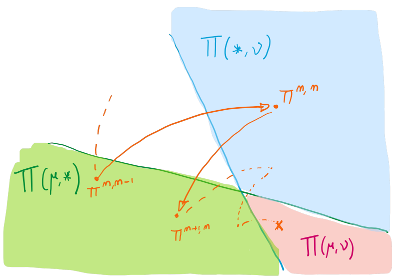

Figure 2.1: More precisely, the sequence $\{\pi^{n,n}\}_{n\in\mathbb{N}^*} \subseteq \Pi(\star,\nu)$ is a sequence that always fits the second marginal constraint, whereas the sequence $\{\pi^{n+1,n}\}_{n\in\mathbb{N}^*} \subseteq \Pi(\mu,\star)$ always fits the first marginal. Clearly, when the algorithm converges (problem which will be addressed in Chapters 5 and 6) the limit point of both sequences will be a coupling between the two marginals, *i.e.*, an element of $\Pi(\mu,\nu)$.

Clearly at each step, as soon as we fit one marginal constraint, we violate the other one. For this reason we define the adjusted marginals produced along Sinkhorn's algorithm as the probability measures

$$\mu^n := (\mathrm{proj}_x)_\sharp\pi^{n,n} \quad\text{and}\quad \nu^n := (\mathrm{proj}_y)_\sharp\pi^{n+1,n} \,. \tag{2.2.21}$$

Lastly, let us remark that Sinkhorn's iterates may be considered as potentials of appropriate Schrödinger problems. Indeed, the decomposition given in (2.2.19) implies that

- the couple (φ^{n+1},ψ^n) corresponds to a couple of Schrödinger potentials (as defined in Theorem 2.2.1) associated to the Schrödinger problem with reference measure $\mathrm{R}_{0,T}$ and with marginals μ and $\nu^n := (\mathrm{proj}_y)_\sharp\pi^{n+1,n}$;

- the couple $(\varphi^{n+1},\psi^{n+1})$ corresponds to a couple of Schrödinger potentials (as defined in Theorem 2.2.1) associated to the Schrödinger problem with reference measure $\mathrm{R}_{0,T}$ and with marginals $\mu^{n+1} := (\mathrm{proj}_x)_\sharp\pi^{n+1,n+1}$ and ν.

2.3 Dynamical formulation and optimal control

Under the additional assumption of bounded densities and supports, SP admits a dynamical formulation. What follows is taken from [GT19], however the reader should pay attention to the different time-rescaling we have adopted here in contrast to the choice made there, which eventually leads to different numerical constants.

Lemma 2.3.1 (Proposition 4.1 in [GT19]). *Let* $(M, \mathsf{d}, \mathfrak{m})$ *satisfy* (CD) *and suppose the marginals* μ, ν *satisfy **A1** and that their densities (w.r.t.* $\mathfrak{m}$*) are bounded and compactly supported. Then it holds*

$$\begin{aligned}\mathcal{C}_T(\mu,\nu) &= \mathscr{H}(\mu|\mathfrak{m}) + \int_0^T \int |\nabla \log P_{T-t}g|^2 \, \mathrm{d}\mathbf{P}_t^T \, \mathrm{d}t \\ &= \mathscr{H}(\nu|\mathfrak{m}) + \int_0^T \int |\nabla \log P_t f|^2 \, \mathrm{d}\mathbf{P}_t^T \, \mathrm{d}t \,,\end{aligned} \tag{2.3.1}$$

where f, g *are the measurable functions defined at* (2.2.1) *whereas the probability measure* $\mathbf{P}_t^T = P_t f \, P_{T-t} g \, \mathfrak{m}$ *is the T-entropic interpolation from* μ *to* ν *at time* t*, i.e.,* $\mathbf{P}_t^T = (X_t)_\# \mathbf{P}^T$.

In [GT19, Proposition 4.1], the dynamical representation formula (2.3.1) for the entropic cost is actually proven under a $\mathsf{CD}(\kappa, N)$ assumption with $N < \infty$. However, the very same argument works also if $(M, \mathsf{d}, \mathfrak{m})$ satisfies $\mathsf{CD}(\kappa, \infty)$ and $\mathfrak{m}(M) = 1$. Indeed, the proof of [GT19, Proposition 4.1] essentially relies on the regularity and integrability of $t \mapsto \rho_t := P_t f \, P_{T-t} g$, $t \mapsto \log P_t f$, $t \mapsto \log P_{T-t} g$ (and of their gradients) and these properties can be extended to the case $N = \infty$ as follows. As concerns the regularity, given that μ, ν have bounded densities and supports, the lower bound (2.2.5) on the heat kernel allows to deduce that $f, g \in \mathrm{L}^\infty(\mathfrak{m})$ with bounded supports too, exactly as in Lemma 2.2.4. Then the smoothing property of P_T entails $\mathcal{C}^\infty$-regularity, see [Gri09, Theorem 3.1]. As regards the integrability, what is needed is the existence of an $\mathrm{L}^1(\mathrm{d}t \otimes \mathfrak{m})$-function dominating, locally in t, $t \mapsto \log(P_t f)\rho_t$, $t \mapsto \log(P_{T-t} g)\rho_t$ and $t \mapsto |\nabla \log P_t f|^2 \rho_t$, $t \mapsto |\nabla \log P_{T-t} g|^2 \rho_t$. By the maximum principle, $t \mapsto \log(P_t f)\rho_t$, $t \mapsto \log(P_{T-t} g)\rho_t$ are dominated (up to multiplicative constants) by $t \mapsto P_{T-t} g$ and $t \mapsto P_t f$, respectively. As for $t \mapsto |\nabla \log P_t f|^2$, $t \mapsto |\nabla \log P_{T-t} g|^2$, the desired domination follows from Hamilton's gradient estimate (2.1.11) and the bounds on $t \mapsto \log(P_t f)\rho_t$, $t \mapsto \log(P_{T-t} g)\rho_t$.

From the dynamical formulation of Lemma 2.3.1 it is immediate the link between the fg-decomposition (or equivalently the Schrödinger potentials φ, ψ) and the stochastic optimal control formulation portrayed in Section 1.3. Indeed, at least in the Euclidean setting, the first identity in (2.3.1) can be seen as (1.3.3)[3] where the control considered is equal to the feedback control

$$\boldsymbol{u}_t^T = \nabla \log P_{T-t} e^{-\psi} (X_t^{\boldsymbol{u}^T}) \,. \tag{2.3.2}$$

[3] Let us remark here that the missing scaling factor $1/2$ in front of the integral in (1.3.3) is due to the presence of the $\sqrt{2}$ factor in front of the Brownian motion in (2.1.1) considered in this chapter.

This suggests we should study the function $\mathcal{U}_t^{T,\psi}(x) := -\log P_{T-t}e^{-\psi}(x)$ and its gradient. We will actually accomplish this in Chapters 5 and 6 in order to prove the exponential convergence of Sinkhorn's algorithm. Let us just point out here that, under some additional regularity assumption on ψ, the function $(\mathcal{U}_t^{T,\psi})_{t\in[0,T]}$ solves the Hamilton-Jacobi-Bellman equation (HJB)

$$\begin{cases} \partial_t u_t + \Delta u_t - \nabla U \cdot \nabla u_t - |\nabla u_t|^2 = 0 \\ u_T = \psi \, . \end{cases} \tag{2.3.3}$$

Due to the role that $\nabla\mathcal{U}_t^{T,\psi}$ and $\nabla\mathcal{U}_t^{T,\varphi}$ play in Lemma 2.3.1 we call these two object forward and respectively backward corrector. The study of the behaviour of their L^2-norms will be the starting point of our discussion in Chapter 3.

Finally, the above link with the stochastic optimal control formulation can indeed be shown to hold, under some geometric regularity assumptions. More precisely, for the classical SP (*i.e.*, the one considered in Chapter 1 with Brownian motions) it holds

Lemma 2.3.2 (Lemma 4.2 in [Con24])**.** *Assume* ***A1*** *and* ***A2****. Then the Schrödinger bridge* $\mathbf{P}^T \in \mathcal{P}(\Omega)$ *coincides with the law of the process*

$$\begin{cases} \mathrm{d}X_t^T = -\nabla\mathcal{U}_t^{T,\psi}(X_t^T)\mathrm{d}t + \mathrm{d}B_t \\ X_0^T \sim \mu \, , \end{cases} \tag{2.3.4}$$

and hence the Schrödinger plan is equal to $\boldsymbol{\pi}^T = \mathcal{L}(X_0^T, X_T^T)$.

This page was intentionally left blank.

Appendix 2

2.A Equivalence with Entropic Optimal Transport problems

Given any cost function $c(\cdot,\cdot) \in \mathrm{L}^1(\mu) \times \mathrm{L}^1(\nu)$ one can always consider the OT problem

$$\inf_{\pi\in\Pi(\mu,\nu)} \int c(x,y)\,\mathrm{d}\pi$$

and its entropic regularisation

$$\mathrm{EOT}_\varepsilon(\mu,\nu) := \inf_{\pi\in\Pi(\mu,\nu)} \int c(x,y)\,\mathrm{d}\pi + \varepsilon\,\mathscr{H}(\pi|\mu\otimes\nu)\,, \tag{2.A.1}$$

for some regularising positive parameter $\varepsilon > 0$. $\mathrm{EOT}_\varepsilon(\mu,\nu)$ is referred to as the *entropic cost* and, when not clear from the context, we will stress its dependence from the cost function by denoting it with $\mathrm{EOT}^c_\varepsilon(\mu,\nu)$.

The above problem shares many common properties with SP. The most remarkable one is that a result similar to our Theorem 2.2.1 holds true also in this case, *i.e.*, it admits a unique optimiser $\boldsymbol{\pi}^\varepsilon \in \Pi(\mu,\nu)$ and its density is of the form

$$\frac{\mathrm{d}\boldsymbol{\pi}^\varepsilon}{\mathrm{d}(\mu\otimes\nu)} = \exp\left(\frac{\varphi_\varepsilon \oplus \psi_\varepsilon - c}{\varepsilon}\right) \qquad \mu\otimes\nu\text{-a.s.} \tag{2.A.2}$$

for two measurable functions $\varphi_\varepsilon \in L^1(\mu)$ and $\psi_\varepsilon \in L^1(\nu)$, to whom we refer as to the (entropic) potentials, and they are unique up to additive shift, *i.e.*, $(\varphi_\varepsilon,\psi_\varepsilon) \mapsto (\varphi_\varepsilon + a, \psi_\varepsilon - a)$ for any $a \in \mathbb{R}$. Henceforth from now on we tacitly assume the validity of the symmetric normalisation (as in (2.2.13)) which in the EOT setting reads as

$$\int \varphi_\varepsilon\,\mathrm{d}\mu = \int \psi_\varepsilon\,\mathrm{d}\nu\,. \tag{2.A.3}$$

Similarly to what we have already explained in Section 1.2, the general SP problem considered in (2.1.3) can also be considered as a generalised EOT problem. However, in general the dependence from $T > 0$ is not explicitly given since p_T is only implicitly given. Therefore if in (2.A.1) we consider $\varepsilon = T$ and allow for an ε-dependent cost, namely $c = -T \log \frac{\mathrm{dR}_{0,T}}{\mathrm{d}(\mu\otimes\nu)}$, than indeed we have $\mathrm{EOT}_\varepsilon(\mu,\nu) = T\,\mathcal{C}_T(\mu,\nu)$ and the two problems share the same optimiser $\boldsymbol{\pi}^\varepsilon \equiv \boldsymbol{\pi}^T$. Lastly on the dual side we would have the identity $T(\varphi^T \oplus \psi^T) = -\varphi_\varepsilon \oplus \psi_\varepsilon$ between the Schrödinger potentials (φ^T, ψ^T), as defined in Theorem 2.2.1, and the entropic potentials, as defined in (2.A.2). This connection allows to translate results from EOT theory into results in SP theory and viceversa. For an extensive introduction to EOT problems and their relation with SP, we refer the reader to the lecture notes [Nut21].

2.B A technical lemma

The proof of the next lemma is taken from [Tam17, Proposition 4.1.5].

Lemma 2.B.1. *Consider the assumption of Theorem 2.2.1, and let V and ${}^\perp W$ be the function spaces defined at* (2.2.10). *Then we have ${}^\perp V_0 \subseteq V_+$.*

Proof. Let us firstly recall the definition of the function spaces

$$\begin{aligned}
V_+ &:= \Big(\mathrm{L}^0(M, \mathfrak{m}_{|A_\mu}) \oplus \mathrm{L}^0(M, \mathfrak{m}_{|A_\nu})\Big) \cap \mathrm{L}^1(M^2, \boldsymbol{\pi}^T)\,,\\
V_0 &:= \{\ell \in \mathrm{L}^\infty(M^2, \boldsymbol{\pi}^T) \colon (\mathrm{proj}_{x_1})_\#(\ell\boldsymbol{\pi}^T) = (\mathrm{proj}_{x_2})_\#(\ell\boldsymbol{\pi}^T) = 0\}\,,\\
V_+^\perp &:= \{\ell \in \mathrm{L}^\infty(M^2, \boldsymbol{\pi}^T) \colon \int u\ell \,\mathrm{d}\boldsymbol{\pi}^T = 0 \quad \forall\, u \in V_+\}\,,\\
{}^\perp V_0 &:= \{u \in \mathrm{L}^1(M^2, \boldsymbol{\pi}^T) \colon \int u\ell \,\mathrm{d}\boldsymbol{\pi}^T = 0 \quad \forall\, \ell \in V_0\}\,.
\end{aligned}$$

For sake of clarity we are going to proceed by steps.

Step 1. We are going to show that V_+ is a closed subset of $\mathrm{L}^1(M^2, \boldsymbol{\pi}^T)$.

Firstly, let us argue that $u \in V_+$ if and only if $u \in \mathrm{L}^1(M^2, \boldsymbol{\pi}^T)$ and for $\mathfrak{m}^{\otimes 4}$−a.e. $(x, x', y, y') \in A_\mu^2 \times A_\nu^2$ it holds

$$u(x,y) + u(x',y') = u(x,y') + u(x',y)\,. \tag{2.B.1}$$

Indeed the *only if* part is trivial; whereas if we assume the above holds true $\mathfrak{m}^{\otimes 4}$-a.e. on $A_\mu^2 \times A_\nu^2$, then Fubini's Theorem guarantees that there exist $(x',y') \in A_\mu \times A_\nu$ such that $\mathfrak{m}^{\otimes 2}$-a.e. on $A_\mu \times A_\nu$ it holds $u = u(\cdot, y') \oplus (u(x',\cdot) - u(x',y')) \in V_+$.

Now, since $\mathfrak{m}_{|A_\mu} \otimes \mathfrak{m}_{|A_\nu} \ll \boldsymbol{\pi}^T$ (cf. (2.2.11)) we deduce that (2.B.1) is a closed condition also in $\mathrm{L}^1(M^2, \boldsymbol{\pi}^T)$, which proves our claim.

Step 2. Next we show that $V_+^\perp \subseteq V_0$ or equivalently that $\mathrm{L}^\infty(M^2, \boldsymbol{\pi}^T) \setminus V_0 \subseteq \mathrm{L}^\infty(M^2, \boldsymbol{\pi}^T) \setminus V_+^\perp$. Hence, let $\ell \in \mathrm{L}^\infty(M^2, \boldsymbol{\pi}^T) \setminus V_0$, *i.e.*, without loss of generalities, we may assume that the first marginal measure $(\mathrm{proj}_{x_1})_\#(\ell\boldsymbol{\pi}^T)$ is non-zero.

If we consider $u := \mathrm{d}(\mathrm{proj}_{x_1})_\#(\ell \boldsymbol{\pi}^T)/\mathrm{d}\mu \oplus 0 \in V_+$, it holds

$$\int u\ell \,\mathrm{d}\boldsymbol{\pi}^T = \int \frac{\mathrm{d}(\mathrm{proj}_{x_1})_\#(\ell \boldsymbol{\pi}^T)}{\mathrm{d}\mu}\,\mathrm{d}(\ell \boldsymbol{\pi}^T) = \int \frac{\mathrm{d}(\mathrm{proj}_{x_1})_\#(\ell \boldsymbol{\pi}^T)}{\mathrm{d}\mu} \circ \mathrm{proj}_{x_1}\,\mathrm{d}(\ell \boldsymbol{\pi}^T)$$
$$= \int \frac{\mathrm{d}(\mathrm{proj}_{x_1})_\#(\ell \boldsymbol{\pi}^T)}{\mathrm{d}\mu}\,\mathrm{d}(\mathrm{proj}_{x_1})_\#(\ell \boldsymbol{\pi}^T) = \int \left(\frac{\mathrm{d}(\mathrm{proj}_{x_1})_\#(\ell \boldsymbol{\pi}^T)}{\mathrm{d}\mu}\right)^2 \mathrm{d}\mu > 0\,,$$

which implies $\ell \notin V_+^\perp$.

Conclusion. Take $u \in \mathrm{L}^1(M^2, \boldsymbol{\pi}^T) \setminus V_+$. Then Hahn-Banach Theorem [Bre10, Theorem 1.7] guarantees the existence of some $\ell \in \mathrm{L}^1(M^2, \boldsymbol{\pi}^T)^* = \mathrm{L}^\infty(M^2, \boldsymbol{\pi}^T)$ such that for any $\tilde{u} \in V_+$ it holds $\int \tilde{u}\ell \,\mathrm{d}\boldsymbol{\pi}^T = 0$ whereas $\int u\ell \,\mathrm{d}\boldsymbol{\pi}^T \neq 0$ (see also the proof of [Bre10, Corollary 1.8]). This particularly implies $\ell \in V_+^\perp \subseteq V_0$. Since $\int u\ell \,\mathrm{d}\boldsymbol{\pi}^T \neq 0$ we can finally deduce that $u \notin {}^\perp V_0$ and therefore that $\mathrm{L}^1(M^2, \boldsymbol{\pi}^T) \setminus V_+ \subseteq \mathrm{L}^1(M^2, \boldsymbol{\pi}^T) \setminus {}^\perp V_0$ or equivalently that ${}^\perp V_0 \subseteq V_+$. $\square$

This page was intentionally left blank.

Chapter 3

Convergence to the Brenier map

In this chapter we further develop the discussion initiated in Section 1.2 where we have linked SP with optimal transport theory. Firstly, inspired by the link with stochastic optimal control, portrayed in Section 1.3, we prove in Section 3.1 some gradient estimates to which we refer to as *corrector estimates*. Then in Section 3.2 we analyse the small-time limit of SP (towards OT) and prove our main result Theorem 3.2.3, *i.e.*, the convergence of the gradients of Schrödinger potentials towards the Brenier map. Lastly, in Section 3.2.3 we provide some quantitative convergence rates estimates in the Euclidean setting under the additional geometric assumption **A**2.

3.1 Corrector estimates

In this section we will show how the (CD) condition implies useful gradient contraction estimates for the Schrödinger potentials. We will refer to them as corrector estimates, in view of the link with stochastic optimal control problems described in Section 2.3.

In order to do that we are going to consider, apart from the curvature-dimension condition (CD) and from **A**1, that

A3. *For* $\mathfrak{p} = \mu, \nu \in \mathcal{P}(M)$ *it holds either*

- $\dfrac{\mathrm{d}\mathfrak{p}}{\mathrm{d}\mathfrak{m}} \in \mathrm{L}^\infty(\mathfrak{m})$ *and is compactly supported, or*
- $\dfrac{\mathrm{d}\mathfrak{p}}{\mathrm{d}\mathfrak{m}}$ *is locally bounded away from zero on* $\mathrm{int}(\mathrm{supp}(\mathfrak{p}))$ *and* $\mathfrak{p}(\partial \mathrm{supp}(\mathfrak{p})) = 0$*;*

Our proof strategy reads as follows. Firstly, we will assume that the marginals have compact support and bounded densities, then we will extend it to the case in which $f, g \in \mathrm{L}^\infty(\mathfrak{m})$, and finally to the case when we assume that the second condition in **A**3 is satisfied by relying on a finite-dimensional approximation argument where the second marginal is fixed while the first marginal constraint is replaced by a finite-dimensional one, similarly to the discussion already given for (2.2.9). More precisely, since (M, d) is separable, we know that there exists a countable dense family of bounded measurable functions $\{\phi_i\}_{i\in\mathbb{N}}$ such that

$$(\mathrm{proj}_{x_1})_\# \pi = \mu \quad \Leftrightarrow \quad \int \phi_i(x) \mathrm{d}\pi(x,y) = \int \phi_i \mathrm{d}\mu \qquad \forall\, i \in \mathbb{N}\,.$$

Therefore for any fixed $K \in \mathbb{N}$ we may define the convex and closed (in total variation) set

$$\mathcal{Q}_K^\nu := \left\{ \pi \in \mathcal{P}(M^2) : (\mathrm{proj}_{x_2})_\# \pi = \nu\,, \int \phi_i(x) \mathrm{d}\pi(x,y) = \int \phi_i \mathrm{d}\mu \text{ for all } i \leq K \right\}$$

and then consider the associated minimisation problem

$$\inf_{\pi \in \mathcal{Q}_K^\nu} \mathscr{H}(\pi | \mathrm{R}_{0,T})\,. \tag{3.1.1}$$

With the next lemma we show that the above problem is well posed and gives an approximation of SP.

Lemma 3.1.1. *Let* $(M, \mathsf{d}, \mathfrak{m})$ *satisfy* (CD) *and suppose that the marginals* μ, ν *satisfy* **A**1. *Then the above minimisation problem* (3.1.1) *admits a unique optimiser* $\pi^K \in \mathcal{Q}_K^\nu$, *whose density is given by*

$$\frac{\mathrm{d}\pi^K}{\mathrm{dR}_{0,T}}(x,y) = f_K(x)\, g_K(y) \qquad \text{with } f_K = C \exp\left(\sum_{i=1}^K \lambda_i\, \phi_i \right) \in \mathrm{L}^\infty(\mathfrak{m}) \tag{3.1.2}$$

for some constant $C > 0$ *and multipliers* $\lambda_i \in \mathbb{R}$. *Moreover,* g_K *converges* $\mathfrak{m}$-*a.e. to* g *and* $P_T f_K$ *converges* $\mathfrak{m}$-*a.e. to* $P_T f$ *on* $\mathrm{supp}(\nu)$. *Finally, the optimal value* $\mathscr{H}(\pi^K | \mathrm{R}_{0,T})$ *converges to* $\mathcal{C}_T(\mu, \nu)$ *as* $K \uparrow +\infty$.

Proof. The existence and uniqueness of the minimizer $\pi^K \in \mathcal{Q}_K^\nu$ is guaranteed by [Nut21, Proposition 1.17][1] Moreover, from the same reference we have $\boldsymbol{\pi}^T \ll \pi^K$ and

$$\pi^K \to \boldsymbol{\pi}^T \text{ in total variation} \quad \text{and} \quad \mathscr{H}(\pi^K | \mathrm{R}_{0,T}) \to \mathscr{H}(\boldsymbol{\pi}^T | \mathrm{R}_{0,T}) = \mathcal{C}_T(\mu, \nu)\,. \tag{3.1.3}$$

Let us now prove (3.1.2). Firstly, notice that if we introduce the marginal $\mu^K := (\mathrm{proj}_{x_1})_\# \pi^K$, then it immediately follows that $\pi^K \in \Pi(\mu^K, \nu) \subseteq \mathcal{Q}_K^\nu$ is the unique optimiser of the Schrödinger problem with marginals μ^K, ν and clearly

[1]Let us just point out here that this result applies to settings where the reference $\mathrm{R}_{0,T}$ is a probability measure. When this is not the case, one can consider the probability measure $\mathrm{R}_{0,T}^\nu(\mathrm{d}x, \mathrm{d}y) := \mathsf{p}_T(x,y)\mathfrak{m}(\mathrm{d}x)\nu(\mathrm{d}y)$ and argue as we already did in the proof of Theorem 2.2.1.

$\mathcal{C}_T(\mu^K,\nu) = \mathscr{H}(\pi^K|\mathrm{R}_{0,T})$. Owing to (3.1.3), $\mathscr{H}(\mu^K|\mathfrak{m}) \leq \mathscr{H}(\pi^K|\mathrm{R}_{0,T})$ is eventually finite for K large enough and therefore Theorem 2.2.1 yields the existence of two positive measurable functions f_K, g_K such that $\mathrm{R}_{0,T}$-a.e. it holds $\frac{\mathrm{d}\pi^K}{\mathrm{d}\mathrm{R}_{0,T}}(x,y) = f_K(x)\,g_K(y)$ with $\log f_K \in \mathrm{L}^1(\mu^K)$ and $\log g_K \in \mathrm{L}^1(\nu)$. Then, in view of the additive property of the relative entropy (1.A.4) we may write for all $\pi \in \mathcal{Q}^\nu_K$

$$\mathscr{H}(\pi|\mathrm{R}_{0,T}) = \mathscr{H}(\nu|\mathfrak{m}) + \int \mathscr{H}\left(\pi(\cdot|\mathrm{proj}_{x_2} = y)|\mathfrak{m}P_T(y)\right)\,\mathrm{d}\nu(y)\,,$$

where $\mathfrak{m}P_T(y) = \mathrm{R}_{0,T}(\cdot|\mathrm{proj}_{x_2} = y) \in \mathcal{P}(M)$ is the probability measure whose density is given by $\mathsf{p}_T(x,y)\mathrm{d}\mathfrak{m}(x)$. Henceforth, if $y \mapsto \pi^K(\mathrm{d}x|y) \in \mathcal{P}(M)$ denotes the stochastic kernel associated to $\pi^K = \pi^K(\mathrm{d}x|y) \otimes \nu(\mathrm{d}y)$ when conditioning on the second variable, for any fixed $y \in \mathrm{supp}(\nu)$ the probability measure $\pi^K(\mathrm{d}x|y)$ clearly minimises

$$\inf_{\mathfrak{q}\in\mathcal{Q}_K(y)} \mathscr{H}(\mathfrak{q}|\mathfrak{m}P_T(y))$$
$$\text{with } \mathcal{Q}_K(y) := \left\{\mathfrak{q} \in \mathcal{P}(M):\ \int \phi_i \mathrm{d}\mathfrak{q} = \int \phi_i \mathrm{d}\pi^K(\cdot|y) \text{ for all } i \leq K\right\}.$$

The existence and uniqueness of solution in the above problem, for any fixed $y \in \mathrm{supp}(\nu)$, are once again ensured by [Nut21, Proposition 1.17]. Moreover, by arguing as in [Nut21, Example 1.18] for any fixed $y \in \mathrm{supp}(\nu)$ we can write

$$\frac{\mathrm{d}\pi^K(\cdot|y)}{\mathrm{d}(\mathfrak{m}P_T(y))}(x) = c(y)\exp\left(\sum_{i=1}^K b_i(y)\phi_i(x)\right)$$

for some constants $c(y) > 0$ and $b_i(y) \in \mathbb{R}$, possibly depending on y. By combining the above expression, the f_K, g_K-decomposition and the Schrödinger system associated to $\mathcal{C}_T(\mu^K,\nu)$ (cf. (2.2.2) in Theorem 2.2.1) we deduce that

$$\begin{aligned} f_K(x)g_K(y) = \frac{\mathrm{d}\pi^K}{\mathrm{d}\mathrm{R}_{0,T}}(x,y) &= \frac{\mathrm{d}(\pi^K(\cdot|y)\otimes\nu)}{\mathrm{d}(\mathfrak{m}P_T\otimes\mathfrak{m})}(x,y) = \frac{\mathrm{d}\nu}{\mathrm{d}\mathfrak{m}}(y)\,\frac{\mathrm{d}\mathfrak{q}^*(y)}{\mathrm{d}(\mathfrak{m}P_T(y))}(x) \\ &= g_K(y)P_Tf_K(y)\,c(y)\exp\left(\sum_{i=1}^K b_i(y)\phi_i(x)\right). \end{aligned}$$

Henceforth, for any $y \in \mathrm{supp}(\nu)$ it holds

$$f_K(x) = P_Tf_K(y)\,c(y)\exp\left(\sum_{i=1}^K b_i(y)\phi_i(x)\right)$$

and since the above left-hand side does not depend on the choice of y, we may choose a fixed $y^* \in \mathrm{supp}(\nu)$ and write

$$f_K(x) = P_Tf_K(y^*)\,c(y^*)\exp\left(\sum_{i=1}^K b_i(y^*)\phi_i(x)\right) \in (0,+\infty)\,.$$

This proves (3.1.2) with $C := P_T f_K(y^*)\, c(y^*)$ and $\lambda_i := b_i(y^*)$.

Finally, we claim that we can fix a renormalisation for the decomposition f_K, g_K (which is unique up to a multiplicative constant) such that g_K converges $\mathfrak{m}$-a.e. to g. In order to do so, note that the convergence in total variation in (3.1.3) implies that (along a non-relabelled subsequence)

$$f_K(x)g_K(y) = \frac{\mathrm{d}\pi^K}{\mathrm{dR}_{0,T}}(x,y) \to \frac{\mathrm{d}\boldsymbol{\pi}^T}{\mathrm{dR}_{0,T}}(x,y) = f(x)g(y), \qquad \mathrm{R}_{0,T}\text{-a.e.}$$

and a fortiori $\mathfrak{m} \otimes \mathfrak{m}$-a.e., thanks to the (CD) assumption and the Gaussian lower bounds (2.2.5) and (2.2.7). If $A \subseteq M \times M$ denotes the subset where the latter limit holds pointwise and A^c its complement, then $\mathfrak{m} \otimes \mathfrak{m}(A^c) = 0$ and Fubini's Theorem implies that for $\mathfrak{m}$-a.e. $x \in M$ it holds $\mathfrak{m}(A^c_x) = 0$, where the section is defined as $A_x := \{y\colon (x,y) \in A\}$. Combining this with $\mu \ll \mathfrak{m}$ we deduce that there exists an element $x^* \in \mathrm{int}(\mathrm{supp}(\mu))$ such that $f(x^*) \neq 0$ and $\mathfrak{m}(A^c_{x^*}) = 0$. We have therefore proven that

$$f_K(x^*)g_K(y) \to f(x^*)g(y) \quad \text{for } \mathfrak{m}\text{-a.e. } y. \tag{3.1.4}$$

By renormalising the f_K, g_K-decomposition such that $f_K(x^*) = f(x^*) \in (0, +\infty)$, (3.1.4) reads as

$$g_K \to g \quad \mathfrak{m}\text{-a.e.}$$

As a direct consequence of this and (2.2.2) we get the $\mathfrak{m}$-a.e. convergence of $P_T f_K$ to $P_T f$ on $\mathrm{supp}(\nu)$, since the marginal ν is always the same at each step $K \in \mathbb{N}$. □

Thanks to the previous result, we are finally able to prove contraction gradient estimates for the Scrhödinger potentials.

Proposition 3.1.2 (Corrector estimates)**.** *Let* $(M, \mathsf{d}, \mathfrak{m})$ *satisfy* (CD)*, suppose the marginals* μ*,* ν *satisfy* ***A*1** *and let* f*,* g *be as in Theorem 2.2.1 and satisfying the normalisation* (2.2.13)*. Then, if* $\frac{\mathrm{d}\nu}{\mathrm{d}\mathfrak{m}}$ *is locally bounded away from* 0 *on* $\mathrm{int}(\mathrm{supp}(\nu))$ *and* $\nu(\partial\mathrm{supp}(\nu)) = 0$ *it holds*

$$\|\nabla \log P_T f\|^2_{\mathrm{L}^2(\nu)} \leq \frac{1}{E_{2\kappa}(T)}\left[\mathcal{C}_T(\mu,\nu) - \mathscr{H}(\nu|\mathfrak{m})\right], \tag{3.1.5}$$

where $E_{2\kappa}$ *is defined as in* (2.1.10)*. In particular, it is part of the statement the fact that* $\log P_T f \in W^{1,2}_{loc}(\mathrm{int}(\mathrm{supp}(\nu)))$ *with* $|\nabla \log P_T f| \in \mathrm{L}^2(\nu)$*.*

Similarly, if $\frac{\mathrm{d}\mu}{\mathrm{d}\mathfrak{m}}$ *is locally bounded away from* 0 *on* $\mathrm{int}(\mathrm{supp}(\mu))$ *and if it holds* $\mu(\partial\mathrm{supp}(\mu)) = 0$*, then* $\log P_T g \in W^{1,2}_{loc}(\mathrm{int}(\mathrm{supp}(\mu)))$ *with* $|\nabla \log P_T g| \in \mathrm{L}^2(\mu)$ *and it holds*

$$\|\nabla \log P_T g\|^2_{\mathrm{L}^2(\mu)} \leq \frac{1}{E_{2\kappa}(T)}\left[\mathcal{C}_T(\mu,\nu) - \mathscr{H}(\mu|\mathfrak{m})\right]. \tag{3.1.6}$$

In particular, both (3.1.5) *and* (3.1.6) *hold true if the marginals satisfy* ***A*3***.*

Proof. We will only prove inequality (3.1.5), as (3.1.6) follows by completely analogous techniques.

As a preliminary step, let us assume that μ, ν have bounded supports and bounded densities $\frac{\mathrm{d}\mu}{\mathrm{d}\mathfrak{m}}, \frac{\mathrm{d}\nu}{\mathrm{d}\mathfrak{m}}$. Under this extra hypothesis, it holds

$$\int |\nabla \log P_T f|^2 \,\mathrm{d}\nu \le e^{-2\kappa(T-t)} \int |\nabla \log P_t f|^2 \,\mathrm{d}\mathbf{P}_t^T, \tag{3.1.7}$$

where we recall that $\mathbf{P}_t^T = P_t f\, P_{T-t} g\, \mathfrak{m}$ is the T-entropic interpolation from μ to ν at time t. If $(M, \mathsf{d}, \mathfrak{m})$ satisfies $\mathsf{CD}(\kappa, N)$ with $N < \infty$, this is a consequence of Grönwall lemma applied to the functions $\alpha(t) := \int |\nabla \log P_t f|^2 \,\mathrm{d}\mathbf{P}_t^T$ and $\beta(t) := e^{2\kappa(T-t)} \int |\nabla \log P_T f|^2 \,\mathrm{d}\nu$, as on the one hand by [Con19, Lemmas 3.6 and 3.7] together with [GT21, Proposition 4.8] (which justifies the computations of [Con19, Lemmas 3.6 and 3.7] in the non-compact and possibly negatively curved setting) we have

$$\alpha \in C^1((0,T]) \qquad \text{and} \qquad \alpha'(t) \le -2\kappa\alpha(t),\ \forall t \in (0,T] \tag{3.1.8}$$

while on the other hand it is readily verified that $\beta' = -2\kappa\beta$; since $\alpha(T) = \beta(T)$, it must hold $\alpha(t) \ge \beta(t)$ for all $t \in (0,T]$, namely (3.1.7).

If instead $(M, \mathsf{d}, \mathfrak{m})$ satisfies $\mathsf{CD}(\kappa, \infty)$ and $\mathfrak{m}(M) = 1$, the reader should refer to [CT21, Lemma 2.2]: this grants the validity of (3.1.8) for

$$\alpha_\delta(t) := c_\delta \int |\nabla \log(P_t f + \delta)|^2 \,(P_t f + \delta)(P_{T-t} g + \delta) \,\mathrm{d}\mathfrak{m},$$

where $\delta > 0$ is any positive number and c_δ is a normalisation constant, so that $c_\delta(P_t f + \delta)(P_{T-t} g + \delta)\mathfrak{m}$ is a probability measure. By considering

$$\beta_\delta(t) := e^{2\kappa(T-t)} \int |\nabla \log(P_T f + \delta)|^2 \,(P_T f + \delta) g \,\mathrm{d}\mathfrak{m},$$

by the same argument as before we obtain $\alpha_\delta(t) \ge \beta_\delta(t)$ for all $t \in (0,T]$ and $\delta > 0$, whence (3.1.7) by passing to the limit as $\delta \downarrow 0$ by the Dominated Convergence Theorem. Indeed, note first that

$$\begin{aligned} |\nabla \log(P_t f + \delta)|^2 (P_t f + \delta)(P_{T-t} g + \delta) &= \frac{|\nabla P_t f|^2}{P_t f + \delta} (P_{T-t} g + \delta) \\ &\le |\nabla \log P_t f|^2 P_t f (P_{T-t} g + \delta) \le C(\kappa, t)(P_t f)|\log P_t f|(P_{T-t} g + 1) \end{aligned}$$

where the last inequality is due to Hamilton's gradient estimate (2.1.11), $C(\kappa, t)$ being some constant that only depends on κ, t. Then, since from Lemma 2.2.4 we know that $f, g \in \mathrm{L}^\infty(\mathfrak{m})$, by the maximum principle we deduce that the previous right-hand side is bounded, hence integrable as $\mathfrak{m}(M) = 1$, and thus provides an admissible dominating function.

Now, by multiplying by $e^{2\kappa(T-t)}$ and integrating over $t \in [0,T]$ in (3.1.7) we deduce that

$$E_{2\kappa}(T) \int |\nabla \log P_T f|^2 \,\mathrm{d}\nu \le \int_0^T \int |\nabla \log P_t f|^2 \,\mathrm{d}\mathbf{P}_t^T \mathrm{d}t = \mathcal{C}_T(\mu, \nu) - \mathscr{H}(\nu | \mathfrak{m}),$$

where the identity is justified by the dynamical formulation of SP in Lemma 2.3.1. Inequality (3.1.5) is thus proved, as well as the last part of the statement.

Now let us remove the additional assumptions on μ, ν in a two-steps procedure.

1st step. In addition to **A**1, assume that μ, ν are such that the associated $f, g \in L^\infty(\mathfrak{m})$. Then fix $\bar{x} \in M$ and introduce $f_n := \mathbf{1}_{B_n(\bar{x})} f$, $g_n := \mathbf{1}_{B_n(\bar{x})} g$, so that (3.1.5) holds true for

$$\mu_n := c_n f_n P_T g_n \, \mathfrak{m} \qquad \text{and} \qquad \nu_n := c_n g_n P_T f_n \, \mathfrak{m},$$

where c_n is a normalisation constant (note that by self-adjointness of P_T it is the same for both measures μ_n and ν_n). Namely

$$\int |\nabla \log P_T f_n|^2 \, \mathrm{d}\nu_n \le \frac{1}{E_{2\kappa}(T)} \Big[\mathcal{C}_T(\mu_n, \nu_n) - \mathscr{H}(\nu_n | \mathfrak{m}) \Big].$$

Observing that $|\nabla \log P_T f_n| \ge |\nabla \log(P_T f_n + \delta)|$ for any $\delta > 0$, the inequality above implies in particular that

$$\int |\nabla \log(P_T f_n + \delta)|^2 \, \mathrm{d}\nu_n \le \frac{1}{E_{2\kappa}(T)} \Big[\mathcal{C}_T(\mu_n, \nu_n) - \mathscr{H}(\nu_n | \mathfrak{m}) \Big] \tag{3.1.9}$$

for all $\delta > 0$. Let us now pass to the limit as $n \to \infty$.

To this end, observe first that $P_t f_n \to P_t f$ and $P_T g_n \to P_T g$ pointwise as $n \to \infty$. Indeed

$$\begin{aligned} |P_T f_n(x) - P_T f(x)| &= \left| \int (\mathbf{1}_{B_n(\bar{x})}(y) - 1) f(y) \mathsf{p}_T(x,y) \, \mathrm{d}\mathfrak{m}(y) \right| \\ &\le \|f\|_{L^\infty} \int |\mathbf{1}_{B_n(\bar{x})}(y) - 1| \mathsf{p}_T(x,y) \, \mathrm{d}\mathfrak{m}(y) \end{aligned}$$

and the right-hand side vanishes as $n \to \infty$ by dominated convergence. This allows to handle the right-hand side of (3.1.9) in the following way. As concerns the term $\mathcal{C}_T(\mu_n, \nu_n)$, on the one hand,

$$\begin{aligned} \int \log f_n \, \mathrm{d}\mu_n \le \int \log f \, \mathrm{d}\mu_n &= \int (\log f)^+ \, \mathrm{d}\mu_n - \int (\log f)^- \, \mathrm{d}\mu_n \\ &\le c_n \int (\log f)^+ \, \mathrm{d}\mu - \int (\log f)^- \, \mathrm{d}\mu_n, \end{aligned}$$

where we have used $f_n \le f$ and, as a consequence of the maximum principle, also $P_T g_n \le P_T g$. This implies that

$$\begin{aligned} \limsup_{n\to\infty} \int \log f_n \, \mathrm{d}\mu_n &\le \int (\log f)^+ \, \mathrm{d}\mu - \liminf_{n\to\infty} \int (\log f)^- \, \mathrm{d}\mu_n \\ &\le \int (\log f)^+ \, \mathrm{d}\mu - \int (\log f)^- \, \mathrm{d}\mu = \int \log f \, \mathrm{d}\mu, \end{aligned}$$

since $c_n \to 1$ and $f_n\, P_T g_n \to f\, P_T g$ pointwise (as already discussed before) together with Fatou Lemma. On the other hand, one can show in an analogous fashion that

$$\limsup_{n\to\infty} \int \log g_n \,\mathrm{d}\nu_n \le \int \log g \,\mathrm{d}\nu.$$

Since

$$\mathcal{C}_T(\mu_n,\nu_n) = \int \log f_n \,\mathrm{d}\mu_n + \int \log g_n \,\mathrm{d}\nu_n + \log c_n,$$

we conclude that

$$\limsup_{n\to\infty} \mathcal{C}_T(\mu_n,\nu_n) \le \mathcal{C}_T(\mu,\nu). \tag{3.1.10}$$

Secondly, as regards $\mathscr{H}(\nu_n|\mathfrak{m})$, note that $g_n P_T f_n \to g P_T f$ $\mathfrak{m}$-a.e. together with $g_n P_T f_n \le g P_T f \in \mathrm{L}^1(\mathfrak{m})$ entails $g_n P_T f_n \to g P_T f$ in $\mathrm{L}^1(\mathfrak{m})$ by dominated convergence, whence $\nu_n \rightharpoonup \nu$. Moreover, as $n \to \infty$ it holds

$$\begin{aligned}
|M_2(\nu_n) - M_2(\nu)| &\le \left| \int \mathsf{d}^2(\cdot,\bar{x}) \,\mathrm{d}\nu_n - \int \mathsf{d}^2(\cdot,\bar{x}) g_n P_T f_n \,\mathrm{d}\mathfrak{m} \right| \\
&\qquad + \left| \int \mathsf{d}^2(\cdot,\bar{x}) g_n P_T f_n \,\mathrm{d}\mathfrak{m} - \int \mathsf{d}^2(\cdot,\bar{x}) \,\mathrm{d}\nu \right| \\
&\le (c_n - 1) M_2(\nu) + \int \mathsf{d}^2(\cdot,\bar{x}) \left| \frac{g_n P_T f_n}{g P_T f} - 1 \right| \mathrm{d}\nu \to 0,
\end{aligned}$$

again by $c_n \to 1$ and dominated convergence (recall indeed that $\nu \in \mathcal{P}_2(M)$ by **A**1 and $g_n P_T f_n \le g P_T f$, so that $\frac{g_n P_T f_n}{g P_T f} \in [0,1]$). Therefore $\mathbf{W}_2(\nu_n,\nu) \to 0$ and, by lower semicontinuity of the entropy w.r.t. $\mathbf{W}_2$-convergence (cf. Lemma 2.1.2),

$$\limsup_{n\to\infty} \Big(-\mathscr{H}(\nu_n|\mathfrak{m}) \Big) \le -\mathscr{H}(\nu|\mathfrak{m}). \tag{3.1.11}$$

Passing to the left-hand side of (3.1.9), fix $k \in \mathbb{N}$ and note that for all $n \ge k$ we have

$$\begin{aligned}
\int |\nabla \log(P_T f_n + \delta)|^2 \,\mathrm{d}\nu_n &\ge \frac{1}{c_k} \int |\nabla \log(P_T f_n + \delta)|^2 \,\mathrm{d}\nu_k \\
&= \frac{1}{c_k} \int_{\bar{B}_k(\bar{x})} |\nabla \log(P_T f_n + \delta)|^2 \,\mathrm{d}\nu_k,
\end{aligned}$$

since $c_n \ge 1$. We now claim that $|\nabla \log(P_T f_n + \delta)| \rightharpoonup G$ in $L^2(\bar{B}_k(\bar{x}),\nu_k)$ for some G such that $|\nabla \log(P_T f + \delta)| \le G$. To this end, observe that by the L^∞-Lipschitz regularisation (2.1.9) it holds $|\nabla P_T f_n| \le C_{T,\kappa} \|f_n\|_{\mathrm{L}^\infty(\mathfrak{m})}$, where $C_{T,\kappa}$ only depends on T and κ given by (CD). Since $f_n \le f$, this implies

$$\int_{\bar{B}_k(\bar{x})} |\nabla \log(P_T f_n + \delta)|^2 \,\mathrm{d}\mathfrak{m} \le \frac{C_{T,\kappa}^2}{\delta^2} \mathfrak{m}(B_k(\bar{x})) \|f\|^2_{\mathrm{L}^\infty(\mathfrak{m})}.$$

The functions $(|\nabla \log(P_T f_n + \delta)|)_{n\in\mathbb{N}}$ are thus equi-bounded in $\mathrm{L}^2(\bar{B}_k(\bar{x}),\mathfrak{m})$ and this implies that, up to subsequences, they converge to some function

$G \in L^2(\bar{B}_k(\bar{x}), \mathfrak{m})$ weakly in $L^2(\bar{B}_k(\bar{x}), \mathfrak{m})$. Since $\log(P_T f_n + \delta)$ converges $\mathfrak{m}$-a.e. to $\log(P_T f + \delta)$, by [AGS14a, Lemma 4.3(b)] $|\nabla \log(P_T f + \delta)| \leq G$. As for $|\nabla \log(P_T f_n + \delta)| \rightharpoonup G$ in $L^2(\bar{B}_k(\bar{x}), \nu_k)$, this directly follows from the weak convergence in $L^2(\bar{B}_k(\bar{x}), \mathfrak{m})$ and the fact that, under the current assumptions, $\frac{d\nu}{d\mathfrak{m}} \in L^\infty(\mathfrak{m})$. Combining the claim with the lower semicontinuity of the $L^2(\nu_k)$-norm w.r.t. weak convergence we obtain

$$\begin{aligned}\liminf_{n\to\infty} \int_{\bar{B}_k(\bar{x})} |\nabla \log(P_T f_n + \delta)|^2 \, d\nu_k &\geq \int_{\bar{B}_k(\bar{x})} G^2 d\nu_k \\ &\geq \int_{\bar{B}_k(\bar{x})} |\nabla \log(P_T f + \delta)|^2 \, d\nu_k.\end{aligned}$$

From this inequality, (3.1.10) and (3.1.11) we end up with

$$\frac{1}{c_k} \int_{\bar{B}_k(\bar{x})} |\nabla \log(P_T f + \delta)|^2 \, d\nu_k \leq \frac{1}{E_{2\kappa}(T)} \Big[\mathcal{C}_T(\mu, \nu) - \mathscr{H}(\nu|\mathfrak{m}) \Big]$$

and it is now sufficient to let first $k \to \infty$ and then $\delta \downarrow 0$. In both cases, the monotone convergence theorem (and the fact that $c_k \to 1$) allows to handle the left-hand side and finally get the validity of (3.1.5) for μ, ν.

2nd step. Now let μ, ν be as in **A**1 and assume that $\frac{d\nu}{d\mathfrak{m}}$ is locally bounded away from 0 on $\mathrm{int}(\mathrm{supp}(\nu))$ and $\nu(\partial \mathrm{supp}(\nu)) = 0$. Fix $K \in \mathbb{N}$, let f_K, g_K be defined as in Lemma 3.1.1 and let $\mu_K := (\mathrm{proj}_{x_1})_\# \pi^K$ be the first marginal of the optimizer π^K associated to (3.1.1). Since this approximation guarantees us only that $f_K \in L^\infty(\mathfrak{m})$, let us fix $n \in \mathbb{N}$ and define $g_K^n := \min\{g_K, n\} \in L^\infty(\mathfrak{m})$ so that the previous step applies to the marginals

$$\mu_n^K := c_{K,n} f_K P_T g_K^n \, \mathfrak{m} \qquad \text{and} \qquad \nu_n^K := c_{K,n} g_K^n P_T f_K \, \mathfrak{m},$$

where $c_{K,n}$ is a normalization constant (again, this is the same for both μ_n^K and ν_n^K). Henceforth, in this case (3.1.5) reads as

$$\int |\nabla \log P_T f_K|^2 \, d\nu_n^K \leq \frac{1}{E_{2\kappa}(T)} \Big[\mathcal{C}_T(\mu_n^K, \nu_n^K) - \mathscr{H}(\nu_n^K|\mathfrak{m}) \Big].$$

Owing to algebraic manipulations, the normalising constant $c_{K,n}$ can be neglected and therefore it holds

$$\begin{aligned}&\int |\nabla \log P_T f_K|^2 \, g_K^n P_T f_K \, d\mathfrak{m} \\ &\qquad \leq \frac{1}{E_{2\kappa}(T)} \Big[\int f_K \log f_K \, P_T g_K^n \, d\mathfrak{m} - \int P_T f_K \log(P_T f_K) \, g_K^n \, d\mathfrak{m} \Big].\end{aligned}$$

Since by Lemma 3.1.1 f_K is bounded away from 0 and ∞, it follows that $\log f_K \in L^1(\mu^K)$ and $\log(P_T f_K) \in L^1(\nu)$. From this, by applying Fatou Lemma to the left-hand side and the Dominated Convergence Theorem to the right-hand one, in the limit as $n \to \infty$ we have

$$\int |\nabla \log P_T f_K|^2 \, g_K P_T f_K \, \mathrm{d}\mathfrak{m} \\ \leq \frac{1}{E_{2\kappa}(T)} \left[\int f_K \log f_K \, P_T g_K \, \mathrm{d}\mathfrak{m} - \int P_T f_K \log(P_T f_K) \, g_K \, \mathrm{d}\mathfrak{m} \right],$$

which is the corrector estimate (3.1.5) associated to $\mathcal{C}_T(\mu^K, \nu)$:

$$\int |\nabla \log P_T f_K|^2 \, \mathrm{d}\nu \leq \frac{1}{E_{2\kappa}(T)} \left[\mathcal{C}_T(\mu^K, \nu) - \mathscr{H}(\nu|\mathfrak{m}) \right]. \tag{3.1.12}$$

We wish now to take the limit as $K \to \infty$. However this point is more subtle, since we are interested in proving that $\nabla \log P_T f$ is well defined also in the limit. At this stage, let us stress out that the extra-assumption on the density $\frac{\mathrm{d}\nu}{\mathrm{d}\mathfrak{m}}$ and on $\mathrm{supp}(\nu)$ has never been used and it is just needed in the following discussion. First, notice that the right-hand side above is uniformly bounded in $K \in \mathbb{N}$ since Lemma 3.1.1 ensures that $\mathcal{C}_T(\mu^K, \nu) = \mathscr{H}(\pi^K|\mathrm{R}_{0,T})$ converges to $\mathcal{C}_T(\mu, \nu)$. As concerns the left-hand side, fix $\bar{x} \in \mathrm{int}(\mathrm{supp}(\nu))$. By assumption there exist $r, \alpha > 0$ such that $\bar{B}_r(\bar{x}) \subset \mathrm{supp}(\nu)$ and $\frac{\mathrm{d}\nu}{\mathrm{d}\mathfrak{m}} \geq \alpha$ $\mathfrak{m}$-a.e. in $\bar{B}_r(\bar{x})$, so that, for any fixed $\delta > 0$ it holds

$$\int |\nabla \log P_T f_K|^2 \, \mathrm{d}\nu \geq \int |\nabla \log(P_T f_K + \delta)|^2 \, \mathrm{d}\nu \geq \alpha \int_{\bar{B}_r(\bar{x})} |\nabla \log(P_T f_K + \delta)|^2 \, \mathrm{d}\mathfrak{m}.$$

As a byproduct of these bounds we have

$$\limsup_{K \to \infty} \int_{\bar{B}_r(\bar{x})} |\nabla \log(P_T f_K + \delta)|^2 \, \mathrm{d}\mathfrak{m} \leq \alpha^{-1} \, E_{2\kappa}(T)^{-1} \left[\mathcal{C}_T(\mu, \nu) - \mathscr{H}(\nu|\mathfrak{m}) \right],$$

which is finite. The functions $(|\nabla \log(P_T f_K + \delta)|)_{K \in \mathbb{N}}$ are thus equi-bounded in $\mathrm{L}^2(\bar{B}_r(\bar{x}), \mathfrak{m})$ and this implies that, up to subsequences, they converge to some function $G_{\bar{x}} \in \mathrm{L}^2(\bar{B}_r(\bar{x}), \mathfrak{m})$ weakly in $\mathrm{L}^2(\bar{B}_r(\bar{x}), \mathfrak{m})$. Moreover, from Lemma 3.1.1 we also have that $\log(P_T f_K + \delta)$ converges $\mathfrak{m}$-a.e. to $\log(P_T f + \delta)$ in $\bar{B}_r(\bar{x})$. Therefore, relying again on [AGS14a, Lemma 4.3(b)] we conclude that $|\nabla \log(P_T f + \delta)| \leq G_{\bar{x}}$ on $\bar{B}_r(\bar{x})$. In particular, it is worth stressing that in this case [AGS14a, Lemma 4.3(b)] also ensures that $\log(P_T f + \delta) \in W^{1,2}(\bar{B}_r(\bar{x}))$, which does not follow from the regularising effect of P_T, because of the possible lack of integrability of f.

To replicate the proof given in the previous step we also need to show that $|\nabla \log(P_T f_K + \delta)| \rightharpoonup G_{\bar{x}}$ in $\mathrm{L}^2(\bar{B}_r(\bar{x}), \nu)$, but this may fail, as $\frac{\mathrm{d}\nu}{\mathrm{d}\mathfrak{m}}$ needs not belong to $\mathrm{L}^\infty(\mathfrak{m})$. For this reason, let us introduce $[\nu]_N := \min\{g_K P_T f_K, N\}\mathfrak{m}$ for $N \in \mathbb{N}$, so that the left-hand side of (3.1.12) can be trivially estimated from below as

$$\int |\nabla \log P_T f_K|^2 \, \mathrm{d}[\nu]_N \leq \int |\nabla \log P_T f_K|^2 \, \mathrm{d}\nu \leq \frac{1}{E_{2\kappa}(T)} \left[\mathcal{C}_T(\mu^K, \nu) - \mathscr{H}(\nu|\mathfrak{m}) \right].$$

Since now $\frac{\mathrm{d}[\nu]_N}{\mathrm{d}\mathfrak{m}} \in \mathrm{L}^\infty(\mathfrak{m})$ by construction, the weak convergence $|\nabla \log(P_T f_K + \delta)| \rightharpoonup G_{\bar{x}}$ in $\mathrm{L}^2(\bar{B}_r(\bar{x}), \mathfrak{m})$ implies the weak convergence towards the same limit in $\mathrm{L}^2(\bar{B}_r(\bar{x}), [\nu]_N)$. Combining these considerations with the lower semicontinuity of the $\mathrm{L}^2([\nu]_N)$-norm, we obtain

$$\begin{aligned} \liminf_{K\to\infty} \int_{\bar{B}_r(\bar{x})} |\nabla \log(P_T f_K + \delta)|^2 \,\mathrm{d}[\nu]_N &\geq \int_{\bar{B}_r(\bar{x})} G_{\bar{x}}^2 \mathrm{d}[\nu]_N \\ &\geq \int_{\bar{B}_r(\bar{x})} |\nabla \log(P_T f + \delta)|^2 \,\mathrm{d}[\nu]_N \end{aligned}$$

whence

$$\int_{\bar{B}_r(\bar{x})} |\nabla \log(P_T f + \delta)|^2 \,\mathrm{d}[\nu]_N \leq \frac{1}{E_{2\kappa}(T)} \left[\mathcal{C}_T(\mu,\nu) - \mathscr{H}(\nu|\mathfrak{m}) \right].$$

Choosing now $(x_k)_{k\in\mathbb{N}} \subset \mathrm{int}(\mathrm{supp}(\nu))$ dense and denoting by r_k the radii associated to each x_k according to the previous construction, so that $\mathrm{int}(\mathrm{supp}(\nu)) = \cup_k \bar{B}_{r_k}(x_k)$, by a diagonal argument there exists a measurable function $G \in \mathrm{L}^2_{loc}(\mathfrak{m})$ such that $|\nabla \log(P_T f_K + \delta)| \rightharpoonup G$ in $\mathrm{L}^2(\bar{B}_{r_k}(x_k), \mathfrak{m})$ and $\mathrm{L}^2(\bar{B}_{r_k}(x_k), [\nu]_N)$ for every $k \in \mathbb{N}$ and, by [AGS14a, Lemma 4.3(b)], $|\nabla \log(P_T f + \delta)| \leq G$ $\mathfrak{m}$-a.e since $\log(P_T f_K + \delta) \to \log(P_T f + \delta)$ $\mathfrak{m}$-a.e. in $\mathrm{supp}(\nu)$. Setting $B_k := \cup_{i=1}^k \bar{B}_{r_i}(x_i)$, by the same reasoning as above (noting that the choice of N does not depend on the point x_k) we obtain for any $k \in \mathbb{N}$ that it holds

$$\int_{B_k} |\nabla \log(P_T f + \delta)|^2 \,\mathrm{d}[\nu]_N \leq \frac{1}{E_{2\kappa}(T)} \left[\mathcal{C}_T(\mu,\nu) - \mathscr{H}(\nu|\mathfrak{m}) \right].$$

Taking the limit as $k \to \infty$, by $\nu(\partial \mathrm{supp}(\nu)) = 0$ we infer that $|\nabla \log(P_T f + \delta)|$ actually belongs to $\mathrm{L}^2([\nu]_N)$. Passing then to the limit as $N \to \infty$ and $\delta \downarrow 0$, by monotonicity and again the fact that $\nu(\partial \mathrm{supp}(\nu)) = 0$ we precisely obtain (3.1.5) for μ, ν as in the statement. □

Notice that, when $\kappa > 0$ the above estimates state that as $T \gg 1$ the corrector norms decrease exponentially fast

$$\|\nabla \log P_T g\|_{\mathrm{L}^2(\mu)}, \|\nabla \log P_T f\|_{\mathrm{L}^2(\nu)} \lesssim \exp(-2\kappa T).$$

Combining this with the heuristic discussion of Section 2.3 suggests that the L^2-norm of the optimal control process in the control formulation of SP (as in Section 1.3) is actually exponentially small as T increases, at least in the first half of the time window, *i.e.*, in $[0, T/2]$. This would mean that the optimal control initially does not steer the controlled diffusion (2.3.4) to the final target ν, but actually let the diffusion process reach on its own its ergodic limit (the equilibrium measure $\mathfrak{m}$) and then, just at the end (let's say after time $t = T/2$), the optimal control starts drifting the diffusion process from the equilibrium towards its final target ν. This suggests the name *corrector estimates* we have adopted for Proposition 3.1.2.

The behaviour we have just suggested is often referred to as *turnpike property* in the (deterministic) control community. Historically, in the stochastic control setting, the corrector estimates have been firstly studied in the long-time limit of SP in order to establish entropic turnpike estimates for Schrödinger bridges [Con19, Theorem 1.4]. In this chapter we have generalised such results and we are going to show that such estimates prove to be a useful sharp tool also when dealing with the small-time limit for SP.

Let us conclude this section, by showing that, at least under the $\mathsf{CD}(\kappa,\infty)$ condition, the corrector estimates can be deduced as a corollary of the reverse log-Sobolev inequality [BGL13, Theorem 5.5.2 (v)] which states that for any positive function h and any $t \geq 0$ it holds

$$P_t(h\log h) - (P_t h)\log(P_t h) \geq E_{2\kappa}(t)\,\frac{|\nabla P_t h|^2}{P_t h}\,. \tag{3.1.13}$$

Indeed it is enough noticing that

$$\begin{aligned}
\|\nabla \log P_T f\|^2_{\mathrm{L}^2(\nu)} &= \int \frac{|\nabla P_T f|^2}{(P_T f)^2}\mathrm{d}\nu \overset{(2.2.2)}{=} \int \frac{|\nabla P_T f|^2}{P_T f}\, g\,\mathrm{d}\mathfrak{m} \\
&\overset{(3.1.13)}{\leq} \frac{1}{E_{2\kappa}(T)}\int g\left[P_T(f\log f) - (P_T f)\log(P_T f)\right]\mathrm{d}\mathfrak{m} \\
&= \frac{1}{E_{2\kappa}(T)}\left[\int (P_T g)(f\log f)\mathrm{d}\mathfrak{m} - \int g\,(P_T f)\log(P_T f)\mathrm{d}\mathfrak{m}\right] \\
&\overset{(2.2.2)}{=} \frac{1}{E_{2\kappa}(T)}\left[\int \log f\,\mathrm{d}\mu - \int \log(P_T f)\mathrm{d}\nu\right].
\end{aligned}$$

Since $\boldsymbol{\pi}^T \in \Pi(\mu,\nu)$ and $\log g \in \mathrm{L}^1(\nu)$, we end up with

$$\begin{aligned}
\|\nabla \log P_T f\|^2_{\mathrm{L}^2(\nu)} \leq & \frac{1}{E_{2\kappa}(T)}\left[\int_{M\times M} \log(f(x)g(y))\mathrm{d}\boldsymbol{\pi}^T(x,y) - \int_M \log(g P_T f)\mathrm{d}\nu\right] \\
= & \frac{1}{E_{2\kappa}(T)}\left[\mathcal{C}_T(\mu,\nu) - \mathscr{H}(\nu|\mathfrak{m})\right].
\end{aligned}$$

This proves (3.1.5). The estimate (3.1.6) can be proven in a similar fashion.

3.2 Small-time asymptotics of Schrödinger problem

In this section we are going to show that the gradients of Schrödinger potentials provide good proxies for the gradients of Kantorovich potentials and Brenier's map for the Optimal Transport problem.

For notations' clarity, we are going to stress out the dependence from the time parameter $T > 0$ and denote with φ^T, ψ^T and f^T, g^T respectively the Schrödinger potentials and the fg-decomposition associated to $\mathcal{C}_T(\mu,\nu)$, *i.e.*, when considering the time horizon $T > 0$.

3.2.1 Primal and zero-th order dual convergence results

Firstly, let us collect here known convergence results on primal and zero-th order dual formulations of SP. As we have already explained in Section 1.2 the link between SP and OT, in the classical Brownian motion case, relies on the key observation that $T \log \mathsf{p}_T(x,y) \propto |x-y|^2$ (up to normalising additive constant, cf. (1.2.4)). Remarkably the (CD) condition allows the validity of a similar estimate, at least in the asymptotics $T \downarrow 0$, even for general diffusion reference processes such as (2.1.1). More precisely, [Nor97, Theorem 1.1] states that uniformly on compact subsets of $M \times M$ it holds

$$T \log \mathsf{p}_T(x,y) \xrightarrow{T\downarrow 0} -\frac{1}{4}\,\mathsf{d}^2(x,y)\,. \tag{3.2.1}$$

This guarantees indeed the Γ-convergence of (rescaled) SP towards OT, and particularly that it holds (cf. [GT21, Remark 5.11] for $\mathsf{CD}(\kappa, N)$, or Remark 3.2.2 below for $\mathsf{CD}(\kappa,\infty)$)

$$\lim_{T\downarrow 0} T\,\mathcal{C}_T(\mu,\nu) = \frac{1}{4}\,\mathbf{W}_2^2(\mu,\nu)\,. \tag{3.2.2}$$

We refer the reader to the bibliographical remarks at the end of this chapter for further references on the primal zero-th order convergence results. Independently from those references, our discussion below will provide a proof of (3.2.2) (cf. Remark 3.2.2).

Similarly, (3.2.1) implies the validity of convergence results also on the dual zero-th order side, *i.e.*, the convergence of the (rescaled) Schrödinger potentials $\{-T\varphi^T\}_{T\in(0,1]}$ and $\{-T\psi^T\}_{T\in(0,1]}$. More precisely the (rescaled) potentials will converge, up to subsequences, to some limit measurable functions known as Kantorovich potentials, which are defined as the solution of the dual formulation of OT[2]

$$\frac{1}{4}\,\mathbf{W}_2^2(\mu,\nu) = \sup_{\varphi\in \mathrm{L}^1(\mu),\,\psi\in\mathrm{L}^1(\nu)\,:\,\varphi\oplus\psi\le\frac{1}{4}\mathsf{d}^2(\cdot,\cdot)} \left(\int \varphi\,\mathrm{d}\mu + \int \psi\,\mathrm{d}\nu\right)\,. \tag{3.2.3}$$

We will denote the optimiser of the above problem, *i.e.*, the Kantorovich potentials, with φ^0 and ψ^0, in complete analogy with the notation adopted for the Schrödinger potentials. Notice that any additive shift of couple of Kantorovich potentials, *i.e.*$(\varphi^0,\psi^0) \mapsto (\varphi^0 + a, \psi^0 - a)$ for any $a \in \mathbb{R}$, is again an optimiser for (3.2.3). Hence it is wise to impose a symmetric normalisation (as done in (2.2.13) for φ^T and ψ^T), which reads as

$$\int \varphi^0\,\mathrm{d}\mu = \int \psi^0\,\mathrm{d}\nu = \frac{1}{8}\,\mathbf{W}_2^2(\mu,\nu)\,. \tag{3.2.4}$$

[2]The presence of the unconventional factor $1/4$ comes from the choice of considering $\sqrt{2}\,\mathrm{d}B_t$ in the reference dynamics (2.1.1), which corresponds to the conventional choice of considering the (CD) condition for the Laplace Beltrami operator Δ_g in the generator L (instead of having $\Delta_g/2$ or $\Delta_g/4$).

As concerns (φ^0, ψ^0), due to the lack of regularity of the OT problem (3.2.3), even after the normalisation (3.2.4) uniqueness of the Kantorovich potentials may fail, in contrast to what happens for the Schrödinger potentials. Some results in this direction are known for some specific examples, for instance in the Euclidean case if at least one marginal is absolutely continuous with respect to the Lebesgue measure and if its support is connected then uniqueness holds [BGN22, Appendix B]. Finally, let us mention that the Brenier map is tightly linked to the Kantorovich potentials since under our assumptions it holds $\mathcal{T} = \mathrm{Id} - 2\nabla\varphi^0$ [Fig07, Proposition 3.1], [FG11, Theorem 1.1] (see also the discussion therein).

The convergence proof that we present here relies on [NW22, Proposition 5.1], where the authors consider EOT problems with time-dependent costs. Here we will consider either $c_T(x,y) := -T\log \mathsf{p}_T(x,y)$ or $\tilde{c}_T := -T\log\frac{\mathrm{dR}_{0,T}}{\mathrm{d}(\mu\otimes\nu)}$ as in Section 2.A. While the latter choice would guarantee that $(-T\,\varphi^T, -T\,\psi^T)$ are exactly equal to the entropic potentials (associated to $\tilde{c}_T$), it is in general more convenient working with c_T because of the dynamical interpretation of the cost as transition kernel of the associated SDE (2.1.1). The reader must then pay attention to the fact that the entropic potentials associated to c_T, which will be denoted as Φ_T and Ψ_T (and satisfying the symmetric normalisation (2.A.3)), differ from the previous ones. Similarly, the EOT optimal value will differ, though the two problems share the same optimiser. Indeed

$$\tilde{c}_T(x,y) = c_T(x,y) + T\log\rho(x) + T\log\sigma(y),$$

and therefore it holds

$$\mathrm{EOT}_T^{c_T}(\mu,\nu) = T\,\mathcal{C}_T(\mu,\nu) - T\,\mathscr{H}(\mu|\mathfrak{m}) - T\,\mathscr{H}(\nu|\mathfrak{m}) \tag{3.2.5}$$

and

$$-T\varphi^T = \Phi_T + T\log\rho \quad \text{and} \quad -T\psi^T = \Psi_T + T\log\sigma\,. \tag{3.2.6}$$

After this premise, let us show the zero-th order convergence in the dual formulation, *i.e.*, the strong convergence of the Schrödinger potentials (possibly along a subsequence) towards Kantorovich potentials.

Lemma 3.2.1 (L^1-convergence of the potentials). *Assume that $(M,\mathsf{d},\mathfrak{m})$ satisfies* (CD) *and that μ,ν satisfy **A1**. Then it holds:*

- *$\{\Phi_T\}_{T\in(0,1]}$ and $\{\Psi_T\}_{T\in(0,1]}$ are strongly precompact in $\mathrm{L}^1(\mu)$ and $\mathrm{L}^1(\nu)$ respectively and their accumulation points are Kantorovich potentials (φ^0,ψ^0) for* (3.2.3)*;*
- *or equivalently, $\{-T\varphi^T\}_{T\in(0,1]}$ and $\{-T\psi^T\}_{T\in(0,1]}$ are strongly precompact in $\mathrm{L}^1(\mu)$ and $\mathrm{L}^1(\nu)$ respectively and their accumulation points are Kantorovich potentials (φ^0,ψ^0) for* (3.2.3).

If the Kantorovich potentials (φ^0, ψ^0) *associated to* (3.2.3) *are unique, then it holds*

$$\Phi_T \to \varphi^0 \text{ strongly in } \mathrm{L}^1(\mu) \quad \text{and} \quad \Psi_T \to \psi^0 \text{ strongly in } \mathrm{L}^1(\nu)\,,$$

and equivalently

$$-T\varphi^T \to \varphi^0 \text{ strongly in } \mathrm{L}^1(\mu) \quad \text{and} \quad -T\psi^T \to \psi^0 \text{ strongly in } \mathrm{L}^1(\nu)\,.$$

Proof. Firstly, notice that the equivalence between the statements follows from (3.2.6) and the finite entropy condition in **A**1. By [NW22, Proposition 5.1], to obtain the desired convergence for Φ_T and Ψ_T it is sufficient to show that $c_T := -T\log \mathsf{p}_T \to \mathsf{d}^2/4$ uniformly on compact subsets as $T \to 0$ (which is (3.2.1)) and that there exists a function $\bar{c}(x,y) = \bar{c}_1(x) + \bar{c}_2(y)$ with $\bar{c}_1 \in L^1(\mu)$ and $\bar{c}_2 \in L^1(\nu)$ such that $c_T \le \bar{c}$ for all T sufficiently small, say $T \le 1$.

Under the $\mathsf{CD}(\kappa, N)$ condition with $N < \infty$, using the heat kernel lower bound (2.1.7) and assuming without loss of generality that $T \le 1$, so that $\log \mathfrak{m}(B_{\sqrt{T}}(x)) \le \log \mathfrak{m}(B_1(x))$, we see that

$$\begin{aligned} c_T(x,y) &\le T\log C_1 + T\log \mathfrak{m}(B_{\sqrt{T}}(x)) + \frac{\mathsf{d}^2(x,y)}{4-\delta} + C_2T^2 \\ &\le T\log C_1 + (T\log \mathfrak{m}(B_1(x)))^+ + \frac{\mathsf{d}^2(x,y)}{4-\delta} + C_2T^2. \end{aligned}$$

By the trivial inequality $\mathsf{d}^2(x,y) \le 2(\mathsf{d}^2(x,z) + \mathsf{d}^2(y,z))$ valid for any $z \in M$ we conclude that $c_T(x,y) \le c'_T(x) + c''_T(y)$ with

$$\begin{aligned} c'_T(x) &:= (T\log \mathfrak{m}(B_1(x)))^+ + \frac{2}{4-\delta}\mathsf{d}^2(x,z)\,, \\ c''_T(y) &:= T\log C_1 + \frac{2}{4-\delta}\mathsf{d}^2(y,z) + C_2T^2\,. \end{aligned}$$

By the fact that $\nu \in \mathcal{P}_2(M)$, it is clear that c''_T can be dominated by a ν-integrable function not depending on T. As regards c'_T it follows from $\mu \in \mathcal{P}_2(M)$ and Lemma 2.1.1 which gives the μ-integrability of the positive part of $\log \mathfrak{m}(B_1(\cdot))$.

On the other hand, if we assume that $(M, \mathsf{d}, \mathfrak{m})$ satisfies $\mathsf{CD}(\kappa, \infty)$ and that $\mathfrak{m}(M) = 1$, then leveraging on the heat kernel lower bound (2.1.8) we see that

$$c_T(x,y) \le \frac{\kappa T}{2(1-e^{-\kappa T})}\mathsf{d}^2(x,y)$$

and in this case the trivial inequality $\mathsf{d}^2(x,y) \le 2(\mathsf{d}^2(x,z) + \mathsf{d}^2(y,z))$, valid for any $z \in M$, readily provides us with functions c'_T, c''_T which are μ- and ν-integrable respectively.

Therefore we can apply [NW22, Proposition 5.1], which concludes our proof. □

Remark 3.2.2. *Let us mention that the previous result gives as a direct consequence a new proof of the small-time limit of the normalised Schrödinger cost, that is, under the same assumptions of Lemma 3.2.1 we have proven the validity of* (3.2.2).

3.2.2 Convergence of the gradients to the Brenier map

Even though in general we may not assume the uniqueness of (normalised) Kantorovich potentials, reason why in Lemma 3.2.1 we couldn't straightforwardly deduce L^1-convergence for the full sequences, when working with the gradients uniqueness of the limits holds. More precisely, we will show the $L^2(\mu)$-convergence of the Schrödinger map $\boldsymbol{\mathcal{T}}^T = \mathrm{Id} + 2T\nabla\varphi^T$ towards the the Brenier map $\boldsymbol{\mathcal{T}} = \mathrm{Id} - 2\nabla\varphi^0$, whose uniqueness indeed holds under our assumptions (cf. [Fig07, Proposition 3.1], [FG11, Theorem 1.1] and discussion therein). We will prove this for marginals with finite Fisher information (w.r.t. $\mathfrak{m}$), which is defined for any $\mathfrak{p} \in \mathcal{P}(M)$ as

$$\mathcal{I}(\mathfrak{p}) := \left\| \nabla \log \frac{\mathrm{d}\mathfrak{p}}{\mathrm{d}\mathfrak{m}} \right\|^2_{L^2(\mathfrak{p})} = \int \left| \nabla \log \frac{\mathrm{d}\mathfrak{p}}{\mathrm{d}\mathfrak{m}} \right|^2 \mathrm{d}\mathfrak{p} \,. \tag{3.2.7}$$

Theorem 3.2.3. *Suppose $(M, \mathsf{d}, \mathfrak{m})$ satisfies* (CD) *and that **A1** holds true. If $\frac{\mathrm{d}\mu}{\mathrm{d}\mathfrak{m}}$ is locally bounded away from* 0 *on* $\mathrm{int}(\mathrm{supp}(\mu))$, *if $\mu(\partial\,\mathrm{supp}(\mu)) = 0$ and if the Fisher information $\mathcal{I}(\mu)$ is finite, then the Schrödinger map $\boldsymbol{\mathcal{T}}^T$ from μ to ν converges strongly in $L^2(\mu)$ to the Brenier map from μ to ν in the small-time limit. Equivalently, there are (φ^0, ψ^0) Kantorovich potentials such that as $T \downarrow 0$*

$$-T\nabla\varphi^T \to \nabla\varphi^0 \quad \text{strongly in } L^2(\mu)\,.$$

A similar result holds for $T\nabla\psi^T$ under the corresponding assumptions on ν.

Proof. For the readers' convenience we divide our proof into four steps.

1st step: weak compactness of the gradients in $L^2(\mu)$. From the identity $\mu = fP_Tg\,\mathfrak{m}$ we deduce that $\log \mathrm{d}\mu/\mathrm{d}\mathfrak{m} = -\varphi^T - \mathcal{U}_0^{T,\psi}$, where $\mathcal{U}_0^{T,\psi} := -\log P_T e^{-\psi^T}$ and hence

$$\begin{aligned}\left\| T\nabla\varphi^T \right\|_{L^2(TM,\mu)} &= \left\| T\nabla \log\rho + T\nabla\mathcal{U}_0^{T,\psi} \right\|_{L^2(TM,\mu)} \\ &\le T\left\| \nabla \log\rho \right\|_{L^2(TM,\mu)} + T\left\| \nabla\mathcal{U}_0^{T,\psi} \right\|_{L^2(TM,\mu)}.\end{aligned}$$

The correctors estimate (3.1.6) allows to control the last term as

$$\left\| \nabla\mathcal{U}_0^{T,\psi} \right\|^2_{L^2(TM,\mu)} \le \frac{1}{E_{2\kappa}(T)} \left[\mathcal{C}_T(\mu,\nu) - \mathscr{H}(\mu|\mathfrak{m}) \right]$$

whence

$$\|T\nabla\varphi^T\|_{L^2(TM,\mu)} \le T\sqrt{\mathcal{I}(\mu)} + \sqrt{\frac{T}{E_{2\kappa}(T)}} \sqrt{T\,\mathcal{C}_T(\mu,\nu) - T\,\mathscr{H}(\mu|\mathfrak{m})}.$$

This inequality together with the fact that $\lim_{T\downarrow 0} \frac{E_{2\kappa}(T)}{T} = 1$, $\lim_{T\downarrow 0} T\mathcal{C}_T(\mu,\nu) = \frac{1}{4}\mathbf{W}_2^2(\mu,\nu)$ (cf. Remark 3.2.2) and $\mathcal{I}(\mu) < +\infty$ implies that

$$\limsup_{T\to 0} \left\| T\nabla\varphi^T \right\|_{L^2(TM,\mu)} \le \frac{\mathbf{W}_2(\mu,\nu)}{2}. \tag{3.2.8}$$

Therefore, the sequence $\{T\nabla\varphi^T\}_{T>0}$ is weakly compact in $L^2(\mu)$.

2nd step: weak convergence of the gradients in $L^2(\mu_k)$. Fix $k \in \mathbb{N}$ and define $\rho_k := \min\{d\mu/d\mathfrak{m}, k\}$, $\mu_k := \rho_k\mathfrak{m}$ (note that μ_k needs not to be a probability measure). We claim that for any k (large enough) and for any weakly convergent subsequence of $\{T\nabla\varphi^T\}_{T>0}$ there exist a Kantorovich potential φ^0 and a subsubsequence $T_n \downarrow 0$ such that $T_n\nabla\varphi^{T_n} \rightharpoonup -\nabla\varphi^0$ in $L^2(TM, \mu_k)$.

Firstly, note that for any subsequence of $\{T\nabla\varphi^T\}_{T>0}$, we can consider the corresponding subsequence of $\{T\varphi^T\}_{T>0}$ and deduce from Lemma 3.2.1 that there exist a Kantorovich potential $\varphi^0 \in L^1(\mu)$ and a subsubsequence $T_n \downarrow 0$ such that $T_n\varphi^{T_n} \to -\varphi^0$ strongly in $L^1(\mu)$, and a fortiori in $L^1(\mu_k)$ because $\mu_k \leq \mu$ trivially. Moreover, the family $\{T\nabla\varphi^T\}_{T>0}$ is weakly compact also in $L^2(TM, \mu_k)$. Hence, given a subsequence of $\{T\nabla\varphi^T\}_{T>0}$ weakly converging to some $\zeta_k \in L^2(TM, \mu_k)$ there exists $T_n \downarrow 0$ such that

$$-T_n\varphi^{T_n} \to \varphi^0 \text{ in } L^1(\mu_k) \qquad \text{and} \qquad T_n\nabla\varphi^{T_n} \rightharpoonup \zeta_k \text{ in } L^2(TM, \mu_k)\,. \tag{3.2.9}$$

Our claim is proven once we show that the weak limit ζ_k actually does not depend on k and coincides with the weak gradient $-\nabla\varphi^0$ $\mathfrak{m}$-a.e.

To this end, fix $x \in \mathrm{int}(\mathrm{supp}(\mu))$. By assumption there exists an open neighbourhood $B \subset \mathrm{supp}(\mu)$ of x and a constant $c > 0$ such that $\rho = d\mu/d\mathfrak{m} \geq c$ $\mathfrak{m}$-a.e. on B. Without loss of generality, we can assume that $B = B(x, r)$ for some radius $r > 0$. In this way compactness in $L^2(B, \mu)$ implies compactness in $L^2(B, \mathfrak{m})$ and, a fortiori, in $L^2(B, \mathrm{vol})$, since $\mathfrak{m} = e^{-U}\mathrm{vol}$ and $c' \leq e^{-U} \leq C'$ on B for some constants c', $C' > 0$. More explicitly,

$$\begin{aligned} c'T^2 \int_B \left|\nabla\varphi^T\right|^2 d\mathrm{vol} \leq T^2 \int_B \left|\nabla\varphi^T\right|^2 d\mathfrak{m} \leq \frac{T^2}{c} \int_B \left|\nabla\varphi^T\right|^2 d\mu \\ \leq \frac{1}{c} \left\|T\nabla\varphi^T\right\|^2_{L^2(\mu)} \end{aligned} \tag{3.2.10}$$

and the right-hand side is uniformly bounded in T by (3.2.8).

Now, inspired by the proof of [BK08, Proposition 2.14], we observe that the Neumann Laplacian on the smooth compact Riemannian manifold with boundary $(\overline{B}, g, \mathrm{vol}|_{\overline{B}})$ has a spectral gap (see e.g. [GHL04]). This is equivalent to the fact that $\mathrm{vol}|_{\overline{B}}$ satisfies a Poincaré inequality, whence in particular

$$\int_B \left|T\varphi^T\right|^2 d\mathrm{vol} \leq \left(\int_B T\varphi^T d\mathrm{vol}\right)^2 + C_P \int_B \left|T\nabla\varphi^T\right|^2 d\mathrm{vol}$$

for some constant $C_P > 0$. Note that the first term on the right-hand side is uniformly bounded in T_n, since by the fact that $\rho_k \geq c$ (provided $k \geq c$) and $c'\mathrm{vol} \leq \mathfrak{m} \leq C'\mathrm{vol}$ $\mathfrak{m}$-a.e. in B, (3.2.9) implies $T_n\varphi^{T_n} \to -\varphi^0$ in $L^1(B, \mathrm{vol})$, so that $(\int_B T_n\varphi^{T_n} d\mathrm{vol})^2$ converges to $(\int_B \varphi^0 d\mathrm{vol})^2$ as $T_n \downarrow 0$. The second one is bounded as well by (3.2.10). Hence we obtain compactness in $L^2(B, \mathrm{vol})$ for the subsequence $\{T_n\varphi^{T_n}\}_{n\in\mathbb{N}}$.

As a consequence, there exist a non-relabelled subsequence and a limit element $\xi \in \mathrm{L}^2(B, \mathrm{vol})$ such that $T_n\varphi^{T_n} \rightharpoonup \xi$ in $\mathrm{L}^2(B, \mathrm{vol})$. On the other hand, by (3.2.9) we know that $T_n\varphi^{T_n} \to -\varphi^0$ in $\mathrm{L}^1(\mu_k)$. This implies that $\xi = -\varphi^0$ μ_k-a.s. on B and, as a byproduct, that

$$T_n\varphi^{T_n} \rightharpoonup -\varphi^0 \text{ in } \mathrm{L}^2(B, \mathrm{vol}) \tag{3.2.11}$$

without passing to a subsequence. Indeed let us consider an arbitrary non-negative $\phi \in C_c^\infty(B)$, so that $\phi \in \mathrm{L}^1 \cap \mathrm{L}^\infty(B, \mathrm{vol})$, and start observing that

$$\left|\int_B \phi(\varphi^0 + \xi)\,\mathrm{dvol}\right| \leq \left|\int_B \phi(\varphi^0 + T_n\varphi^{T_n})\,\mathrm{dvol}\right| + \left|\int_B \phi(\xi - T_n\varphi^{T_n})\,\mathrm{dvol}\right|.$$

The second term vanishes since $T_n\varphi^{T_n} \rightharpoonup \xi$ in $\mathrm{L}^2(B, \mathrm{vol})$. As for the first one, note that

$$\begin{aligned}
\left|\int_B \phi(\varphi^0 + T_n\varphi^{T_n})\,\mathrm{dvol}\right| &\leq \|\phi\|_\infty \int_B \left|\varphi^0 + T_n\varphi^{T_n}\right|\,\mathrm{dvol} \\
&\leq \frac{1}{c'}\|\phi\|_\infty \int_B \left|\varphi^0 + T_n\varphi^{T_n}\right|\,\mathrm{d}\mathfrak{m} \\
&\leq \frac{1}{cc'}\|\phi\|_\infty \int_B \left|\varphi^0 + T_n\varphi^{T_n}\right|\,\mathrm{d}\mu_k \\
&\leq \frac{1}{cc'}\|\phi\|_\infty \left\|\varphi^0 + T_n\varphi^{T_n}\right\|_{\mathrm{L}^1(\mu_k)}
\end{aligned}$$

where for the second and third inequality we have used once more that $\mathfrak{m} \geq c'\mathrm{vol}$ and $\rho_k \geq c$ $\mathfrak{m}$-a.e. on B respectively. Hence, from $T_n\varphi^{T_n} \to -\varphi^0$ in $\mathrm{L}^1(\mu_k)$ it follows the weak convergence $T_n\varphi^{T_n} \rightharpoonup -\varphi^0$ in $\mathrm{L}^2(B, \mathrm{vol})$, without passing to any subsubsequence.

Now we are ready to prove the claim of this second step. Let us take $\beta \in C_c^\infty(B, TM)$ and observe that the assumption $\mathcal{I}(\mu) < \infty$ can be equivalently restated as $\nabla\sqrt{\rho} \in \mathrm{L}^2(TM, \mathfrak{m})$. As a consequence, and by definition of ρ_k, $\nabla\sqrt{\rho_k} \in \mathrm{L}^2(TM, \mathfrak{m})$ too and this fact together with $\rho_k \leq k$ justifies the validity of the chain rule $\nabla\rho_k = \nabla((\sqrt{\rho_k})^2) = 2\sqrt{\rho_k}\nabla\sqrt{\rho_k}$ and proves that $|\nabla\rho_k| \in \mathrm{L}^2(B, \mathfrak{m})$, whence the applicability of the integration by parts formula, so that

$$T\int \langle\nabla\varphi^T, \beta\rangle\,\mathrm{d}\mu_k = -T\int \varphi^T \mathrm{div}(\beta)\,\mathrm{d}\mu_k - T\int \varphi^T\langle\beta, \nabla\rho_k\rangle\,\mathrm{d}\mathfrak{m}.$$

On the one hand, for the left-hand side we know that

$$\lim_{T_n\downarrow 0} T_n\int \langle\nabla\varphi^{T_n}, \beta\rangle\,\mathrm{d}\mu_k = \int \langle\zeta_k, \beta\rangle\,\mathrm{d}\mu_k = \int \langle\rho_k\,\zeta_k, \beta\rangle\,\mathrm{d}\mathfrak{m}.$$

As concerns the right-hand side, the fact that $\mathrm{div}(\beta)$ is bounded in B and the strong convergence in (3.2.9) ensure that

$$\lim_{T_n\downarrow 0} T_n\int \varphi^{T_n}\mathrm{div}(\beta)\,\mathrm{d}\mu_k = -\int \varphi^0\mathrm{div}(\beta)\,\mathrm{d}\mu_k = -\int \varphi^0\rho_k\,\mathrm{div}(\beta)\,\mathrm{d}\mathfrak{m}.$$

For the limit of the second term, $\langle \beta, \nabla \rho_k \rangle \in \mathrm{L}^2(\mathfrak{m})$ together with $\mathrm{supp}(\beta) \subset B$, $e^{-U/2} \leq \sqrt{C'}$ in B and $T\varphi^T \rightharpoonup -\varphi^0$ in $\mathrm{L}^2(B, \mathrm{vol})$, proved in (3.2.11) above, allow us to deduce that

$$\lim_{T_n \downarrow 0} T_n \int \varphi^{T_n} \langle \beta, \nabla \rho_k \rangle \, \mathrm{d}\mathfrak{m} = -\int \varphi^0 \langle \beta, \nabla \rho_k \rangle \, \mathrm{d}\mathfrak{m}.$$

Hence, after rearrangement,

$$\begin{aligned} \int \varphi^0 \rho_k \, \mathrm{div}(\beta) \, \mathrm{d}\mathfrak{m} &= \int \langle \rho_k \, \zeta_k, \beta \rangle \, \mathrm{d}\mathfrak{m} - \int \varphi^0 \langle \beta, \nabla \rho_k \rangle \, \mathrm{d}\mathfrak{m} \\ &= \int \langle \beta, \rho_k \, \zeta_k - \varphi^0 \, \nabla \rho_k \rangle \, \mathrm{d}\mathfrak{m}, \end{aligned}$$

which means that $\nabla(\varphi^0 \rho_k) = \varphi^0 \nabla \rho_k - \rho_k \, \zeta_k$ on B by definition of weak gradient.

To conclude that $\zeta_k = -\nabla \varphi^0$ on B, recall that $\rho_k \in W^{1,2}(B)$. Using the lower bound $\rho_k \geq c$ $\mathfrak{m}$-a.e. on B, by chain rule we get that also $\rho_k^{-1} \in W^{1,2}(B)$ and therefore, by Leibniz rule,

$$\nabla \varphi^0 = \nabla(\varphi^0 \rho_k \cdot \rho_k^{-1}) = \frac{\nabla(\varphi^0 \rho_k)}{\rho_k} + \varphi^0 \rho_k \nabla(\rho_k^{-1}) = -\zeta_k \qquad \mathfrak{m}\text{-a.e. on } B.$$

Since B was obtained starting from an arbitrary $x \in \mathrm{int}(\mathrm{supp}(\mu))$, we conclude that $\zeta_k = -\nabla \varphi^0$ $\mathfrak{m}$-a.e. in $\mathrm{int}(\mathrm{supp}(\mu))$ and therefore, up to changing the weak limit on a μ-null set we have proven that

$$T_n \nabla \varphi^{T_n} \rightharpoonup -\nabla \varphi^0 \text{ in } \mathrm{L}^2(TM, \mu_k)\,. \tag{3.2.12}$$

3rd step: weak convergence of the gradients in $\mathrm{L}^2(\mu)$. We now claim that the result above can be improved to

$$T_n \nabla \varphi^{T_n} \rightharpoonup -\nabla \varphi^0 \text{ in } \mathrm{L}^2(TM, \mu)\,.$$

To this end, fix $\beta \in \mathrm{L}^2 \cap \mathrm{L}^\infty(TM, \mu)$ and start observing

$$\begin{aligned} \left| \int \langle T_n \nabla \varphi^{T_n} + \nabla \varphi^0, \beta \rangle \, \mathrm{d}\mu \right| \leq & \left| T_n \int \langle \nabla \varphi^{T_n}, \beta \rangle (\rho - \rho_k) \, \mathrm{d}\mathfrak{m} \right| \\ & + \left| \int \langle T_n \nabla \varphi^{T_n} + \nabla \varphi^0, \beta \rangle \, \mathrm{d}\mu_k \right| + \left| \int \langle \nabla \varphi^0, \beta \rangle (\rho - \rho_k) \, \mathrm{d}\mathfrak{m} \right|. \end{aligned}$$

Recalling that $\mathbf{W}_2(\mu, \nu) = 2 \left\| \nabla \varphi^0 \right\|_{\mathrm{L}^2(TM, \mu)}$ (see [AGS14a, Theorem 10.3], paying attention to the different rescaling), we see that the third term on the right-hand side vanishes as $k \to \infty$ by dominated convergence theorem. As concerns the first one, let us first estimate it as follows

$$\begin{aligned} \left| T_n \int \langle \nabla \varphi^{T_n}, \beta \rangle (\rho - \rho_k) \, \mathrm{d}\mathfrak{m} \right| &\leq T_n \int |\langle \nabla \varphi^{T_n}, \beta \rangle| \Big(1 - \frac{\rho_k}{\rho} \Big) \, \mathrm{d}\mu \\ \leq & \|\beta\|_\infty \int T_n |\nabla \varphi^{T_n}| \Big(1 - \frac{\rho_k}{\rho} \Big) \, \mathrm{d}\mu \leq \|\beta\|_\infty \|T_n \nabla \varphi^{T_n}\|_{\mathrm{L}^2(TM, \mu)} \Big\| 1 - \frac{\rho_k}{\rho} \Big\|_{\mathrm{L}^2(\mu)} \end{aligned}$$

and then observe that $\|T_n \nabla \varphi^{T_n}\|_{L^2(TM,\mu)}$ is bounded as $T_n \downarrow 0$ by (3.2.8), while $\left\|1 - \frac{\rho_k}{\rho}\right\|_{L^2(\mu)} \to 0$ as $k \to \infty$ by dominated convergence. This allows to deduce that, given $\varepsilon > 0$ there exists k' independent of T_n so that

$$\left| \int \langle T_n \nabla \varphi^{T_n} + \nabla \varphi^0, \beta \rangle \, d\mu \right| \leq \left| \int \langle T_n \nabla \varphi^{T_n} + \nabla \varphi^0, \beta \rangle \, d\mu_{k'} \right| + 2\varepsilon.$$

Taking now the limit as $T_n \downarrow 0$, the right-hand side converges to 2ε by (3.2.12) and the arbitrariness of ε together with the density of $L^2 \cap L^\infty(TM, \mu)$ as subspace of $L^2(TM, \mu)$ allows to conclude that $T_n \nabla \varphi^{T_n} \rightharpoonup -\nabla \varphi^0$ in $L^2(TM, \mu)$ as claimed.

4th step: strong convergence of the gradients in $L^2(\mu)$. The previous step and the uniqueness of the Brenier map $\mathcal{T} = \mathrm{Id} - 2\nabla \varphi^0$ grant that the whole sequence of gradients converges

$$T \nabla \varphi^T \rightharpoonup -\nabla \varphi^0 \text{ in } L^2(TM, \mu_k), \qquad \forall k \in \mathbb{N},$$

which automatically implies

$$\liminf_{T \to 0} \left\| T \nabla \varphi^T \right\|_{L^2(TM,\mu_k)} \geq \left\| \nabla \varphi^0 \right\|_{L^2(TM,\mu_k)}.$$

Recalling again that $\mathbf{W}_2(\mu, \nu) = 2 \left\| \nabla \varphi^0 \right\|_{L^2(TM,\mu)}$, we deduce from (3.2.8) that

$$\lim_{T \to 0} \left\| T \nabla \varphi^T \right\|_{L^2(TM,\mu)} = \left\| \nabla \varphi^0 \right\|_{L^2(TM,\mu)}.$$

Since in a Hilbert space such as $L^2(TM, \mu)$ weak convergence plus convergence of the norm implies strong convergence, we obtain the desired conclusion. □

Remark 3.2.4. *The previous proof runs exactly in the same way if we consider the entropic potentials (Φ_T, Ψ_T) instead of the rescaled Schrödinger potentials $(T\varphi^T, T\psi^T)$. Moreover, the weak compactness in the first step can be proven without assuming finite Fisher information $\mathcal{I}(\mu)$, since*

$$\|T \nabla \Phi_T\|_{L^2(\mu)} = \left\| T \nabla \varphi^T + T \nabla \log d\mu / d\mathfrak{m} \right\|_{L^2(\mu)} = T \left\| \nabla \mathcal{U}_0^{T,\psi} \right\|_{L^2(\mu)}.$$

However, even though we are considering the entropic potentials (Φ_T, Ψ_T), the assumption of finite Fisher information $\mathcal{I}(\mu) < +\infty$ is needed in the second step where we identify the weak limit of $\{\nabla \Phi_{T_n}\}_{n \in \mathbb{N}}$ with $\nabla \varphi^0$.

Corollary 3.2.5. *Under the same assumptions of Theorem 3.2.3, if μ satisfies a Poincaré inequality, then the convergence of the potentials in Lemma 3.2.1 holds true also in $L^2(\mu)$.*

3.2.3 Quantitative convergence of gradients

In this last section we are going to show how functional inequalities, combined with Theorem 2.2.5 and Lemma 2.3.2, can be employed when proving quantitative convergence rates for the gradients in the classical SP setting. Remarkably, this problem is tightly linked with the suboptimality of the Schrödinger plan $\boldsymbol{\pi}^T$ for the OT problem (1.2.7).

Theorem 3.2.6 (Suboptimality of $\boldsymbol{\pi}^T$ for $\mathbf{W}_2$)**.** *Assume **A**1, **A**2 and that $\mu \in \mathcal{P}_{2+\delta}(\mathbb{R}^d)$ for some $\delta > 0$. Let $\boldsymbol{\pi}^T \in \Pi(\mu,\nu)$ be the Schrödinger plan for the classical SP* (1.2.2)*. Then, there exists a constant $C_\mu > 0$ (depending on μ and $\delta > 0$) such that*

$$\int |x-y|^2 \,\mathrm{d}\boldsymbol{\pi}^T - \mathbf{W}_2^2(\mu,\nu)$$
$$\leq \begin{cases} T\left[2\,\mathrm{Ent}(\nu) + d\,\log\left(\frac{32\pi\,C_\mu}{d}\sqrt{1+\frac{\beta_\mu}{\alpha_\nu}}\right) + d\right] & \text{if } \beta_\mu \leq \alpha_\nu, \\ T\left[2\,\mathrm{Ent}(\nu) + d\,\log\left(\frac{32\pi\,C_\mu}{d}\,\frac{\beta_\mu}{\alpha_\nu}\,\sqrt{1+\frac{\beta_\mu}{\alpha_\nu}}\right) + d\right] & \text{if } \alpha_\nu < \beta_\mu < +\infty, \\ 2\,d\,T\,\log(1/T) + O(T) & \text{if } \beta_\mu = +\infty. \end{cases}$$

Before proving this result, let us briefly explain the role of the constant C_μ, as well as the role of the finite $2+\delta$-moment assumption for the marginal μ.

In our proof strategy we are going to rely on explicit convergence rates for EOT towards OT as the parameter T vanishes (that is a quantitative version of what we have shown in (3.2.2)). To this end, we will apply [EN22a, Theorem 3.8] that requires the OT plan $\boldsymbol{\pi}^0$ to satisfy a quantisation property, which can be summarised as being well-approximated by probability measures supported on a finite number of points. More precisely, we say that a measure $\mathfrak{p} \in \mathcal{P}(\mathcal{X})$ satisfies the quantisation property with constant $C \geq 0$ and rate $\alpha > 0$ if for all $n \geq 1$ we have

$$\exists\, \mathfrak{p}^n \in \mathcal{P}^n(\mathcal{X}) \,:\, \mathbf{W}_2(\mathfrak{p}^n,\mathfrak{p}) \leq C\,n^{-\alpha}\,, \qquad (\mathrm{quant}_2(C,\alpha))$$

where $\mathcal{P}^n(\mathcal{X})$ is the set of probability measures on $\mathcal{X}$ supported on at most n points. A nice property of $(\mathrm{quant}_2(C,\alpha))$ is that its validity is preserved along Lipschitz transport maps, *i.e.*, if $\mathfrak{p}$ satisfies $(\mathrm{quant}_2(C,\alpha))$ and $\mathcal{T}$ is a Lipschitz transformation, then the pushforward $\mathcal{T}_\#\mathfrak{p}$ satisfies $(\mathrm{quant}_2(C,\alpha))$ with the same rate $\alpha > 0$ and with constant $C\,\mathrm{Lip}(\mathcal{T})$.

Notice that under **A**2, Caffarelli's contraction Theorem guarantees that the OT map $\boldsymbol{\mathcal{T}}$ is $\sqrt{\beta_\mu/\alpha_\nu}$-Lipschitz (see also [CP23]). As a consequence of this, since $\boldsymbol{\pi}^0 = (\mathrm{Id},\boldsymbol{\mathcal{T}})_\#\mu$, if the first marginal μ satisfies $(\mathrm{quant}_2(C,\alpha))$, then $\boldsymbol{\pi}^0$ does as well with same rate and with constant $C\sqrt{1+\beta_\mu/\alpha_\nu}$.

Therefore, in order to apply [EN22a, Theorem 3.8] we just need the first marginal to satisfy $(\mathrm{quant}_2(C,\alpha))$. If we further assume that μ has finite $2+\delta$-moment (for some positive $\delta > 0$) than a proof of this can be found in [GL00, Theorem 6.2], where the authors prove (quant$_2(C,\alpha)$) for $\alpha = 1/d$ and (asymptotic) constant

$$C_\mu := \lim_{n\to\infty} n^{1/d} \inf_{\mu^n \in \mathcal{P}^n(\mathbb{R}^d)} \mathbf{W}_2(\mu^n,\mu) \,. \tag{3.2.13}$$

See also [GL00, Corollary 6.7] for non-asymptotic explicit bounds for C_μ (with dependence from δ and from the $2+\delta$-moment of μ).

In conclusion, **A**2 and $\mu \in \mathcal{P}_{2+\delta}(\mathbb{R}^d)$ guarantee that the OT plan $\boldsymbol{\pi}^0$ satisfies (quant$_2(C,\alpha)$) with rate $\alpha = 1/d$ and constant $C = C_\mu\sqrt{1+\beta_\mu/\alpha_\nu}$ and therefore they further guarantee the validity of [EN22a, Theorem 3.8].

Proof of Theorem 3.2.6. Firstly, let us recall that according to Lemma 2.3.2 we can consider the stochastic interpretation $\boldsymbol{\pi}^T = \mathcal{L}(X_0^T, X_T^T)$ where

$$\begin{cases} \mathrm{d}X_t^T = -\nabla \mathcal{U}_t^{T,\psi^T}(X_t^T)\mathrm{d}t + \mathrm{d}B_t \\ X_0^T \sim \mu \,. \end{cases} \tag{3.2.14}$$

Therefore we can immediately deduce that

$$\begin{aligned} \int |x-y|^2\,\mathrm{d}\boldsymbol{\pi}^T &= \mathbb{E}[|X_0^T - X_T^T|^2] = \mathbb{E}\left[\left|-\int_0^T \nabla\mathcal{U}_t^{T,\psi^T}(X_t^T)\,\mathrm{d}t + B_T\right|^2\right] \\ &= T + \mathbb{E}\left[\left|\int_0^T \nabla\mathcal{U}_t^{T,\psi^T}(X_t^T)\,\mathrm{d}t\right|^2\right] - 2\,\mathbb{E}\left[B_T\cdot\int_0^T \nabla\mathcal{U}_t^{T,\psi^T}(X_t^T)\,\mathrm{d}t\right] \,, \end{aligned} \tag{3.2.15}$$

where the expectation is taken under the law $\mathbf{P}^T$ of the process (3.2.14). Next, since $(\mathcal{U}_t^{T,\psi^T})_{t\in[0,T]}$ solves HJB (2.3.3)[3], a straightforward application of Ito's formula shows that $(\nabla\mathcal{U}_t^{T,\psi^T}(X_t^T))_{t\in[0,T]}$ is a martingale (cf. [Con24, Proof of Theorem 2.1]) and satisfies

$$\mathrm{d}\nabla\mathcal{U}_t^{T,\psi^T}(X_t^T) = \nabla^2\mathcal{U}_t^{T,\psi^T}(X_t^T)\,\mathrm{d}B_t \,.$$

As a consequence of this, and from the independence between X_0^T and the

[3] Pay attention that here we consider Brownian motions, therefore there is no drift, *i.e.*, $U = 0$ and there is a factor $1/2$ in front of the Laplacian.

Brownian motion, the last term in (3.2.15) can be rewritten as

$$\begin{aligned}
&\mathbb{E}\left[B_T \cdot \int_0^T \nabla \mathcal{U}_t^{T,\psi^T}(X_t^T)\,\mathrm{d}t\right] \\
&\quad = T\,\mathbb{E}\left[B_T \cdot \nabla \mathcal{U}_0^{T,\psi^T}(X_0^T)\right] + \mathbb{E}\left[B_T \cdot \int_0^T \int_0^t \nabla^2 \mathcal{U}_s^{T,\psi^T}(X_s^T)\,\mathrm{d}B_s\,\mathrm{d}t\right] \\
&\quad = \int_0^T \mathbb{E}\left[(B_T - B_t)\cdot \int_0^t \nabla^2 \mathcal{U}_s^{T,\psi^T}(X_s^T)\,\mathrm{d}B_s\right]\mathrm{d}t \\
&\qquad\qquad + \int_0^T \mathbb{E}\left[\int_0^t \mathrm{d}B_s \cdot \int_0^t \nabla^2 \mathcal{U}_s^{T,\psi^T}(X_s^T)\,\mathrm{d}B_s\right]\mathrm{d}t \\
&\quad = \int_0^T \int_0^t \mathbb{E}\left[\mathrm{Tr}(\nabla^2 \mathcal{U}_s^{T,\psi^T}(X_s^T))\right]\mathrm{d}s\,\mathrm{d}t
\end{aligned}$$

where the last step follows from the independence between the Brownian increment $B_T - B_t$ and $(\nabla^2 \mathcal{U}_s^{T,\psi^T}(X_s^T))_{s\in[0,t]}$ and Ito Isometry. Therefore, so far we have shown that

$$\begin{aligned}
&\int |x-y|^2\,\mathrm{d}\boldsymbol{\pi}^T \\
&= T + \mathbb{E}\left[\left|\int_0^T \nabla \mathcal{U}_t^{T,\psi^T}(X_t^T)\,\mathrm{d}t\right|^2\right] - 2\int_0^T \int_0^t \mathbb{E}\left[\mathrm{Tr}(\nabla^2 \mathcal{U}_s^{T,\psi^T}(X_s^T))\right]\mathrm{d}s\,\mathrm{d}t\,.
\end{aligned} \tag{3.2.16}$$

Now, observe that Jensen's inequality implies that

$$\begin{aligned}
\mathbb{E}\left[\left|\int_0^T \nabla \mathcal{U}_t^{T,\psi^T}(X_t^T)\,\mathrm{d}t\right|^2\right] \le T\,\mathbb{E}\left[\int_0^T |\nabla \mathcal{U}_t^{T,\psi^T}(X_t^T)|^2\,\mathrm{d}t\right] &= 2\,T\,\mathscr{H}(\mathbf{P}^T|\mathrm{R}_\mu^T) \\
&= 2\,T\,\mathscr{H}(\boldsymbol{\pi}^T|\mu\otimes \mathsf{p}_T)
\end{aligned} \tag{3.2.17}$$

with $\mathrm{R}_\mu^T \in \mathcal{P}(\Omega)$ being the law of a Brownian motion started with distribution μ, $\mathsf{p}_T(\cdot,\cdot)$ being the corresponding heat kernel and hence $\mu\otimes\mathsf{p}_T = (X_0, X_T)_{\#}\mathrm{R}_\mu^T \in \mathcal{P}(\mathbb{R}^d\times\mathbb{R}^d)$. Then we may rewrite the above right hand side as

$$\begin{aligned}
&2\,T\,\mathscr{H}(\boldsymbol{\pi}^T|\mu\otimes\mathsf{p}_T) \\
&\quad = 2\,T\,\mathscr{H}(\boldsymbol{\pi}^T|\mu\otimes\nu) + 2\,T\,\mathscr{H}(\mu\otimes\nu|\mu\otimes\mathrm{Leb}) - 2\,T\int \log\frac{\mathrm{d}(\mu\otimes\mathsf{p}_T)}{\mathrm{d}(\mu\otimes\nu)}\,\mathrm{d}\boldsymbol{\pi}^T \\
&\quad = 2\,T\,\mathscr{H}(\boldsymbol{\pi}^T|\mu\otimes\nu) + 2\,T\,\mathrm{Ent}(\nu) - 2\,T\int \log \mathsf{p}_T(x,y)\,\mathrm{d}\boldsymbol{\pi}^T(x,y) \\
&\quad = 2\,T\,\mathscr{H}(\boldsymbol{\pi}^T|\mu\otimes\nu) + 2\,T\,\mathrm{Ent}(\nu) - d\,T\,\log\left(\frac{1}{2\pi T}\right) + \int |x-y|^2\,\mathrm{d}\boldsymbol{\pi}^T \\
&\quad = \theta\,\mathrm{EOT}_{2T/\theta}^{\mathsf{d}^2/\theta}(\mu,\nu) + 2\,T\,\mathrm{Ent}(\nu) - d\,T\,\log\left(\frac{1}{2\pi T}\right),
\end{aligned} \tag{3.2.18}$$

where $\theta > 0$ is some fixed parameter (to be determined later) whereas the term $\mathrm{EOT}^{\mathrm{d}^2/\theta}_{2T/\theta}(\mu,\nu)$ is the entropic cost as defined in (2.A.1) when considering the cost function $c(x,y) = |x-y|^2/\theta$. Then, the above algebraic manipulation allows us to rely on [EN22a, Theorem 3.8] that guarantees that it holds

$$\mathrm{EOT}^{\mathrm{d}^2/\theta}_{2T/\theta}(\mu,\nu) - \frac{1}{\theta}\,\mathbf{W}_2^2(\mu,\nu) \le \frac{d\,T}{\theta}\,\log\left(\frac{\theta}{2T}\right) + T\,\frac{32\,C_\mu}{\theta^2}\sqrt{1+\beta_\mu/\alpha_\nu}\,, \tag{3.2.19}$$

where $C_\mu > 0$ is the quantisation constant as introduced in (3.2.13).

By combining (3.2.17) and (3.2.18) with (3.2.19) we get

$$\begin{aligned}\mathbb{E}&\left[\left|\int_0^T \nabla\mathcal{U}_t^{T,\psi^T}(X_t^T)\,\mathrm{d}t\right|^2\right]\\ &\qquad\le \mathbf{W}_2^2(\mu,\nu) + T\left[2\,\mathrm{Ent}(\nu) + d\,\log(\theta\pi) + \frac{32\,C_\mu}{\theta}\sqrt{1+\beta_\mu/\alpha_\nu}\right],\end{aligned}$$

and from (3.2.16), by minimising over $\theta > 0$ we conclude that

$$\begin{aligned}\int |x-y|^2\,\mathrm{d}\boldsymbol{\pi}^T - \mathbf{W}_2^2(\mu,\nu) &\le T\left[2\,\mathrm{Ent}(\nu) + d\,\log\left(\frac{32\pi\,C_\mu}{d}\sqrt{1+\frac{\beta_\mu}{\alpha_\nu}}\right) + d\right]\\ &\qquad -2\int_0^T\int_0^t \mathbb{E}\left[\mathrm{Tr}(\nabla^2\mathcal{U}_s^{T,\psi^T}(X_s^T))\right]\mathrm{d}s\,\mathrm{d}t\,.\end{aligned}$$

Now, recall that Theorem 2.2.5 states that for any $t \in [0,T]$ the following lower-bound holds

$$\begin{aligned}\nabla^2\mathcal{U}_t^{T,\psi^T} &\ge \frac{\alpha_{\psi^T}}{1+(T-t)\alpha_{\psi^T}}\\ \text{with } \alpha_{\psi^T} &= \begin{cases}\frac{1}{2}\left(\alpha_\nu + \sqrt{\alpha_\nu^2 + 4\alpha_\nu/(\beta_\mu\,T^2)}\right) - T^{-1} & \text{if } \beta_\mu < +\infty\,,\\ \alpha_\nu - T^{-1} & \text{for } \beta_\mu = +\infty\,.\end{cases}\end{aligned} \tag{3.2.20}$$

Therefore if $\alpha_\nu \ge \beta_\mu$ (which forces the OT map $\boldsymbol{\mathcal{T}}$ to be 1-Lipschitz), then α_{ψ^T} is non-negative and henceforth $\nabla^2\mathcal{U}_t^{T,\psi^T} \ge 0$ for any $t \in [0,T]$. This is enough to conclude the proof when $\alpha_\nu \ge \beta_\mu$.

On the contrary when $\alpha_\nu < \beta_\mu$, then eventually for T small enough (*i.e.*, for $T < \alpha_\nu^{-1} - \beta_\mu^{-1}$) we have $\alpha_{\psi^T} < 0$. In this case(3.2.20) gives uniformly in $x \in \mathbb{R}^d$

$$\mathrm{Tr}(\nabla^2\mathcal{U}_t^{T,\psi^T}(x)) \ge \frac{d\,\alpha_{\psi^T}}{1+(T-t)\alpha_{\psi^T}}$$

from which we may deduce that

$$
\begin{aligned}
-2\int_0^T\int_0^t \mathbb{E}\Big[\operatorname{Tr}(\nabla^2\mathcal{U}_s^{T,\psi^T}(X_s^T))\Big]\mathrm{d}s\,\mathrm{d}t &\le -2\,d\int_0^T\int_0^t \frac{\alpha_{\psi^T}}{1+(T-s)\alpha_{\psi^T}}\,\mathrm{d}s\,\mathrm{d}t\\
&= -2\,d\,T\,\log(1+T\,\alpha_{\psi^T})+2\,d\int_0^T\log(1+(T-t)\alpha_{\psi^T})\,\mathrm{d}t\\
&\le -2\,d\,T\,\log(1+T\,\alpha_{\psi^T}) = -2\,d\,T\,\log\Big(\frac{T\,\alpha_\nu}{2}+\sqrt{\frac{T^2\,\alpha_\nu^2}{4}+\frac{\alpha_\nu}{\beta_\mu}}\Big)\\
&\le \begin{cases} 2\,d\,T\,\log\sqrt{\frac{\beta_\mu}{\alpha_\nu}} = d\,T\,\log(\beta_\mu/\alpha_\nu) & \text{if } \beta_\mu<+\infty\,,\\ 2\,d\,T\log(1/(T\alpha_\nu)) & \text{if } \beta_\mu=+\infty\,.\end{cases}
\end{aligned}
$$

This concludes the proof. □

Remark 3.2.7. *En passant, by combining* (3.2.18), (3.2.19) *and again minimising over* $\theta>0$ *we have actually shown that*

$$
\begin{aligned}
T\,\mathcal{C}_T(\mu,\nu)-\frac{1}{2}\mathbf{W}_2^2(\mu,\nu) &= T\,\mathscr{H}(\mathbf{P}^T|\mathrm{R}_{0,T})-\frac{1}{2}\mathbf{W}_2^2(\mu,\nu)\\
&= T\,\mathrm{Ent}(\mu)+T\,\mathscr{H}(\mathbf{P}^T|\mathrm{R}_\mu^T)-\frac{1}{2}\mathbf{W}_2^2(\mu,\nu)\\
&\le T\Big[\mathrm{Ent}(\mu)+\mathrm{Ent}(\nu)+\frac{d}{2}\log\Big(\frac{32\pi\,C_\mu}{d}\sqrt{1+\frac{\beta_\mu}{\alpha_\nu}}\Big)+\frac{d}{2}\Big]\,,
\end{aligned}
$$

which is a first quantitative version of the zero-th order convergence (3.2.2).

This shows that the mismatch between the Schrödinger cost and the (squared) Wasserstein distance is of order T*, unlike what happens for the EOT cost where the convergence is slower (of order* $T\log(1/T)$*, cf. [EN22a]). Nevertheless, in the previous theorem we have shown that the suboptimality of the SP/EOT plan* $\boldsymbol{\pi}^T$ *w.r.t. the OT problem once again is of order* T *(at least when* $\beta<+\infty$*).*

We would like to stress that our proof strategy clearly explains this difference between orders of convergence. Indeed, the correct order of convergence is T*, that is the one that captures the suboptimality of* $\boldsymbol{\pi}^T$*. Then, when applying the quantisation estimates [EN22a], we get an extra factor of the order* $T\,d\,\log(1/T)$ *which comes from the fact that* $\boldsymbol{\pi}^0$ *satisfies* ($\mathrm{quant}_2(C,\alpha)$) *with rate* $\alpha=1/d$*. This latter term is slower and dominates the convergence rate of EOT towards OT.*

Remarkably, when considering the Schrödinger cost the extra factor $T\,d\,\log(1/T)$ *coming from the quantisation perfectly matches the normalising constant coming from the partition function of the Gaussian (cf.* (1.2.5)*), which is missing in the EOT formulation.*

This suggests that even for more general EOT problems the convergence rate obtained in [EN22a] is affected from the normalising constant of the reference measure $\mathrm{R}\propto\exp(-c/\varepsilon)$ *and that therefore it might be more interesting analysing the suboptimality rate for the optimal entropic plan, instead of focusing solely on the costs.*

We are finally ready to prove the quantitative convergence of the gradients of Schrödinger potentials. Our approach relies on an explicit expression available in the Brownian motion case that reads as

$$\nabla \log P_T e^{-\psi^T}(x) = T^{-1} \int (y-x)\, \pi_T^{x,\psi^T}(\mathrm{d}y) \tag{3.2.21}$$

where

$$\pi_T^{x,\psi^T}(\mathrm{d}y) \propto \exp\Big(-|x-y|^2/_{2T} - \psi^T(y)\Big)\,\mathrm{d}y$$

is the conditional law of the Schrödinger plan $\boldsymbol{\pi}^T$, i.e.

$$\boldsymbol{\pi}^T(\mathrm{d}x\mathrm{d}y) = \mu(\mathrm{d}x) \otimes \pi_T^{x,\psi^T}(\mathrm{d}y)\,. \tag{3.2.22}$$

We will provide a proof of (3.2.21) (under some additional regularity assumptions in Proposition 6.A.2 in Chapter 6). For now, we refer the reader to [Con24, Proposition 5.2] where the above identity is proven by further assuming that

$$\exists\, \gamma,\, \varepsilon > 0 \quad \text{s.t.} \quad \int \exp(\gamma |x|^{1+\varepsilon})\mathrm{d}\mu < +\infty\,; \tag{3.2.23}$$

assumption required there in order to justify the differentiation under integral sign. Notice that (3.2.23) is met for a wide class of marginals (e.g., asymptotically log-concave marginals, cf. Lemma 6.A.1) and that this further guarantees $\mu \in \mathcal{P}_{2+\delta}$. Let us also mention here that $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ satisfying **A**2 with $\beta_\mu < +\infty$ guarantees the Fisher information $\mathcal{I}(\mu)$ to be finite. Therefore the assumptions of Theorem 3.2.3 are met in what follows.

Theorem 3.2.8. *Assume A1, A2 with $\beta_\mu < +\infty$ and further assume* (3.2.23)*.*

$$\begin{aligned}
&\Big\|(-T\nabla\varphi^T) - \nabla\varphi^0\Big\|^2_{\mathrm{L}^2(\mu)} \\
&\qquad \le \begin{cases} T\sqrt{\frac{\beta_\mu}{\alpha_\nu}}\left[2\,\mathrm{Ent}(\nu) + d\,\log\Big(\frac{32\pi\, C_\mu}{d}\sqrt{1+\frac{\beta_\mu}{\alpha_\nu}}\Big) + d\right] & \text{if } \beta_\mu \le \alpha_\nu, \\ T\sqrt{\frac{\beta_\mu}{\alpha_\nu}}\left[2\,\mathrm{Ent}(\nu) + d\,\log\Big(\frac{32\pi\, C_\mu}{d}\,\frac{\beta_\mu}{\alpha_\nu}\,\sqrt{1+\frac{\beta_\mu}{\alpha_\nu}}\Big) + d\right] & \text{if } \alpha_\nu < \beta_\mu\,. \end{cases}
\end{aligned}$$

Proof. Fix $S \in (0,T)$ and let φ^S, ψ^S be the Schrödinger potentials associated to the SP problem with time horizon $[0,S]$. By combining (2.2.3) with (3.2.21) we know that for any $x \in \mathrm{supp}(\mu)$ it holds

$$\begin{aligned}
|T\nabla\varphi^T - S\nabla\varphi^S|^2(x) &= |T\,\nabla\log P_T\exp(-\psi^T) - S\,\nabla\log P_S\exp(-\psi^S)|^2(x) \\
&= \left|\int y\,\mathrm{d}\pi_T^{x,\psi^T} - \int y\,\mathrm{d}\pi_S^{x,\psi^S}\right|^2 \le \mathbf{W}_2^2(\pi_T^{x,\psi^T},\, \pi_S^{x,\psi^S}) \\
&\le \frac{4}{\alpha_\nu + \sqrt{\alpha_\nu^2 + 4\alpha_\nu/(\beta_\mu\, S^2)}}\,\mathscr{H}(\pi_T^{x,\psi^T}|\pi_S^{x,\psi^S})
\end{aligned}$$

where in the last step we have applied the Talagrand transportation cost inequality [BGL13, Corollary 9.3.2], which holds since Theorem 2.2.5 implies

$$\nabla^2(-\log \pi_S^{x,\psi^S}(y)) = \nabla_y^2(|x-y|^2/_{2S} + \psi^S(y)) \geq \frac{1}{2}\left(\alpha_\nu + \sqrt{\alpha_\nu^2 + 4\alpha_\nu/(\beta_\mu\, S^2)}\right).$$

By integrating over μ, recalling (3.2.22) and by applying the additive property of relative entropy (1.A.4), we deduce that

$$\begin{aligned}
&\|T\nabla\varphi^T - S\nabla\varphi^S\|^2_{\mathrm{L}^2(\mu)} \\
&\quad \leq \frac{4}{\alpha_\nu + \sqrt{\alpha_\nu^2 + 4\alpha_\nu/(\beta_\mu\, S^2)}}\left[\mathscr{H}(\mu|\mu) + \int \mathscr{H}(\pi_\varepsilon^{x,\psi^\varepsilon}|\pi_S^{x,\psi^S})\,\mathrm{d}\mu(x)\right] \\
&\quad = \frac{2\sqrt{\beta_\mu/\alpha_\nu}}{\frac{\sqrt{\beta_\mu\,\alpha_\nu}\,S}{2} + \sqrt{\frac{\alpha_\nu\,\beta_\mu\,S^2}{4}+1}}\, S\,\mathscr{H}(\boldsymbol{\pi}^T|\boldsymbol{\pi}^S) \leq 2\sqrt{\beta_\mu/\alpha_\nu}\, S\,\mathscr{H}(\boldsymbol{\pi}^T|\boldsymbol{\pi}^S)\,.
\end{aligned}$$

Then, from Corollary 2.2.3 (Pythagoras Theorem for entropic projections) we deduce that

$$\begin{aligned}
\|T\nabla\varphi^T - S\nabla\varphi^S\|^2_{\mathrm{L}^2(\mu)} &\leq 2\sqrt{\beta_\mu/\alpha_\nu}\, S\,\mathscr{H}(\boldsymbol{\pi}^T|\boldsymbol{\pi}^S) \\
&= 2\sqrt{\beta_\mu/\alpha_\nu}\left(S\,\mathscr{H}(\boldsymbol{\pi}^T|\mathrm{R}_{0,S}) - S\,\mathscr{H}(\boldsymbol{\pi}^S|\mathrm{R}_{0,S})\right) \\
&\overset{(1.2.5)}{=} 2\,S\sqrt{\beta_\mu/\alpha_\nu}\left(\mathscr{H}(\boldsymbol{\pi}^T|\mu\otimes\nu) + \mathrm{Ent}(\mu) + \mathrm{Ent}(\nu) + \frac{d}{2}\log(2\pi S)\right) \\
&\qquad + \sqrt{\beta_\mu/\alpha_\nu}\left(\int |x-y|^2\,\mathrm{d}\boldsymbol{\pi}^T - 2\,S\,\mathcal{C}_S(\mu,\nu)\right).
\end{aligned}$$

Letting $S \downarrow 0$, thanks to Theorem 3.2.3, by recalling that the Schrödinger cost converges to the Wasserstein distance (cf. (3.2.2) paying attention to the different scaling, or also Remark 3.2.7) we finally deduce that

$$\|(-T\nabla\varphi^T) - \nabla\varphi^0\|^2_{\mathrm{L}^2(\mu)} \leq \sqrt{\beta_\mu/\alpha_\nu}\left(\int |x-y|^2\,\mathrm{d}\boldsymbol{\pi}^T - \mathbf{W}_2^2(\mu,\nu)\right).$$

This conclude our proof since it is enough combining the above estimate with the suboptimality estimate proven in Theorem 3.2.6. □

Bibliographical Remarks

For both SP and EOT the small-time and small-noise limits have been extensively studied in the literature. For what concerns the convergence of the entropic cost to the squared Wasserstein distance, the first result is that of [Mik04], eventually generalised in [Léo12a]. As a further step in the analysis of this convergence, in [ADPZ11] (in dimension one), in [EMR15] (in greater dimensions) and most recently in [Pal19], the first-order Taylor expansion of the entropic cost has been investigated showing that the first-order coefficient is given by the sum of the relative entropies of the marginals. Under stronger assumptions on the marginals, [CT21] determined the second-order Taylor expansion of the entropic cost showing that the second-order term is related to the average of the Fisher information along the geodesic between the marginals (an analogous result has also been obtained slightly later in [CRL$^+$20]). Most recently, in [CPT23] the authors have analysed the small-noise limit in the EOT setting under very general assumptions on the cost function c (which allow for non-uniqueness of the optimal transport coupling for instance). Particularly, in the small-noise limit, they have compared the EOT cost with the OT cost, showing that in general this error is of the order $\varepsilon \log(\varepsilon^{-1})$ and that this lower bound is sharp. These results were obtained for compact marginals in [CPT23] and then they have been further generalised in [EN22b] for unbounded settings.

Alongside this line of work, in [Mik04] and [BGN22] it is shown that in the small-time and small-noise limits the optimal solutions to SP and EOT respectively converge to the optimal coupling of the OT problem. Particularly, [BGN22] show that EOT and SP plans satisfy a Large Deviation Principle, result that has been recently generalised on the path space to Schrödinger bridges in [Kat24] where the author further establishes exponential continuity for Brownian bridges.

When it comes to the convergence of dual optimiser, in the EOT setting it is proven in [NW22] that the entropic potentials $(\varphi_\varepsilon, \psi_\varepsilon)$ converge to the Kantorovich potentials (when the latter are unique) associated to the Monge-Kantorovich problem with cost c

$$\sup_{\varphi \in L^1(\mu),\, \psi \in L^1(\nu)\,:\, \varphi \oplus \psi \leq c} \left(\int \varphi \, \mathrm{d}\mu + \int \psi \, \mathrm{d}\nu \right) .$$

More precisely, in [NW22] the authors prove in the small-noise limit $\varepsilon \downarrow 0$ that the sequences $\{\varphi_\varepsilon\}_{\varepsilon>0}$ and $\{\psi_\varepsilon\}_{\varepsilon>0}$ are (strongly) compact in $\mathrm{L}^1(\mu)$ and $\mathrm{L}^1(\nu)$ respectively, and that their accumulation points are optimiser of the Monge-Kantorovich problem. Adapting their proof strategy and relying on [Nor97], we have shown under a curvature-dimension condition a similar convergence statement for SP in Lemma 3.2.1. Heuristically speaking, this follows from the connection between EOT and SP given by taking $\varepsilon = T$ and $c = -T \log \frac{\mathrm{d}\mathrm{R}_{0,T}}{\mathrm{d}(\mu \otimes \nu)}$.

Regarding the convergence of the gradients of the Schrödinger potentials, to the best of our knowledge Theorem 3.2.3 is a novelty in the setting we are dealing with. Indeed, closely related results have been obtained in [PNW21], but in a more restrictive setting (the quadratic Euclidean EOT problem) and under strong regularity assumptions. Namely, uniform bounds on the Hessian of the Kantorovich potential φ^0 are required to prove a modified version of the aforementioned convergence. Furthermore, the authors work in a very specific setting where the marginals μ, ν are compactly supported and with densities globally bounded away from 0 and ∞ on their supports. Moreover they show the convergence of the gradients of the potentials associated to a modified EOT where μ and ν are replaced by empirical measures associated to an n-sample; therefore they take a regularisation parameter ε depending on the batch size n and then consider the limit $n \to \infty$. In the very recent work [MS23] finally appeared a quantitative convergence result close to our Theorem 3.2.8, where indeed the authors study the suboptimality of the Schrödinger plan $\boldsymbol{\pi}^T$ w.r.t. the classical OT quadratic problem as we did in Theorem 3.2.6.

Almost all the results presented in this chapter are based on the published paper [CCGT23], with the exception of the suboptimality and quantitative convergence theorems of Section 3.2.3. Indeed, these are results I got during my third year of PhD, but that never got published since [MS23] got similar results concurrently. The two approaches are different (the one presented in Section 3.2.3 being more stochastic in nature), however [MS23] got better convergence results (in Theorem 3.2.6 we get a rate of the order $T\log(1/T)$ when $\beta_\mu = +\infty$, whereas they always manage to get a rate of order T).

Chapter 4

Quantitative stability for the Schrödinger problem

In this chapter we are interested in explicitly quantifying how sensitive are the optimal costs and optimal plans in SP to variations of the marginal constraints. For notations' clarity we will explicitly stress out the dependence of the optimal Schrödinger coupling with respect to the couple of marginals and denote it with $\pi^{\mu\to\nu,T}$. Similarly, (f,g) will denote the fg-decomposition for $\mathcal{C}_T(\mu,\nu)$ (cf. Theorem 2.2.1) while $(\bar f,\bar g)$ will stand for the decomposition associated to $\mathcal{C}_T(\bar\mu,\bar\nu)$; we further implicitly assume that both are normalised according to (2.2.13). Lastly, we will denote with ρ, σ the densities of μ, ν w.r.t. $\mathfrak{m}$, whereas $\bar\rho$, $\bar\sigma$ will be the densities of $\bar\mu$, $\bar\nu$.

We are going to show (cf. Theorem 4.2.2) that the curvature-dimension condition (CD) implies a rather general and explicit stability result in terms of the symmetric relative entropy, i.e.

$$\mathscr{H}^{\mathrm{sym}}(\mu,\bar\mu) := \mathscr{H}(\mu|\bar\mu) + \mathscr{H}(\bar\mu|\mu)\,, \tag{4.0.1}$$

and in terms of a negative-order weighted homogeneous Sobolev norm, which is defined for any signed measure ν as follows:

$$\|\nu\|_{\dot{\mathrm{H}}^{-1}(\mu)} := \sup\Big\{ |\langle h,\nu\rangle| \,:\, \|h\|_{\dot{\mathrm{H}}^1(\mu)} \le 1 \Big\}\,, \quad \text{where } \|h\|^2_{\dot{\mathrm{H}}^1(\mu)} := \int |\nabla h|^2 \,\mathrm{d}\mu\,.$$

This dual norm on the space of signed measures encodes the linearised behaviour of the Wasserstein distance $\mathbf{W}_2$ for infinitesimal perturbations (see e.g. [Pey18] and references therein). For instance, if $\mu \in \mathcal{P}_2(\mathbb{R}^d)$, $\mu \ll$ Leb and $\bar\mu^\varepsilon = (1+\varepsilon h)\mu$ for some $h \in \mathrm{L}^\infty(\mu)$ with $\int_{\mathbb{R}^d} h\mathrm{d}\mu = 0$, then [Vil03, Theorem 7.26] implies

$$\|\mu - \bar\mu^\varepsilon\|_{\dot{\mathrm{H}}^{-1}(\mu)} = \varepsilon\, \|h\,\mu\|_{\dot{\mathrm{H}}^{-1}(\mu)} \quad \text{and} \quad \|h\,\mu\|_{\dot{\mathrm{H}}^{-1}(\mu)} \le \liminf_{\varepsilon\to 0} \frac{\mathbf{W}_2(\mu,\bar\mu^\varepsilon)}{\varepsilon}\,.$$

Moreover, [Pey18] provides non-asymptotic comparisons between this Sobolev norm and the Wasserstein distance. In particular, it always holds $\mathbf{W}_2(\mu,\bar{\mu}) \leq 2\,\|\mu-\bar{\mu}\|_{\dot{\mathrm{H}}^{-1}(\mu)}$, and when $(M,\mathsf{d},\mathrm{vol})$ has non-negative Ricci curvature and if the densities $\frac{\mathrm{d}\mu}{\mathrm{dvol}}$ and $\frac{\mathrm{d}\bar{\mu}}{\mathrm{dvol}}$ are bounded away from 0 and ∞, then the norm of the difference $\|\mu-\bar{\mu}\|_{\dot{\mathrm{H}}^{-1}(\mu)}$ is equivalent to the Wasserstein distance $\mathbf{W}_2(\mu,\bar{\mu})$.

Apart from the curvature-dimension condition (CD) and from **A**1 and **A**3, we further assume the following integrability condition on the reference measure:

$$\exists\, r>0: \qquad \int \mathfrak{m}\left(B_{\sqrt{T}}(x)\right)^{rT} e^{r\,\mathsf{d}^2(x,z)}\mathrm{d}\mathfrak{m}(x) < +\infty \quad \forall z\in M\,. \tag{I}$$

The reason why we assume the above condition is that our computations will heavily rely on integrating the Schrödinger potentials of one problem against the marginal constraints of the other, which requires enough integrability for the former.

Notice that if $\mathfrak{m}$ is a probability measure ($\mathfrak{m}(M)=1$), the previous assumption is met as soon as

$$e^{r\,\mathsf{d}^2(x,z_0)} \in \mathrm{L}^1(\mathfrak{m}) \quad \text{for some r} > 0 \text{ and } \mathrm{z}_0 \in \mathrm{M}\,, \tag{4.0.2}$$

which is true for instance under a positive curvature condition $\mathsf{CD}(\kappa,\infty)$ with $\kappa>0$. Indeed the latter implies a logarithmic Sobolev inequality with parameter κ^{-1} [BGL13, Corollary 5.7.1] and then by means of Herbst's argument [BGL13, Proposition 5.4.1] we get $e^{r\,\mathsf{d}^2(x,z_0)}\in\mathrm{L}^1(\mathfrak{m})$ for any $r<\frac{\kappa}{2}$. Moreover, as soon as $\mathfrak{m}$ is a probability measure the above condition (I) is time-independent and therefore it allows to consider the small-time limit for SP.

Condition (4.0.2) should also be compared with Conditions (6.8) and (6.9) in [Nut21] in the EOT setting, by considering the cost function $c(x,y)=-T\log\mathsf{p}_T$, combined with a Gaussian heat kernel lower bound (cf. (2.2.6) and (2.2.7)). There, the authors are interested in getting uniform bounds on the Schrödinger potentials along Sinkhorn's iterates. However let us stress out that [Nut21] requires $e^{r\,\mathsf{d}^2(x,y)}\in\mathrm{L}^1(\mu\otimes\nu)$ for some $r>\frac{1}{T}$, which in particular does not suit the most interesting EOT regime, i.e. the small-noise (or equivalently small-time) limit. On the contrary, our condition only requires $r>0$ independently from the time-window $[0,T]$ and moreover we are able to pass the integrability assumption on the equilibrium measure $\mathfrak{m}$. This suits more the *stability setting* since we would like to keep the assumptions on the marginals as light as possible. We should further mention here that in the most recent work [NW23] the authors manage to prove the uniform integrability of the potentials along Sinkhorn's algorithm by solely requiring $e^{r\,\mathsf{d}^2(x,y)}\in\mathrm{L}^1(\mu\otimes\nu)$ for some $r>0$ overcoming the small-noise issue present in [Nut21]. In conclusion, it is not surprising that our stability results require (I), since we need enough integrability for the potentials, which is analogous to requiring $e^{r\,\mathsf{d}^2(x,y)}\in\mathrm{L}^1(\mu\otimes\nu)$ in the EOT setting.

4.1 A log-integrability Lemma

Let us start by recalling a few interesting facts about Orlicz spaces, from which we will deduce integrability results for our potentials.

Let $\mathfrak{q}$ be a probability measure on a measurable space (Ω, Σ) and consider the Young functions

$$\theta(t) := e^t - 1\,, \qquad \theta^*(s) = \begin{cases} s\,\log s - s + 1 & \text{if } s > 0, \\ 1 & \text{if } s = 0\,. \end{cases}$$

where $\theta^*(s) := \sup_{t\in\mathbb{R}}\{st - \theta(t)\}$. The Orlicz space associated to the Young function θ, denoted by $\mathrm{L}_\theta(\mathfrak{q})$, is defined as the space of measurable functions $f : \Omega \to \mathbb{R}$ with finite Luxemburg norm

$$\|f\|_\theta := \inf\left\{ b > 0\colon \int \theta\left(\frac{|f|}{b}\right) \mathrm{d}\mathfrak{q} \le 1 \right\}\,.$$

To be precise we should make explicit reference to the underlying probability $\mathfrak{q}$, when talking about Luxemburg norms; however we will omit such reference when it is clear from the context.

When dealing with finite relative entropies, the Orlicz space associated to θ^* (defined analogously as for θ) plays a natural role. Indeed, for any probability $\mathfrak{p}$ with $\mathscr{H}(\mathfrak{p}|\mathfrak{q}) < +\infty$ we have

$$\begin{aligned} \left\|\frac{\mathrm{d}\mathfrak{p}}{\mathrm{d}\mathfrak{q}}\right\|_{\theta^*} &= \inf\left\{ b > 0\colon \int \theta^*\left(\frac{1}{b}\frac{\mathrm{d}\mathfrak{p}}{\mathrm{d}\mathfrak{q}}\right) \mathrm{d}\mathfrak{q} \le 1 \right\} \\ &= \inf\left\{ b > 0\colon \frac{1}{b}\mathscr{H}(\mathfrak{p}|\mathfrak{q}) - \frac{1}{b}\,(1 + \log b) + 1 \le 1 \right\} \\ &= \inf\left\{ b > 0\colon \mathscr{H}(\mathfrak{p}|\mathfrak{q}) - 1 \le \log b \right\} = e^{\mathscr{H}(\mathfrak{p}|\mathfrak{q}) - 1}\,. \end{aligned}$$

Finally, let us recall that for any $f \in \mathrm{L}_\theta(\mathfrak{q})$ and $g \in \mathrm{L}_{\theta^*}(\mathfrak{q})$ we have

$$\int |fg|\,\mathrm{d}\mathfrak{q} \le 2\,\|f\|_\theta\,\|g\|_{\theta^*}\,, \tag{4.1.1}$$

which is a consequence of the trivial inequality

$$st \le \theta(t) + \theta^*(s) \quad \forall s, t > 0\,.$$

After this digression, let us prove the following lemma, which is pivotal in our reasoning since it allows to deduce the log-integrability of any positive measurable function under the only hypotheses of finite relative entropy and some positive and negative integrability of the same function.

Lemma 4.1.1. *Let h be a positive measurable function and $\mathfrak{p}, \mathfrak{q} \in \mathcal{P}(\Omega)$ such that $\mathscr{H}(\mathfrak{p}|\mathfrak{q}) < \infty$. If there exist p, $q > 0$ such that $h \in \mathrm{L}^q(\mathfrak{q})$ and $h^{-1} \in \mathrm{L}^p(\mathfrak{q})$, then*

$\log h \in \mathrm{L}^1(\mathfrak{p})$. *More precisely, it holds*

$$\int |\log h| \,\mathrm{d}\mathfrak{p} \leq 2\, e^{\mathscr{H}(\mathfrak{p}|\mathfrak{q})-1} \left[\frac{1}{1 \wedge p \wedge q} \vee \log_2 \Big(\mathfrak{q}\{h \geq 1\}^{1-\frac{1}{q}} \, \|h\|_{\mathrm{L}^q(\mathfrak{q})} + \mathfrak{q}\{h < 1\}^{1-\frac{1}{p}} \left\| h^{-1} \right\|_{\mathrm{L}^p(\mathfrak{q})} \Big) \right], \tag{4.1.2}$$

and also

$$\int |\log h| \,\mathrm{d}\mathfrak{p} \leq \frac{2\, e^{\mathscr{H}(\mathfrak{p}|\mathfrak{q})-1}}{p \wedge q} \left[1 \vee \log_2 \left(\|h\|_{\mathrm{L}^q(\mathfrak{q})}^{p\wedge q} + \left\| h^{-1} \right\|_{\mathrm{L}^p(\mathfrak{q})}^{p \wedge q} \right) \right]. \tag{4.1.3}$$

Proof. By means of the Orlicz-Young inequality (4.1.1) we have

$$\int |\log h| \,\mathrm{d}\mathfrak{p} = \int \left| \frac{\mathrm{d}\mathfrak{p}}{\mathrm{d}\mathfrak{q}} \log h \right| \mathrm{d}\mathfrak{q} \leq 2 \, \|\log h\|_\theta \, \left\| \frac{\mathrm{d}\mathfrak{p}}{\mathrm{d}\mathfrak{q}} \right\|_{\theta^*} = 2 \, \|\log h\|_\theta \, e^{\mathscr{H}(\mathfrak{p}|\mathfrak{q})-1} \,. \tag{4.1.4}$$

Therefore it suffices to show that the above Luxemburg norm $\|\log h\|_\theta$ is finite. To this end, notice that for any $b > 0$ we have

$$\int \theta \left(\frac{|\log h|}{b} \right) \mathrm{d}\mathfrak{q} = \int \theta \left(\left| \log h^{\frac{1}{b}} \right| \right) \mathrm{d}\mathfrak{q} = \int_{\{h \geq 1\}} h^{\frac{1}{b}} \mathrm{d}\mathfrak{q} + \int_{\{h<1\}} h^{-\frac{1}{b}} \mathrm{d}\mathfrak{q} - 1 \,. \tag{4.1.5}$$

Before proceeding, let us also point out that the above right-hand side is always non-negative since it is greater or equal to $\mathfrak{q}\{h \geq 1\} + \mathfrak{q}\{h < 1\} - 1 = 0$.

Now consider any parameter $\delta \in (0, \infty)$ (to be fixed later) and fix $b \geq \frac{1}{\delta \wedge p \wedge q}$. For the sake of notation, let $P := \{h \geq 1\}$ and its complementary $N := \{h < 1\}$. In what follows we are going to assume that both $\mathfrak{q}(P) > 0$ and $\mathfrak{q}(N) > 0$; the case $\mathfrak{q}(P) \in \{0, 1\}$ can be treated in the same fashion. Now, introduce the probability measures on these sets induced by $\mathfrak{q}$, *i.e.*

$$\mathrm{d}\mathfrak{q}_{|_P}(x) := \frac{\mathrm{d}\mathfrak{q}(x)}{\mathfrak{q}(P)} \qquad \text{and} \qquad \mathrm{d}\mathfrak{q}_{|_N}(x) := \frac{\mathrm{d}\mathfrak{q}(x)}{\mathfrak{q}(N)} \,.$$

Since $qb \geq 1$, from Jensen's inequality we deduce that

$$\begin{aligned} \int_{\{h \geq 1\}} h^{\frac{1}{b}} \mathrm{d}\mathfrak{q} = \mathfrak{q}(P) \int_P h^{\frac{1}{b}} \mathrm{d}\mathfrak{q}_{|_P} = \mathfrak{q}(P) \int_P h^{q \, \frac{1}{qb}} \mathrm{d}\mathfrak{q}_{|_P} \leq \mathfrak{q}(P) \left(\int_P h^q \mathrm{d}\mathfrak{q}_{|_P} \right)^{\frac{1}{qb}} \\ = \mathfrak{q}(P)^{1-\frac{1}{qb}} \left(\int_P h^q \mathrm{d}\mathfrak{q} \right)^{\frac{1}{qb}} \leq \mathfrak{q}(P)^{1-\frac{1}{qb}} \, \|h\|_{\mathrm{L}^q(\mathfrak{q})}^{\frac{1}{b}} \,. \end{aligned}$$

Similarly, from $pb \geq 1$ we deduce that

$$\begin{aligned}\int_{\{h<1\}} \left(\frac{1}{h}\right)^{\frac{1}{b}} \mathrm{d}\mathfrak{q} = \mathfrak{q}(N) \int_N h^{-\frac{1}{b}} \mathrm{d}\mathfrak{q}_{|N} &= \mathfrak{q}(N) \int_N h^{-p\,\frac{1}{pb}} \mathrm{d}\mathfrak{q}_{|N} \\ \leq \mathfrak{q}(N) \left(\int_N h^{-p} \mathrm{d}\mathfrak{q}_{|N}\right)^{\frac{1}{pb}} &= \mathfrak{q}(N)^{1-\frac{1}{pb}} \left(\int_N h^{-p} \mathrm{d}\mathfrak{q}\right)^{\frac{1}{pb}} \\ \leq \mathfrak{q}(N)^{1-\frac{1}{pb}} \left\|h^{-1}\right\|^{\frac{1}{b}}_{\mathrm{L}^p(\mathfrak{q})} .\end{aligned}$$

For the sake of brevity let $\lambda := \mathfrak{q}(P)$. Since also $\delta b \geq 1$, from (4.1.5) and the concavity of $x^{\frac{1}{\delta b}}$ we deduce

$$\begin{aligned}\int \theta\left(\frac{|\log h|}{b}\right) \mathrm{d}\mathfrak{q} &\leq \lambda^{1-\frac{1}{qb}} \|h\|^{\frac{1}{b}}_{\mathrm{L}^q(\mathfrak{q})} + (1-\lambda)^{1-\frac{1}{pb}} \left\|h^{-1}\right\|^{\frac{1}{b}}_{\mathrm{L}^p(\mathfrak{q})} - 1 \\ &= \lambda \left(\frac{\|h\|^{\delta}_{\mathrm{L}^q(\mathfrak{q})}}{\lambda^{\frac{\delta}{q}}}\right)^{\frac{1}{\delta b}} + (1-\lambda)\left(\frac{\|h^{-1}\|^{\delta}_{\mathrm{L}^p(\mathfrak{q})}}{(1-\lambda)^{\frac{\delta}{p}}}\right)^{\frac{1}{\delta b}} - 1 \\ &\leq \left(\lambda^{1-\frac{\delta}{q}} \|h\|^{\delta}_{\mathrm{L}^q(\mathfrak{q})} + (1-\lambda)^{1-\frac{\delta}{p}} \left\|h^{-1}\right\|^{\delta}_{\mathrm{L}^p(\mathfrak{q})}\right)^{\frac{1}{\delta b}} - 1 \leq 1 ,\end{aligned}$$

where the last inequality holds as soon as

$$b \geq \frac{1}{\delta} \log_2 \left(\lambda^{1-\frac{\delta}{q}} \|h\|^{\delta}_{\mathrm{L}^q(\mathfrak{q})} + (1-\lambda)^{1-\frac{\delta}{p}} \left\|h^{-1}\right\|^{\delta}_{\mathrm{L}^p(\mathfrak{q})}\right) .$$

Therefore, if we take

$$\bar{b} := \frac{1}{\delta \wedge p \wedge q} \vee \frac{1}{\delta} \log_2 \left(\lambda^{1-\frac{\delta}{q}} \|h\|^{\delta}_{\mathrm{L}^q(\mathfrak{q})} + (1-\lambda)^{1-\frac{\delta}{p}} \left\|h^{-1}\right\|^{\delta}_{\mathrm{L}^p(\mathfrak{q})}\right) ,$$

we deduce that $\int \theta\left(\frac{|\log h|}{\bar{b}}\right) \mathrm{d}\mathfrak{q} \leq 1$, and hence

$$\|\log h\|_{\theta} \leq \bar{b} = \frac{1}{\delta \wedge p \wedge q} \vee \frac{1}{\delta} \log_2 \left(\lambda^{1-\frac{\delta}{q}} \|h\|^{\delta}_{\mathrm{L}^q(\mathfrak{q})} + (1-\lambda)^{1-\frac{\delta}{p}} \left\|h^{-1}\right\|^{\delta}_{\mathrm{L}^p(\mathfrak{q})}\right) . \tag{4.1.6}$$

From this last result and (4.1.4) it follows that $\log h \in \mathrm{L}^1(\mathfrak{q})$ and that

$$\begin{aligned}\int |\log h| \, \mathrm{d}\mathfrak{p} \leq 2\, e^{\mathscr{H}(\mathfrak{p}|\mathfrak{q})-1} \Bigg[\frac{1}{\delta \wedge p \wedge q} \vee \frac{1}{\delta} \log_2 \Bigg(&\lambda^{1-\frac{\delta}{q}} \|h\|^{\delta}_{\mathrm{L}^q(\mathfrak{q})} \\ &+ (1-\lambda)^{1-\frac{\delta}{p}} \left\|h^{-1}\right\|^{\delta}_{\mathrm{L}^p(\mathfrak{q})}\Bigg)\Bigg] .\end{aligned}$$

By taking $\delta = 1$ we immediately get (4.1.2). In order to get the second bound, take $\delta := p \wedge q$. Then (4.1.6) reads as

$$\|\log h\|_\theta \le \frac{1}{p \wedge q}\left[1 \vee \log_2\left(\lambda^{1-\frac{p\wedge q}{q}}\|h\|_{\mathrm{L}^q(\mathfrak{q})}^{p\wedge q} + (1-\lambda)^{1-\frac{p\wedge q}{p}}\left\|h^{-1}\right\|_{\mathrm{L}^p(\mathfrak{q})}^{p\wedge q}\right)\right]. \tag{4.1.7}$$

Since $\lambda = \mathfrak{q}(P) \in [0,1]$ we deduce that

$$\begin{aligned}
&\log_2\left(\mathfrak{q}(P)^{1-\frac{p\wedge q}{q}}\|h\|_{\mathrm{L}^q(\mathfrak{q})}^{p\wedge q} + (1-\mathfrak{q}(P))^{1-\frac{p\wedge q}{p}}\left\|h^{-1}\right\|_{\mathrm{L}^p(\mathfrak{q})}^{p\wedge q}\right)\\
&\qquad \le \log_2 \sup_{\lambda\in[0,1]}\left(\lambda^{1-\frac{p\wedge q}{q}}\|h\|_{\mathrm{L}^q(\mathfrak{q})}^{p\wedge q} + (1-\lambda)^{1-\frac{p\wedge q}{p}}\left\|h^{-1}\right\|_{\mathrm{L}^p(\mathfrak{q})}^{p\wedge q}\right)\\
&\qquad = \begin{cases}\log_2\left(\|h\|_{\mathrm{L}^q(\mathfrak{q})}^{p} + \|h^{-1}\|_{\mathrm{L}^p(\mathfrak{q})}^{p}\right) & \text{if } p \le q\\ \log_2\left(\|h\|_{\mathrm{L}^q(\mathfrak{q})}^{q} + \|h^{-1}\|_{\mathrm{L}^p(\mathfrak{q})}^{q}\right) & \text{if } q < p\end{cases}\\
&\qquad = \log_2\left(\|h\|_{\mathrm{L}^q(\mathfrak{q})}^{p\wedge q} + \left\|h^{-1}\right\|_{\mathrm{L}^p(\mathfrak{q})}^{p\wedge q}\right),
\end{aligned}$$

and hence

$$\|\log h\|_\theta \le \frac{1}{p\wedge q}\left[1 \vee \log_2\left(\|h\|_{\mathrm{L}^q(\mathfrak{q})}^{p\wedge q} + \left\|h^{-1}\right\|_{\mathrm{L}^p(\mathfrak{q})}^{p\wedge q}\right)\right].$$

As a byproduct of the above bound and (4.1.4) we get (4.1.3). □

Remark 4.1.2. *When $\mathfrak{q}\{h \ge 1\} \in \{0,1\}$, the bound* (4.1.3) *can be improved. Indeed from* (4.1.7) *we straightforwardly deduce that*

$$\int |\log h|\,\mathrm{d}\mathfrak{p} \le \begin{cases} 2\,e^{\mathscr{H}(\mathfrak{p}|\mathfrak{q})-1}\left(\dfrac{1}{p\wedge q} \vee \log_2\|h\|_{\mathrm{L}^q(\mathfrak{q})}\right) & \text{if } \mathfrak{q}\{h\ge 1\} = 1,\\[2ex] 2\,e^{\mathscr{H}(\mathfrak{p}|\mathfrak{q})-1}\left(\dfrac{1}{p\wedge q} \vee \log_2\left\|h^{-1}\right\|_{\mathrm{L}^p(\mathfrak{q})}\right) & \text{if } \mathfrak{q}\{h\ge 1\} = 0.\end{cases}$$

Corollary 4.1.3. *If $p \wedge q \le 1$ in the previous Lemma, then it holds*

$$\int |\log h|\,\mathrm{d}\mathfrak{p} \le 2\,e^{\mathscr{H}(\mathfrak{p}|\mathfrak{q})-1}\left[\frac{1}{p\wedge q} + \left(\log_2 \frac{\|h\|_{\mathrm{L}^q(\mathfrak{q})} + \|h^{-1}\|_{\mathrm{L}^p(\mathfrak{q})}}{2}\right)^+\right].$$

4.2 From corrector to quantitative stability estimates

In this section we show how from the corrector estimates (*i.e.*, Proposition 3.1.2) it is possible to derive novel quantitative stability estimates for SP.

Before moving to the proof of our main result (Theorem 4.2.2 below) let us give some L^p-bounds (with respect to the reference measure $\mathfrak{m}$) for $(P_T g)^{-1}$ and $(P_T f)^{-1}$, that combined with Lemma 4.1.1 will be pivotal for the validity of all the computations performed hereafter. Moreover, since it will frequently appear in the forthcoming technical bounds, let us recall here that the EOT entropic cost defined at (3.2.5) associated to $c_T(x,y) := -T \log \mathsf{p}_T(x,y)$ equals

$$\mathrm{EOT}_T^{c_T}(\mu,\nu) = T\,\mathcal{C}_T(\mu,\nu) - T\,\mathscr{H}(\mu|\mathfrak{m}) - T\,\mathscr{H}(\nu|\mathfrak{m})\,.$$

Lemma 4.2.1. *Assume* (CD), (I) *and* ***A1***. *Then* $(P_T g)^{-1} \in L^p(\mathfrak{m})$ *and* $(P_T f)^{-1} \in L^p(\mathfrak{m})$ *with* $p = r\,T$. *More precisely, it holds*

$$\left\|(P_T g)^{-1}\right\|_p \leq C_1 \exp\left[\frac{M_2(\nu)}{T} - \frac{\mathrm{EOT}_T^{c_T}(\mu,\nu)}{2T} + C_2\,T\right],$$

$$\left\|(P_T f)^{-1}\right\|_p \leq C_1 \exp\left[\frac{M_2(\mu)}{T} - \frac{\mathrm{EOT}_T^{c_T}(\mu,\nu)}{2T} + C_2\,T\right],$$

where the constants $C_1 > 0$ *and* $C_2 \geq 0$ *do not depend on the marginals* μ,ν *and neither on* f *or* g. *Moreover if* $\kappa \geq 0$ *or if* $\mathsf{CD}(\kappa,\infty)$ *with* $\mathfrak{m}(M) = 1$ *holds, then* $C_2 = 0$.

Proof. We will only prove the first inequality since the proof of the second runs exactly in the same fashion. Firstly, notice that from Jensen's inequality it follows

$$\begin{aligned}
\log P_T g(x) &= \log \int g(y)\mathsf{p}_T(x,y)\mathrm{d}\mathfrak{m}(y) = \log \int g(y)\mathsf{p}_T(x,y)\sigma(y)^{-1}\mathrm{d}\nu(y) \\
&\geq \int \log g\,\mathrm{d}\nu - \int \log \sigma \mathrm{d}\nu + \int \log \mathsf{p}_T(x,y)\,\mathrm{d}\nu(y) \\
&= \frac{1}{2}\left[\mathcal{C}_T(\mu,\nu) - \mathscr{H}(\mu|\mathfrak{m}) - \mathscr{H}(\nu|\mathfrak{m})\right] + \int \log \mathsf{p}_T(x,y)\,\mathrm{d}\nu(y) \\
&= \frac{\mathrm{EOT}_T^{c_T}(\mu,\nu)}{2T} + \int \log \mathsf{p}_T(x,y)\,\mathrm{d}\nu(y)\,,
\end{aligned}$$

where the last step follows from (2.2.13). On the one hand, if $(M, \mathsf{d}, \mathfrak{m})$ satisfies $\mathsf{CD}(\kappa,\infty)$ for some $\kappa \in \mathbb{R}$ and if $\mathfrak{m}(M) = 1$, then from (2.2.6) we deduce the lower bound

$$\log P_T g(x) \geq \frac{\mathrm{EOT}_T^{c_T}(\mu,\nu)}{2T} - \frac{\mathsf{d}^2(x,z_0)}{T} - \frac{M_2(\nu)}{T}\,.$$

Therefore if we take $p = r\,T$ we have, owing to (I),

$$\begin{aligned}\left\|(P_T g)^{-1}\right\|_p^p &= \int (P_T g)^{-p}\mathrm{d}\mathfrak{m} = \int e^{-p\log P_T g}\mathrm{d}\mathfrak{m}\\ &\leq \exp\left[p\left(\frac{M_2(\nu)}{T} - \frac{\mathrm{EOT}_T^{c_T}(\mu,\nu)}{2T}\right)\right]\int e^{\frac{p}{T}\mathsf{d}^2(x,z_0)}\mathrm{d}\mathfrak{m}(x)\\ &\leq C_1^p \exp\left[p\left(\frac{M_2(\nu)}{T} - \frac{\mathrm{EOT}_T^{c_T}(\mu,\nu)}{2T}\right)\right].\end{aligned}$$

On the other hand if $\mathsf{CD}(\kappa, N)$ holds with $N < +\infty$, the lower bound (2.2.7) leads to

$$\log P_T g(x) \geq \frac{\mathrm{EOT}_T^{c_T}(\mu,\nu)}{2T} - \log\left[C_1\,\mathfrak{m}\left(B_{\sqrt{T}}(x)\right)\right] - C_2\,T - \frac{\mathsf{d}^2(x,z_0) + M_2(\nu)}{T}.$$

Therefore if we consider again $p = r\,T$, up to relabelling the positive constant C_1 at each step, we have

$$\begin{aligned}&\left\|(P_T g)^{-1}\right\|_p^p = \int (P_T g)^{-p}\mathrm{d}\mathfrak{m} = \int e^{-p\log P_T g}\mathrm{d}\mathfrak{m}\\ &\leq C_1^p\, e^{\left(\frac{M_2(\nu)}{T} - \frac{\mathrm{EOT}_T^{c_T}(\mu,\nu)}{2T} + C_2\,T\right)p}\int \mathfrak{m}\left(B_{\sqrt{T}}(x)\right)^p e^{\frac{p}{T}\mathsf{d}^2(x,z_0)}\mathrm{d}\mathfrak{m}(x)\\ &\leq C_1^p \exp\left[p\left(\frac{M_2(\nu)}{T} - \frac{\mathrm{EOT}_T^{c_T}(\mu,\nu)}{2T} + C_2\,T\right)\right].\end{aligned}$$

This concludes our proof. □

We are now ready to state and prove the quantitative stability estimate for the optimiser of SP. It is worth pointing out that Theorems 4.2.2 and 4.2.7 below require **A**3 only because the latter is required for the corrector estimates. Moreover, the integrability condition (I) does not play any role in the constants appearing in the following stability bound. Indeed one could, at least formally, perform the computations that lead to (4.2.1) and (4.2.2) without this integrability condition.

Theorem 4.2.2. *Let* $(M, \mathsf{d}, \mathfrak{m})$ *satisfy* (CD) *and* (I). *For any* (μ,ν) *and* $(\bar\mu,\bar\nu)$ *satisfy-*

*ing **A**1 and **A**3 it holds*

$$\begin{aligned}\sqrt{E_{2\kappa}(T)}\,\mathscr{H}^{\mathrm{sym}}(\boldsymbol{\pi}^{\mu\to\nu,T},\boldsymbol{\pi}^{\bar\mu\to\bar\nu,T}) \leq\ & \sqrt{E_{2\kappa}(T)}\Big(\mathscr{H}^{\mathrm{sym}}(\mu,\bar\mu)+\mathscr{H}^{\mathrm{sym}}(\nu,\bar\nu)\Big)\\ &+\sqrt{\mathcal{C}_T(\mu,\nu)-\mathscr{H}(\mu|\mathfrak{m})}\,\|\mu-\bar\mu\|_{\dot{\mathrm{H}}^{-1}(\mu)}\\ &+\sqrt{\mathcal{C}_T(\mu,\nu)-\mathscr{H}(\nu|\mathfrak{m})}\,\|\nu-\bar\nu\|_{\dot{\mathrm{H}}^{-1}(\nu)}\\ &+\sqrt{\mathcal{C}_T(\bar\mu,\bar\nu)-\mathscr{H}(\bar\mu|\mathfrak{m})}\,\|\bar\mu-\mu\|_{\dot{\mathrm{H}}^{-1}(\bar\mu)}\\ &+\sqrt{\mathcal{C}_T(\bar\mu,\bar\nu)-\mathscr{H}(\bar\nu|\mathfrak{m})}\,\|\bar\nu-\nu\|_{\dot{\mathrm{H}}^{-1}(\bar\nu)}\end{aligned} \tag{4.2.1}$$

where $E_{2\kappa}$ is defined as in (2.1.10)*. Moreover, under the same assumptions it holds*

$$\begin{aligned}&\sqrt{E_{2\kappa}(T)}\,\mathscr{H}^{\mathrm{sym}}(\boldsymbol{\pi}^{\mu\to\nu,T},\boldsymbol{\pi}^{\bar\mu\to\bar\nu,T})\\ &\quad\leq\Big[\sqrt{\mathcal{I}(\mu)}+\sqrt{\mathcal{C}_T(\mu,\nu)-\mathscr{H}(\mu|\mathfrak{m})}\Big]\,\|\mu-\bar\mu\|_{\dot{\mathrm{H}}^{-1}(\mu)}\\ &\quad+\Big[\sqrt{\mathcal{I}(\bar\mu)}+\sqrt{\mathcal{C}_T(\bar\mu,\bar\nu)-\mathscr{H}(\bar\mu|\mathfrak{m})}\Big]\,\|\bar\mu-\mu\|_{\dot{\mathrm{H}}^{-1}(\bar\mu)}\\ &\quad+\Big[\sqrt{\mathcal{I}(\nu)}+\sqrt{\mathcal{C}_T(\mu,\nu)-\mathscr{H}(\nu|\mathfrak{m})}\Big]\,\|\nu-\bar\nu\|_{\dot{\mathrm{H}}^{-1}(\nu)}\\ &\quad+\Big[\sqrt{\mathcal{I}(\bar\nu)}+\sqrt{\mathcal{C}_T(\bar\mu,\bar\nu)-\mathscr{H}(\bar\nu|\mathfrak{m})}\Big]\,\|\bar\nu-\nu\|_{\dot{\mathrm{H}}^{-1}(\bar\nu)}\,.\end{aligned} \tag{4.2.2}$$

Proof. Let us start with proving the first estimate (4.2.1). We can assume that $\mathscr{H}^{\mathrm{sym}}(\mu,\bar\mu)$, $\mathscr{H}^{\mathrm{sym}}(\nu,\bar\nu)$ are both finite, otherwise our claim is trivial. Firstly, notice that we can write

$$\begin{aligned}&\mathscr{H}(\boldsymbol{\pi}^{\mu\to\nu,T}|\boldsymbol{\pi}^{\bar\mu\to\bar\nu,T})-\mathscr{H}(\nu|\bar\nu)-\mathscr{H}(\mu|\bar\mu)\\ &\quad=\int\log\frac{f(x)g(y)}{\bar f(x)\bar g(y)}\,\mathrm{d}\boldsymbol{\pi}^{\mu\to\nu,T}(x,y)-\int\log\frac{gP_Tf}{\bar g P_T\bar f}\mathrm{d}\nu-\int\log\frac{fP_Tg}{\bar f P_T\bar g}\mathrm{d}\mu\\ &\quad=\int\log\frac{P_T\bar g(x)\,P_T\bar f(y)}{P_Tg(x)\,P_Tf(y)}\,\mathrm{d}\boldsymbol{\pi}^{\mu\to\nu,T}(x,y)\\ &\quad=\int\Big[\log P_T\bar g(x)+\log P_T\bar f(y)-\log P_Tg(x)-\log P_Tf(y)\Big]\mathrm{d}\boldsymbol{\pi}^{\mu\to\nu,T}(x,y)\,.\end{aligned}$$

We claim that the above right-hand side is equal to

$$\int\log P_T\bar g\mathrm{d}\mu+\int\log P_T\bar f\mathrm{d}\nu-\int\log P_Tg\mathrm{d}\mu-\int\log P_Tf\mathrm{d}\nu\,. \tag{4.2.3}$$

Since $\mathscr{H}(\mu|\mathfrak{m})$, $\mathscr{H}(\nu|\mathfrak{m})$ are both finite, thanks to Theorem 2.2.1 we already know that $\log P_T g \in \mathrm{L}^1(\mu)$ and $\log P_T f \in \mathrm{L}^1(\nu)$ and that under our normalisation (2.2.13) they read as

$$-\int \log P_T g \mathrm{d}\mu = -\int \log P_T f \mathrm{d}\nu = \frac{\mathrm{EOT}_T^{c_T}(\mu,\nu)}{2T} \in (-\infty,+\infty)\,.$$

Hence for the claim to be true it only remains to prove that

$$(\log P_T\bar{g})^- \in \mathrm{L}^1(\mu) \quad \text{and} \quad \big(\log P_T\bar{f}\big)^- \in \mathrm{L}^1(\nu)\,, \tag{4.2.4}$$

because this implies that $\int \log P_T\bar{g}\mathrm{d}\mu$, $\int \log P_T\bar{f}\mathrm{d}\nu > -\infty$, whence the fact that (4.2.3) is a well-defined summation. We will prove (4.2.4) by relying on Lemma 4.2.1 and Lemma 4.1.1. We consider two different cases.

1st case: $\mathfrak{m} \in \mathcal{P}(M)$**.** Consider the positive measurable function

$$h = \mathbf{1}_{\{P_T\bar{g}<1\}}\, P_T\bar{g} + \mathbf{1}_{\{P_T\bar{g}\geq 1\}}$$

and notice that $(\log P_T\bar{g})^- = -\log h$. Since $\mathfrak{m}(M) = 1$ and $h \leq 1$ we have $h \in \mathrm{L}^1(\mathfrak{m})$; moreover $h^{-1} \in \mathrm{L}^{rT}(\mathfrak{m})$ since

$$\begin{aligned}\left\|h^{-1}\right\|_{rT}^{rT} = \int h^{-rT}\,\mathrm{d}\mathfrak{m} &= \int \left[\mathbf{1}_{\{P_T\bar{g}<1\}}(P_T\bar{g})^{-rT} + \mathbf{1}_{\{P_T\bar{g}\geq 1\}}\right]\mathrm{d}\mathfrak{m} \\ &\leq \left\|(P_T\bar{g})^{-1}\right\|_{rT}^{rT} + \mathfrak{m}\{P_T\bar{g}\geq 1\}\end{aligned}$$

and the right-hand side is finite because $\mathfrak{m}\{P_T\bar{g}\geq 1\} \leq 1$ and $(P_T\bar{g})^{-1} \in \mathrm{L}^{rT}(\mathfrak{m})$ by Lemma 4.2.1. Therefore we may apply Lemma 4.1.1 with $\mathfrak{q} = \mathfrak{m}$ and $\mathfrak{p} = \mu$ and deduce

$$\int |\log h|\,\mathrm{d}\mu \leq 2\,e^{\mathscr{H}(\mu|\mathfrak{m})-1}\left\{\frac{1}{rT} \vee \log_2\left[1 + \left(1 + \left\|(P_T\bar{g})^{-1}\right\|_{rT}^{rT}\right)^{\frac{1}{rT}}\right]\right\} < +\infty\,.$$

Hence, we have proven that

$$0 \leq \int (\log P_T\bar{g})^-\,\mathrm{d}\mu = \int |\log h|\,\mathrm{d}\mu < +\infty\,,$$

which is equivalent to $(\log P_T\bar{g})^- \in \mathrm{L}^1(\mu)$.

2nd case: $\mathfrak{m} \notin \mathcal{P}(M)$**.** The proof is similar to the previous case, however we work with the probability measure $\mathfrak{m}_W$ (defined via (1.A.3)) instead of $\mathfrak{m}$. Indeed, since we are assuming $\mathscr{H}(\mu|\mathfrak{m}) < +\infty$, there exists a positive measurable function $W\colon M \to [0,+\infty)$ such that

$$z_W = \int e^{-W}\mathrm{d}\mathfrak{m} < +\infty\,, \quad W \in \mathrm{L}^1(\mu) \quad \text{and} \quad \mathscr{H}(\mu|\mathfrak{m}_W) < +\infty\,,$$

where $\mathfrak{m}_W := z_W^{-1} e^{-W} \mathfrak{m} \in \mathcal{P}(M)$. By considering the same positive measurable function

$$h = \mathbf{1}_{\{P_T\bar{g}<1\}}\, P_T\bar{g} + \mathbf{1}_{\{P_T\bar{g}\geq 1\}}$$

we are guaranteed once again that $\|h\|_{\mathrm{L}^1(\mathfrak{m}_W)} \leq 1$ and that

$$\begin{aligned}\left\|h^{-1}\right\|^{rT}_{\mathrm{L}^{rT}(\mathfrak{m}_W)} &= \int h^{-rT}\,\mathrm{d}\mathfrak{m}_W = \int \left[\mathbf{1}_{\{P_T\bar{g}<1\}}\,(P_T\bar{g})^{-rT} + \mathbf{1}_{\{P_T\bar{g}\geq 1\}}\right]\,\mathrm{d}\mathfrak{m}_W \\ &\leq \left\|(P_T\bar{g})^{-1}\right\|^{rT}_{\mathrm{L}^{rT}(\mathfrak{m}_W)} + \mathfrak{m}_W\,\{P_T\bar{g}\geq 1\} \leq \left\|(P_T\bar{g})^{-1}\right\|^{rT}_{\mathrm{L}^{rT}(\mathfrak{m}_W)} + 1\,,\end{aligned}$$

By arguing in the same way as in the second half of Lemma 4.2.1 (since $\mathfrak{m}$ is not a probability, recall that we are under the $\mathsf{CD}(\kappa, N)$ condition with $N < +\infty$) we see that

$$\begin{aligned}\left\|(P_T\bar{g})^{-1}\right\|^{rT}_{\mathrm{L}^{rT}(\mathfrak{m}_W)} &= \int (P_T\bar{g})^{-rT}\mathrm{d}\mathfrak{m}_W \\ &\leq C_1^{rT}\, e^{\left(\frac{M_2(\bar{\nu})}{T} - \frac{\mathrm{EOT}_T^{c_T}(\bar{\mu},\bar{\nu})}{2T} + C_2\,T\right) rT} \int \mathfrak{m}\left(B_{\sqrt{T}}(x)\right)^{rT} e^{r\,\mathsf{d}^2(x,z_0)}\mathrm{d}\mathfrak{m}_W(x),\end{aligned}$$

which is finite because of the integrability condition (I) and from the fact that $W \geq 0$. Therefore we have shown that $\|(P_T\bar{g})^{-1}\|^{rT}_{\mathrm{L}^{rT}(\mathfrak{m}_W)}$ is finite and once again from Lemma 4.1.1 (this time with $\mathfrak{q} = \mathfrak{m}_W$ and $\mathfrak{p} = \mu$) we deduce that

$$\int |\log h|\,\mathrm{d}\mu \leq 2\,e^{\mathscr{H}(\mu|\mathfrak{m}_W)-1}\left\{\frac{1}{rT} \vee \log_2\left[1 + \left(1 + \left\|(P_T\bar{g})^{-1}\right\|^{rT}_{\mathrm{L}^{rT}(\mathfrak{m}_W)}\right)^{\frac{1}{rT}}\right]\right\}$$

is finite. Hence, we have proven that

$$0 \leq \int (\log P_T\bar{g})^-\,\mathrm{d}\mu = \int |\log h|\,\mathrm{d}\mu < +\infty,$$

which is equivalent to $(\log P_T\bar{g})^- \in \mathrm{L}^1(\mu)$.

By arguing in the same way we can also prove that $\left(\log P_T\bar{f}\right)^- \in \mathrm{L}^1(\nu)$, whence the validity of (4.2.4). This implies $\int \log P_T\bar{g}\mathrm{d}\mu$, $\int \log P_T\bar{f}\mathrm{d}\nu > -\infty$ and that (4.2.3) is a well-defined summation. As a consequence it follows

$$\begin{aligned}&\mathscr{H}(\boldsymbol{\pi}^{\mu\to\nu,T}|\boldsymbol{\pi}^{\bar{\mu}\to\bar{\nu},T}) - \mathscr{H}(\nu|\bar{\nu}) - \mathscr{H}(\mu|\bar{\mu}) \\ &\quad= \int \log P_T\bar{g}\,\mathrm{d}\mu + \int \log P_T\bar{f}\,\mathrm{d}\nu - \int \log P_T g\,\mathrm{d}\mu - \int \log P_T f\,\mathrm{d}\nu\,.\end{aligned} \tag{4.2.5}$$

By exchanging the roles of μ, ν and $\bar{\mu}, \bar{\nu}$ we can also write

$$\begin{aligned}&\mathscr{H}(\boldsymbol{\pi}^{\bar{\mu}\to\bar{\nu},T}|\boldsymbol{\pi}^{\mu\to\nu,T}) - \mathscr{H}(\bar{\nu}|\nu) - \mathscr{H}(\bar{\mu}|\mu) \\ &\quad= \int \log P_T g\,\mathrm{d}\bar{\mu} + \int \log P_T f\,\mathrm{d}\bar{\nu} - \int \log P_T\bar{g}\,\mathrm{d}\bar{\mu} - \int \log P_T\bar{f}\,\mathrm{d}\bar{\nu}\,,\end{aligned} \tag{4.2.6}$$

which added to (4.2.5) gives

$$
\begin{aligned}
\mathscr{H}^{\mathrm{sym}}(\boldsymbol{\pi}^{\mu\to\nu,T},\boldsymbol{\pi}^{\bar\mu\to\bar\nu,T}) &= \mathscr{H}^{\mathrm{sym}}(\mu,\bar\mu)+\mathscr{H}^{\mathrm{sym}}(\nu,\bar\nu)-\int\log P_T g\,\mathrm{d}(\mu-\bar\mu)\\
&\quad-\int\log P_T\bar g\,\mathrm{d}(\bar\mu-\mu)-\int\log P_T f\,\mathrm{d}(\nu-\bar\nu)-\int\log P_T\bar f\,\mathrm{d}(\bar\nu-\nu)\\
&\le \mathscr{H}^{\mathrm{sym}}(\mu,\bar\mu)+\mathscr{H}^{\mathrm{sym}}(\nu,\bar\nu)+\|\nabla\log P_T g\|_{\mathrm{L}^2(\mu)}\,\|\mu-\bar\mu\|_{\dot{\mathrm{H}}^{-1}(\mu)}\\
&\quad+\|\nabla\log P_T\bar g\|_{\mathrm{L}^2(\bar\mu)}\,\|\bar\mu-\mu\|_{\dot{\mathrm{H}}^{-1}(\bar\mu)}+\|\nabla\log P_T f\|_{\mathrm{L}^2(\nu)}\,\|\nu-\bar\nu\|_{\dot{\mathrm{H}}^{-1}(\nu)}\\
&\quad+\|\nabla\log P_T\bar f\|_{\mathrm{L}^2(\bar\nu)}\,\|\bar\nu-\nu\|_{\dot{\mathrm{H}}^{-1}(\bar\nu)}\,.
\end{aligned}
$$

Let us point out that so far we have not relied on the assumption **A**3, which is needed now when applying the corrector estimates. Indeed, given the above bound, inequality (4.2.1) follows from the corrector estimates from Proposition 3.1.2.

The proof of (4.2.2) under the assumption of finite Fisher information is postponed after the next *mixed integrability* result, which complements what we have proven so far and which is necessary for the integral computations that will follow. □

Corollary 4.2.3. *Let* $(M,\mathsf{d},\mathfrak{m})$ *satisfy* (CD) *and* (I). *For any* (μ,ν) *and* $(\bar\mu,\bar\nu)$ *satisfying A1,* ***A****3 and such that* $\|\mu-\bar\mu\|_{\dot{\mathrm{H}}^{-1}(\mu)}$, $\|\bar\mu-\mu\|_{\dot{\mathrm{H}}^{-1}(\bar\mu)}$, $\|\nu-\bar\nu\|_{\dot{\mathrm{H}}^{-1}(\nu)}$, *and* $\|\bar\nu-\nu\|_{\dot{\mathrm{H}}^{-1}(\bar\nu)}$ *are all finite and with* $\mathscr{H}^{\mathrm{sym}}(\mu,\bar\mu)$, $\mathscr{H}^{\mathrm{sym}}(\nu,\bar\nu)<\infty$, *we have*

$$\log P_T g,\ \log P_T\bar g\in\mathrm{L}^1(\mu)\cap\mathrm{L}^1(\bar\mu)\quad\text{and}\quad\log P_T f,\ \log P_T\bar f\in\mathrm{L}^1(\nu)\cap\mathrm{L}^1(\bar\nu)\,.$$

As a consequence we deduce that also

$$\log f,\ \log\bar f\in\mathrm{L}^1(\mu)\cap\mathrm{L}^1(\bar\mu)\quad\text{and}\quad\log g,\ \log\bar g\in\mathrm{L}^1(\nu)\cap\mathrm{L}^1(\bar\nu)\,.$$

Proof. As mentioned in the previous proof we already know, thanks to Theorem 2.2.1, that

$$\log P_T g\in\mathrm{L}^1(\mu),\ \log P_T f\in\mathrm{L}^1(\nu),\ \log P_T\bar g\in\mathrm{L}^1(\bar\mu),\ \text{and}\ \log P_T\bar f\in\mathrm{L}^1(\bar\nu)\,.$$

Now, given our assumptions, the left-hand side of (4.2.5) is finite (thanks to Theorem 4.2.2) and since we have $\log P_T g\in\mathrm{L}^1(\mu)$, $\log P_T f\in\mathrm{L}^1(\nu)$ and we have further shown that $\int\log P_T\bar g\mathrm{d}\mu$, $\int\log P_T\bar f\mathrm{d}\nu\in(-\infty,+\infty]$ (cf. above proof of Theorem 4.2.2), we deduce that

$$\log P_T\bar g\in\mathrm{L}^1(\mu)\quad\text{and}\quad\log P_T\bar f\in\mathrm{L}^1(\nu)\,.$$

Similarly, working with (4.2.6) we get $\log P_T g\in\mathrm{L}^1(\bar\mu)$ and $\log P_T f\in\mathrm{L}^1(\bar\nu)$.

Finally, from what we have just proven above and since $\mathscr{H}^{\mathrm{sym}}(\mu,\bar\mu)$ and $\mathscr{H}^{\mathrm{sym}}(\nu,\bar\nu)$ are both finite, we deduce

$$\log\frac{f}{\bar f}\in\mathrm{L}^1(\mu)\cap\mathrm{L}^1(\bar\mu)\quad\text{and}\quad\log\frac{g}{\bar g}\in\mathrm{L}^1(\nu)\cap\mathrm{L}^1(\bar\nu)\,,$$

which combined with $\log f \in \mathrm{L}^1(\mu)$, $\log g \in \mathrm{L}^1(\nu)$, $\log \bar f \in \mathrm{L}^1(\bar\mu)$ and $\log \bar g \in \mathrm{L}^1(\bar\nu)$ gives our final assertion. □

Resuming the proof of Theorem 4.2.2. Let us now prove (4.2.2). Without loss of generality we may assume that the Fisher information of the marginals and the negative Sobolev norms are all finite, otherwise the claimed estimate is trivial. Let us start by showing that this guarantees the finiteness of the symmetric relative entropies $\mathscr{H}^{\mathrm{sym}}(\mu,\bar\mu)$, $\mathscr{H}^{\mathrm{sym}}(\nu,\bar\nu)$. Indeed, since $\mathcal{I}(\mu) < +\infty$ we have $\nabla \log \frac{\mathrm{d}\mu}{\mathrm{d}\mathfrak{m}} \in \mathrm{L}^2(\mu)$ and therefore by the definition of $\|\mu-\bar\mu\|_{\dot{\mathrm{H}}^{-1}(\mu)}$ we deduce

$$\left|\int \log \frac{\mathrm{d}\mu}{\mathrm{d}\mathfrak{m}}\,\mathrm{d}(\mu-\bar\mu)\right| \le \sqrt{\mathcal{I}(\mu)}\ \|\mu-\bar\mu\|_{\dot{\mathrm{H}}^{-1}(\mu)} < +\infty\,.$$

Thanks to the above bound and to the finiteness of $\mathscr{H}(\mu|\mathfrak{m})$ and $\mathscr{H}(\bar\mu|\mathfrak{m})$ we are allowed to write

$$\begin{aligned}\int \log\frac{\mathrm{d}\bar\mu}{\mathrm{d}\mu}\,\mathrm{d}\bar\mu &= \int\left[\frac{\mathrm{d}\bar\mu}{\mathrm{d}\mathfrak{m}}\log\frac{\mathrm{d}\bar\mu}{\mathrm{d}\mathfrak{m}} + \left(\frac{\mathrm{d}\mu}{\mathrm{d}\mathfrak{m}}-\frac{\mathrm{d}\bar\mu}{\mathrm{d}\mathfrak{m}}\right)\log\frac{\mathrm{d}\mu}{\mathrm{d}\mathfrak{m}} - \frac{\mathrm{d}\mu}{\mathrm{d}\mathfrak{m}}\log\frac{\mathrm{d}\mu}{\mathrm{d}\mathfrak{m}}\right]\mathrm{d}\mathfrak{m}\\ &= \mathscr{H}(\bar\mu|\mathfrak{m}) - \mathscr{H}(\mu|\mathfrak{m}) + \int \log\frac{\mathrm{d}\mu}{\mathrm{d}\mathfrak{m}}\,\mathrm{d}(\mu-\bar\mu) < +\infty\,,\end{aligned}$$

which reads as $\mathscr{H}(\bar\mu|\mu) < +\infty$. Let us just mention that above the Radon-Nikodym derivative is well defined since, under our assumption, $\bar\mu \sim \mu$ are equivalent. Indeed, if there were a Borel subset $A \subset M$ such that $\mu(A) = 0$ and $\bar\mu(A) > 0$, then by choosing as test function h (a mollified version of) $\mathbf{1}_A$ we would get $\|\mu-\bar\mu\|_{\dot{\mathrm{H}}^{-1}(\mu)} = +\infty$ which we are assuming to be finite; therefore $\bar\mu \ll \mu$. Similarly one can prove $\mu \ll \bar\mu$, and hence $\bar\mu \sim \mu$. We have thus proven that $\mathscr{H}(\bar\mu|\mu)$ is finite. By reasoning in the same fashion one can prove that $\mathscr{H}(\mu|\bar\mu)$, $\mathscr{H}(\bar\nu|\nu)$ and $\mathscr{H}(\nu|\bar\nu)$ are also finite, whence the finiteness of the symmetric entropies.

At this stage, the proof is similar to the one already presented above. It is indeed enough to notice that

$$\begin{aligned}&\mathscr{H}^{\mathrm{sym}}(\boldsymbol{\pi}^{\mu\to\nu,T},\boldsymbol{\pi}^{\bar\mu\to\bar\nu,T})\\ &= \int \log f\,\mathrm{d}(\mu-\bar\mu) + \int \log\bar f\,\mathrm{d}(\bar\mu-\mu) + \int \log g\,\mathrm{d}(\nu-\bar\nu) + \int\log\bar g\,\mathrm{d}(\bar\nu-\nu)\\ &\le \|\nabla\log f\|_{\mathrm{L}^2(\mu)}\,\|\mu-\bar\mu\|_{\dot{\mathrm{H}}^{-1}(\mu)} + \|\nabla\log\bar f\|_{\mathrm{L}^2(\bar\mu)}\,\|\bar\mu-\mu\|_{\dot{\mathrm{H}}^{-1}(\bar\mu)}\\ &+ \|\nabla\log g\|_{\mathrm{L}^2(\nu)}\,\|\nu-\bar\nu\|_{\dot{\mathrm{H}}^{-1}(\nu)} + \|\nabla\log\bar g\|_{\mathrm{L}^2(\bar\nu)}\,\|\bar\nu-\nu\|_{\dot{\mathrm{H}}^{-1}(\bar\nu)}\,,\end{aligned}$$

and that we can bound the L^2-norm of the gradient of the potentials in terms of the Fisher information, e.g.

$$\begin{aligned}\|\nabla\log f\|_{\mathrm{L}^2(\mu)} &\le \left\|\nabla\log\frac{\mathrm{d}\mu}{\mathrm{d}\mathfrak{m}}\right\|_{\mathrm{L}^2(\mu)} + \|\nabla\log P_T g\|_{\mathrm{L}^2(\mu)}\\ &= \sqrt{\mathcal{I}(\mu)} + \|\nabla\log P_T g\|_{\mathrm{L}^2(\mu)}\end{aligned}$$

and analogously for the other three summands. □

Remark 4.2.4. *The cost $\mathcal{C}_T(\mu,\nu)$ appearing on the right-hand sides of* (4.2.1) *and* (4.2.2) *can be bounded by the independent coupling $\mathscr{H}(\mu\otimes\nu|\mathrm{R}_{0,T})$ and henceforth (by combining* (2.2.8), (2.2.7), *and* (2.1.14)*) by the quantity*

$$T\,\mathcal{C}_T(\mu,\nu) \le T\,\mathscr{H}(\mu|\mathfrak{m}) + T\,\mathscr{H}(\nu|\mathfrak{m}) + C(T+T^2) + M_2(\mu) + M_2(\nu) \\ +C\,T\Big[M_1(\mu)+M_1(\nu)\Big]\,,$$

where $M_1(\mu)$, $M_1(\nu)$ and $M_2(\mu)$, $M_2(\nu)$ denote the first and second moments of μ and ν and the constant $C \ge 0$ is independent of the marginals (and it is equal to 0 *if $(M,\mathsf{d},\mathfrak{m})$ satisfies a $\mathsf{CD}(\kappa,\infty)$ condition with $\mathfrak{m}(M)=1$). We have kept the explicit dependence on the cost in* (4.2.1) *because this gives a bound which is sharper compared to the one with the independent coupling $\mathscr{H}(\mu\otimes\nu|\mathrm{R}_{0,T})$, especially in the small-time limit $T\downarrow 0$.*

Remark 4.2.5 (Sharpness in the long-time regime)**.** *Under a non-negative curvature condition we know that in the long-time limit $\boldsymbol{\pi}^{\mu\to\nu,T} \rightharpoonup \mu\otimes\nu$ (cf. [CT21, Lemma 3.1]) and hence we expect*

$$\mathscr{H}^{\mathrm{sym}}(\boldsymbol{\pi}^{\mu\to\nu,T},\boldsymbol{\pi}^{\bar\mu\to\bar\nu,T}) \to \mathscr{H}^{\mathrm{sym}}(\mu\otimes\nu,\bar\mu\otimes\bar\nu) = \mathscr{H}^{\mathrm{sym}}(\mu,\bar\mu) + \mathscr{H}^{\mathrm{sym}}(\nu,\bar\nu)\,.$$

This shows that the previous bound (4.2.1) *under a non-negative curvature condition is* sharp. *Indeed, if $\kappa\ge 0$* (4.2.1) *implies*

$$\limsup_{T\to\infty} \mathscr{H}^{\mathrm{sym}}(\boldsymbol{\pi}^{\mu\to\nu,T},\boldsymbol{\pi}^{\bar\mu\to\bar\nu,T}) \le \mathscr{H}^{\mathrm{sym}}(\mu,\bar\mu) + \mathscr{H}^{\mathrm{sym}}(\nu,\bar\nu)$$

and the above convergence is exponentially fast (of order $\approx e^{-\kappa T}/\sqrt{T}$). This is another confirmation that under the corrector estimates and a non-negative curvature condition we are able to efficiently describe the exact behaviour of the Schrödinger problem, as it has already been shown in the context of the entropic turnpike estimates (e.g., [Con19, Theorem 1.4] in the classic setting)

A stability result can also be stated at the level of the optimal value of the Schrödinger problem, after normalising it as follows

$$T\,\mathcal{C}_T(\mu,\nu) - T\,\mathscr{H}(\mu|\mathfrak{m}) - T\,\mathscr{H}(\nu|\mathfrak{m}) = \mathrm{EOT}_T^{c_T}(\mu,\nu)\,,$$

which corresponds to the EOT entropic cost defined at (3.2.5) associated to $c_T(x,y) := -T\log\mathsf{p}_T(x,y)$. Before moving to the proof of stability estimates for $\mathrm{EOT}_T^{c_T}(\mu,\nu)$, we need another technical result given by the following

Lemma 4.2.6. *Let $(M,\mathsf{d},\mathfrak{m})$ satisfy* (CD) *and* (I). *For any (μ,ν) and $(\bar\mu,\bar\nu)$ satisfying **A**1, **A**3 and such that $\|\mu-\bar\mu\|_{\dot{\mathrm{H}}^{-1}(\mu)}$, $\|\bar\mu-\mu\|_{\dot{\mathrm{H}}^{-1}(\bar\mu)}$, $\|\nu-\bar\nu\|_{\dot{\mathrm{H}}^{-1}(\nu)}$, and $\|\bar\nu-\nu\|_{\dot{\mathrm{H}}^{-1}(\bar\nu)}$ are all finite and with $\mathscr{H}^{\mathrm{sym}}(\mu,\bar\mu)$, $\mathscr{H}^{\mathrm{sym}}(\nu,\bar\nu)<\infty$, it holds*

$$\int \log\frac{P_T f}{P_T\bar f}\,\mathrm{d}\nu \le \int \log\frac{f}{\bar f}\,\mathrm{d}\mu \qquad \textit{and} \qquad \int \log\frac{P_T g}{P_T\bar g}\,\mathrm{d}\mu \le \int \log\frac{g}{\bar g}\,\mathrm{d}\nu\,.$$

Analogous bounds hold when we exchange the roles between (μ,ν,f,g) and $(\bar\mu,\bar\nu,\bar f,\bar g)$.

Proof. Since $\mathscr{H}(\nu|\bar\nu) \le \mathscr{H}(\boldsymbol{\pi}^{\mu\to\nu,T}|\boldsymbol{\pi}^{\mu\to\bar\nu,T})$, by means of the fg-decomposition the inequality reads as

$$\int \log\frac{P_T f}{P_T\bar f}\mathrm{d}\nu + \int \log\frac{g}{\bar g}\mathrm{d}\nu \le \int \log\frac{f}{\bar f}\mathrm{d}\mu + \int\log\frac{g}{\bar g}\mathrm{d}\nu\,,$$

which yields the first inequality. The other bound can be proven in the same way. □

By following the same line of reasoning applied for the stability of the optimal plans, we can finally deduce the following

Theorem 4.2.7. *Let $(M,\mathsf{d},\mathfrak{m})$ satisfy* (CD) *and* (I). *For any (μ,ν) and $(\bar\mu,\bar\nu)$ satisfying **A1**, **A3**, and such that $\|\mu-\bar\mu\|_{\dot{\mathrm{H}}^{-1}(\mu)}$, $\|\bar\mu-\mu\|_{\dot{\mathrm{H}}^{-1}(\bar\mu)}$, $\|\nu-\bar\nu\|_{\dot{\mathrm{H}}^{-1}(\nu)}$, and $\|\bar\nu-\nu\|_{\dot{\mathrm{H}}^{-1}(\bar\nu)}$ are all finite and with $\mathscr{H}^{\mathrm{sym}}(\mu,\bar\mu)$, $\mathscr{H}^{\mathrm{sym}}(\nu,\bar\nu)<\infty$, it holds*

$$\begin{aligned}\mathrm{EOT}_T^{c_T}(\mu,\nu) - \mathrm{EOT}_T^{c_T}(\bar\mu,\bar\nu) \le T\Big[\mathscr{H}(\bar\mu|\mu)\wedge\mathscr{H}(\bar\nu|\nu)\Big]&\\ + \frac{T}{\sqrt{E_{2\kappa}(T)}}\sqrt{\mathcal{C}_T(\mu,\nu)-\mathscr{H}(\mu|\mathfrak{m})}\,\|\mu-\bar\mu\|_{\dot{\mathrm{H}}^{-1}(\mu)}&\\ + \frac{T}{\sqrt{E_{2\kappa}(T)}}\sqrt{\mathcal{C}_T(\mu,\nu)-\mathscr{H}(\nu|\mathfrak{m})}\,\|\nu-\bar\nu\|_{\dot{\mathrm{H}}^{-1}(\nu)}&\,,\end{aligned}$$

where $E_{2\kappa}$ is defined as in (2.1.10), *from which it follows*

$$\begin{aligned}\frac{\sqrt{E_{2\kappa}(T)}}{T}\,\big|\mathrm{EOT}_T^{c_T}(\bar\mu,\bar\nu)-\mathrm{EOT}_T^{c_T}(\mu,\nu)\big| \le \sqrt{E_{2\kappa}(T)}\Big[\mathscr{H}^{\mathrm{sym}}(\mu,\bar\mu)\wedge\mathscr{H}^{\mathrm{sym}}(\nu,\bar\nu)\Big]&\\ + \sqrt{\mathcal{C}_T(\mu,\nu)-\mathscr{H}(\mu|\mathfrak{m})}\,\|\mu-\bar\mu\|_{\dot{\mathrm{H}}^{-1}(\mu)} + \sqrt{\mathcal{C}_T(\mu,\nu)-\mathscr{H}(\nu|\mathfrak{m})}\,\|\nu-\bar\nu\|_{\dot{\mathrm{H}}^{-1}(\nu)}&\\ + \sqrt{\mathcal{C}_T(\bar\mu,\bar\nu)-\mathscr{H}(\bar\mu|\mathfrak{m})}\,\|\bar\mu-\mu\|_{\dot{\mathrm{H}}^{-1}(\bar\mu)} + \sqrt{\mathcal{C}_T(\bar\mu,\bar\nu)-\mathscr{H}(\bar\nu|\mathfrak{m})}\,\|\bar\nu-\nu\|_{\dot{\mathrm{H}}^{-1}(\bar\nu)}&\,.\end{aligned}$$

Proof. First of all, since the assumptions of Corollary 4.2.3 are met, the following computations are all well defined. With this said, let us start by noticing that

$$\begin{aligned}\mathrm{EOT}_T^{c_T}(\mu,\nu) &= T\,\mathcal{C}_T(\mu,\nu) - T\,\mathscr{H}(\mu|\mathfrak{m}) - T\,\mathscr{H}(\nu|\mathfrak{m})\\ &= -T\int\log P_T g\,\mathrm{d}\mu - T\int \log P_T f\,\mathrm{d}\nu\,,\end{aligned}$$

and therefore

$$\begin{aligned}\mathrm{EOT}_T^{c_T}(\bar\mu,\bar\nu)-\mathrm{EOT}_T^{c_T}(\mu,\nu) &= T\int\log P_T g\,\mathrm{d}\mu + T\int\log P_T f\,\mathrm{d}\nu\\ &\quad - T\int\log P_T\bar g\,\mathrm{d}\bar\mu - T\int\log P_T\bar f\,\mathrm{d}\bar\nu\\ &= T\int\log\frac{P_T g}{P_T\bar g}\mathrm{d}\mu + T\int\log\frac{P_T f}{P_T\bar f}\mathrm{d}\nu\\ &\quad - T\int\log P_T\bar g\,\mathrm{d}(\bar\mu-\mu) - T\int\log P_T\bar f\,\mathrm{d}(\bar\nu-\nu)\,.\end{aligned}$$

Applying Lemma 4.2.6 we deduce that

$$\begin{aligned}
\mathrm{EOT}_T^{c_T}(\bar\mu,\bar\nu) - \mathrm{EOT}_T^{c_T}(\mu,\nu) &\le T\int \log\frac{P_T g}{P_T\bar g}\mathrm{d}\mu + T\int \log\frac{f}{\bar f}\mathrm{d}\mu \\
&\quad -T\int \log P_T\bar g\,\mathrm{d}(\bar\mu-\mu) - T\int \log P_T\bar f\,\mathrm{d}(\bar\nu-\nu) \\
&= T\int \log\frac{\mathrm{d}\mu}{\mathrm{d}\bar\mu}\mathrm{d}\mu - T\int \log P_T\bar g\,\mathrm{d}(\bar\mu-\mu) - T\int \log P_T\bar f\,\mathrm{d}(\bar\nu-\nu) \\
&= T\mathscr{H}(\mu|\bar\mu) - T\int \log P_T\bar g\,\mathrm{d}(\bar\mu-\mu) - T\int \log P_T\bar f\,\mathrm{d}(\bar\nu-\nu)\,.
\end{aligned} \tag{4.2.7}$$

Notice that above we could have used Lemma 4.2.6 on the integral $\int\log\frac{P_T g}{P_T\bar g}\mathrm{d}\mu$ and therefore we would have got $\mathscr{H}(\nu|\bar\nu)$ instead of $\mathscr{H}(\mu|\bar\mu)$. Therefore we can state that

$$\begin{aligned}
\mathrm{EOT}_T^{c_T}(\bar\mu,\bar\nu) - \mathrm{EOT}_T^{c_T}(\mu,\nu) &\le T\Big[\mathscr{H}(\mu|\bar\mu)\wedge\mathscr{H}(\nu|\bar\nu)\Big] - T\int \log P_T\bar g\,\mathrm{d}(\bar\mu-\mu) \\
&\quad -T\int \log P_T\bar f\,\mathrm{d}(\bar\nu-\nu) \\
&\le T\Big[\mathscr{H}(\mu|\bar\mu)\wedge\mathscr{H}(\nu|\bar\nu)\Big] + T\left\|\nabla\log P_T\bar g\right\|_{\mathrm{L}^2(\bar\mu)}\left\|\bar\mu-\mu\right\|_{\dot{\mathrm{H}}^{-1}(\bar\mu)} \\
&\quad + T\left\|\nabla\log P_T\bar f\right\|_{\mathrm{L}^2(\bar\nu)}\left\|\bar\nu-\nu\right\|_{\dot{\mathrm{H}}^{-1}(\bar\nu)}\,.
\end{aligned}$$

By exchanging the roles between (μ,ν) and $(\bar\mu,\bar\nu)$ and by arguing in the same way we also get

$$\begin{aligned}
\mathrm{EOT}_T^{c_T}(\mu,\nu) - \mathrm{EOT}_T^{c_T}(\bar\mu,\bar\nu) \le\ & T\Big[\mathscr{H}(\bar\mu|\mu)\wedge\mathscr{H}(\bar\nu|\nu)\Big] \\
&+ T\left\|\nabla\log P_T g\right\|_{\mathrm{L}^2(\mu)}\left\|\mu-\bar\mu\right\|_{\dot{\mathrm{H}}^{-1}(\mu)} + T\left\|\nabla\log P_T f\right\|_{\mathrm{L}^2(\nu)}\left\|\nu-\bar\nu\right\|_{\dot{\mathrm{H}}^{-1}(\nu)}\,,
\end{aligned}$$

and therefore we deduce

$$\begin{aligned}
\left|\mathrm{EOT}_T^{c_T}(\bar\mu,\bar\nu) - \mathrm{EOT}_T^{c_T}(\mu,\nu)\right| &\le T\Big[\mathscr{H}^{\mathrm{sym}}(\mu,\bar\mu)\wedge\mathscr{H}^{\mathrm{sym}}(\nu,\bar\nu)\Big] \\
&\quad + T\left\|\nabla\log P_T g\right\|_{\mathrm{L}^2(\mu)}\left\|\mu-\bar\mu\right\|_{\dot{\mathrm{H}}^{-1}(\mu)} + T\left\|\nabla\log P_T f\right\|_{\mathrm{L}^2(\nu)}\left\|\nu-\bar\nu\right\|_{\dot{\mathrm{H}}^{-1}(\nu)} \\
&\quad + T\left\|\nabla\log P_T\bar g\right\|_{\mathrm{L}^2(\bar\mu)}\left\|\bar\mu-\mu\right\|_{\dot{\mathrm{H}}^{-1}(\bar\mu)} + T\left\|\nabla\log P_T\bar f\right\|_{\mathrm{L}^2(\bar\nu)}\left\|\bar\nu-\nu\right\|_{\dot{\mathrm{H}}^{-1}(\bar\nu)}\,.
\end{aligned}$$

Given the above, the thesis follows from the corrector estimates (cf. Proposition 3.1.2). □

Remark 4.2.8. *Notice that in the cases where we change just one marginal (e.g. $\mu = \bar\mu$), we can get rid of the relative entropy between the marginals in the above result. Therefore in order to get a bound without the symmetric relative entropy of the marginals it is enough to consider the case where one of the marginals is frozen by means of the trivial inequality*

$$\begin{aligned}
\left|\mathrm{EOT}_T^{c_T}(\bar\mu,\bar\nu) - \mathrm{EOT}_T^{c_T}(\mu,\nu)\right| \le\ & \left|\mathrm{EOT}_T^{c_T}(\bar\mu,\bar\nu) - \mathrm{EOT}_T^{c_T}(\mu,\bar\nu)\right| \\
&+ \left|\mathrm{EOT}_T^{c_T}(\mu,\bar\nu) - \mathrm{EOT}_T^{c_T}(\mu,\nu)\right|\,.
\end{aligned}$$

Let us conclude by showing that, under the finite Fisher information assumption, we may write a stability estimate for the Schrödinger costs without involving the symmetric relative entropies in the right-hand side.

Proposition 4.2.9. *Let* $(M, \mathsf{d}, \mathfrak{m})$ *satisfy* (CD) *and* (I). *For any* (μ,ν) *and* $(\bar\mu,\bar\nu)$ *satisfying* ***A***1 *and* ***A***3 *we have*

$$\begin{aligned}
&|\mathcal{C}_T(\bar\mu,\bar\nu) - \mathcal{C}_T(\mu,\nu)| \\
&\qquad \le \frac{1}{\sqrt{E_{2\kappa}(T)}}\left[\sqrt{\mathcal{I}(\mu)} + \sqrt{\mathcal{C}_T(\mu,\nu) - \mathscr{H}(\mu|\mathfrak{m})}\right] \|\mu-\bar\mu\|_{\dot{\mathrm{H}}^{-1}(\mu)} \\
&\qquad + \frac{1}{\sqrt{E_{2\kappa}(T)}}\left[\sqrt{\mathcal{I}(\bar\mu)} + \sqrt{\mathcal{C}_T(\bar\mu,\bar\nu) - \mathscr{H}(\bar\mu|\mathfrak{m})}\right] \|\bar\mu-\mu\|_{\dot{\mathrm{H}}^{-1}(\bar\mu)} \\
&\qquad + \frac{1}{\sqrt{E_{2\kappa}(T)}}\left[\sqrt{\mathcal{I}(\nu)} + \sqrt{\mathcal{C}_T(\mu,\nu) - \mathscr{H}(\nu|\mathfrak{m})}\right] \|\nu-\bar\nu\|_{\dot{\mathrm{H}}^{-1}(\nu)} \\
&\qquad + \frac{1}{\sqrt{E_{2\kappa}(T)}}\left[\sqrt{\mathcal{I}(\bar\nu)} + \sqrt{\mathcal{C}_T(\bar\mu,\bar\nu) - \mathscr{H}(\bar\nu|\mathfrak{m})}\right] \|\bar\nu-\nu\|_{\dot{\mathrm{H}}^{-1}(\bar\nu)}\,,
\end{aligned}$$

where $E_{2\kappa}$ *is defined as in* (2.1.10).

Proof. The proof runs like the one given in Theorem 4.2.7 observing that, similarly to (4.2.7), we can write

$$\mathcal{C}_T(\bar\mu,\bar\nu) - \mathcal{C}_T(\mu,\nu) \le \int \log \bar f \,\mathrm{d}(\bar\mu-\mu) + \int \log \bar g\,\mathrm{d}(\bar\nu-\nu) - \Big[\mathscr{H}(\mu|\bar\mu) \vee \mathscr{H}(\nu|\bar\nu)\Big]$$

and therefore it holds

$$\begin{aligned}
&|\mathcal{C}_T(\bar\mu,\bar\nu) - \mathcal{C}_T(\mu,\nu)| \\
&\qquad \le \|\nabla \log f\|_{\mathrm{L}^2(\mu)}\, \|\mu-\bar\mu\|_{\dot{\mathrm{H}}^{-1}(\mu)} + \|\nabla \log \bar f\|_{\mathrm{L}^2(\bar\mu)}\, \|\bar\mu-\mu\|_{\dot{\mathrm{H}}^{-1}(\bar\mu)} \\
&\qquad + \|\nabla \log g\|_{\mathrm{L}^2(\nu)}\, \|\nu-\bar\nu\|_{\dot{\mathrm{H}}^{-1}(\nu)} + \|\nabla \log \bar g\|_{\mathrm{L}^2(\bar\nu)}\, \|\bar\nu-\nu\|_{\dot{\mathrm{H}}^{-1}(\bar\nu)}\,.
\end{aligned}$$

□

Notice that in the above result we are able to consider the stability directly between the Schrödinger costs $\mathcal{C}_T(\mu,\nu)$ and $\mathcal{C}_T(\bar\mu,\bar\nu)$. This is indeed due to the fact that we are working in a finite Fisher information setting.

Let us further mention that the Lipschitz bounds proven in Theorem 4.2.2 and Proposition 4.2.9 are well behaved in the small-time limit $T \downarrow 0$. Indeed, since we have $4T\mathcal{C}_T(\mu,\nu) \to \mathbf{W}_2^2(\mu,\nu)$ (cf. (3.2.2)), if (CD), **A**1, **A**3 and (I) (for some fixed time window $T_0 > 0$) hold true and if $\mathscr{H}^{\mathrm{sym}}(\mu,\bar\mu)$, $\mathscr{H}^{\mathrm{sym}}(\nu,\bar\nu)$ are both finite, then after rescaling it holds

$$\begin{aligned}
&\limsup_{T\to 0} 2T\,\mathscr{H}^{\mathrm{sym}}(\boldsymbol{\pi}^{\mu\to\nu,T},\boldsymbol{\pi}^{\bar\mu\to\bar\nu,T}) \\
&\qquad\qquad \le \mathbf{W}_2(\mu,\nu)\left[\|\mu-\bar\mu\|_{\dot{\mathrm{H}}^{-1}(\mu)}+\|\nu-\bar\nu\|_{\dot{\mathrm{H}}^{-1}(\nu)}\right] \\
&\qquad\qquad\quad + \mathbf{W}_2(\bar\mu,\bar\nu)\left[\|\bar\mu-\mu\|_{\dot{\mathrm{H}}^{-1}(\bar\mu)}+\|\bar\nu-\nu\|_{\dot{\mathrm{H}}^{-1}(\bar\nu)}\right], \\
&\limsup_{T\to 0} 2\,|\mathrm{EOT}_T^{c_T}(\bar\mu,\bar\nu)-\mathrm{EOT}_T^{c_T}(\mu,\nu)| \\
&\qquad\qquad \le \mathbf{W}_2(\mu,\nu)\left[\|\mu-\bar\mu\|_{\dot{\mathrm{H}}^{-1}(\mu)}+\|\nu-\bar\nu\|_{\dot{\mathrm{H}}^{-1}(\nu)}\right] \\
&\qquad\qquad\quad + \mathbf{W}_2(\bar\mu,\bar\nu)\left[\|\bar\mu-\mu\|_{\dot{\mathrm{H}}^{-1}(\bar\mu)}+\|\bar\nu-\nu\|_{\dot{\mathrm{H}}^{-1}(\bar\nu)}\right].
\end{aligned} \tag{4.2.8}$$

The prefactor T on the left-hand side of the first displacement should not be surprising. For instance, it also appears in the quantitative stability bounds for the optimisers of EOT (cf. [EN22b, Theorem 3.11]), for which there is a vast literature nowadays. The most general and recent result we are aware of is the one in [EN22b], where the authors manage to prove Lipschitzianity with respect to the p-Wasserstein distance provided that the cost function satisfies a certain abstract condition introduced therein. Our estimates (cf. Theorem 4.2.7 and Proposition 4.2.9) are less tight than theirs, however our setting does not satisfy their abstract condition in general. In fact, we cannot expect in general to bound the symmetric entropy of two optimal couplings since, even assuming the weak convergence of the Schrödinger optimizer $\boldsymbol{\pi}^{\mu\to\nu,T}$ to the optimal $\mathbf{W}_2$-coupling $\boldsymbol{\pi}_0^{\mu,\nu}$, it may happen that $\mathscr{H}^{\mathrm{sym}}(\boldsymbol{\pi}_0^{\mu,\nu},\boldsymbol{\pi}_0^{\bar\mu,\bar\nu}) = +\infty$ while the right-hand side stays finite. For instance, consider $M=(0,1)$ with the marginals μ, ν, $\bar\mu$ equal to the Lebesgue measure on $(0,1)$ and $\mathrm{d}\bar\nu(x) = (\log 2)2^x\,\mathrm{d}x$. Then $\mathscr{H}^{\mathrm{sym}}(\nu,\bar\nu)$ and the left-hand sides of our bounds are finite but $\mathscr{H}^{\mathrm{sym}}(\boldsymbol{\pi}_0^{\mu,\nu},\boldsymbol{\pi}_0^{\mu,\bar\nu}) = +\infty$ since the two optimal couplings $\boldsymbol{\pi}_0^{\mu,\nu}$ and $\boldsymbol{\pi}_0^{\mu,\bar\nu}$ have disjoint supports (the first is supported on the diagonal while the latter is supported on the graph of $L(x) = \log_2 x$). For sake of completeness, the finiteness of $\|\nu-\bar\nu\|_{\dot{\mathrm{H}}^{-1}(\nu)}$ is due to [Pey18, Theorem 5], while the finiteness of $\|\bar\nu-\nu\|_{\dot{\mathrm{H}}^{-1}(\bar\nu)}$ follows from the former and [Pey18, Lemma 2].

Finally, since the negative Sobolev norm is intimately related to the $\mathbf{W}_2$-distance, our bounds may lead to a $\mathbf{W}_2$-Lipschitz estimate.

4.2.1 Application to the Entropic Optimal Transport problem with quadratic cost

In this section we translate the stability results stated in Theorem 4.2.2 and Theorem 4.2.7 to the Euclidean EOT setting with quadratic cost

$$\mathrm{EOT}_\varepsilon^{\mathrm{d}^2}(\mu,\nu) := \inf_{\pi\in\Pi(\mu,\nu)} \int |x-y|^2\,\mathrm{d}\pi + \varepsilon\,\mathscr{H}(\pi|\mu\otimes\nu)\,, \qquad \varepsilon>0\,. \tag{4.2.9}$$

We will manage to do that under sufficiently general conditions, so that our results will apply to any couple of marginals $\mu,\nu\in\mathcal{P}_2(\mathbb{R}^d)$ with finite relative entropy w.r.t. the Lebesgue measure Leb on $\mathbb{R}^d$ and with densities w.r.t. Leb locally bounded away from 0 on their support and such that $\mu(\partial\mathrm{supp}(\mu)) = \nu(\partial\mathrm{supp}(\nu)) = 0$; or with bounded and compactly supported densities. For later reference, let $\pi_\varepsilon^{\mu,\nu}$ denote the minimizer (which exists and is unique since $\mu,\nu\in\mathcal{P}_2(\mathbb{R}^d)$, cf. [Nut21, Theorem 4.2]).

In what follows we are going to consider a Schrödinger problem equivalent to the above (4.2.9), where the underlying stochastic dynamics is given by the law of the Ornstein–Uhlenbeck process

$$\mathrm{d}X_t = -\kappa X_t\mathrm{d}t + \sqrt{2}\,\mathrm{d}B_t\,, \tag{4.2.10}$$

$(B_t)_t$ being a d-dimensional Brownian motion and $\kappa>0$ a curvature parameter (whose value will be specified later). The above SDE admits as unique invariant measure the Gaussian distribution $\mathfrak{m}\sim\mathcal{N}(0,\kappa^{-1}\mathrm{Id})$, i.e. $\mathrm{d}\mathfrak{m}(x)\sim e^{-\frac{\kappa}{2}|x|^2}\mathrm{d}x$, which satisfies the curvature condition $\mathsf{CD}(\kappa,\infty)$ and hence a logarithmic Sobolev inequality with parameter κ^{-1} [BGL13, Corollary 5.7.1]. Then, Herbst's argument [BGL13, Proposition 5.4.1] implies that $\mathfrak{m}$ satisfies the integrability condition (I) for any $r<\kappa/2$. Notice that for any choice of $\kappa>0$ we have

$$\mathscr{H}(\mu|\mathfrak{m}) = \mathrm{Ent}(\mu) + \frac{\kappa}{2}\,M_2(\mu) + \frac{d}{2}\log\frac{2\pi}{\kappa} < +\infty \tag{4.2.11}$$

and similarly $\mathscr{H}(\nu|\mathfrak{m})<+\infty$.

Now, let us consider a final time in SP such that $\frac{\varepsilon\kappa}{4} = \sinh(\kappa T)$, i.e.

$$T := \frac{1}{\kappa}\log\left(\frac{\varepsilon\kappa}{4} + \sqrt{\frac{\varepsilon^2\kappa^2}{16}+1}\right) \tag{4.2.12}$$

and let $\mathrm{R}_{0,T} = \mathcal{L}(X_0,X_T)$ denote the joint law at times 0 and T of the Ornstein–Uhlenbeck process solving (4.2.10) started at the invariant measure $\mathfrak{m}$, whose density is given by the transition kernel

$$\mathsf{p}_T(x,y) = \frac{\mathrm{d}\mathrm{R}_{0,T}}{\mathrm{d}(\mathfrak{m}\otimes\mathfrak{m})}(x,y) = \frac{1}{(1-e^{-2\kappa T})^{\frac{d}{2}}}\,\exp\left\{-\frac{|x|^2-2e^{\kappa T}\,x\cdot y+|y|^2}{\frac{2}{\kappa}(e^{2\kappa T}-1)}\right\}\,. \tag{4.2.13}$$

Then, thanks to our choice of T, for any coupling $\pi \in \Pi(\mu,\nu)$ it holds

$$\begin{aligned}
\int_{\mathbb{R}^{2d}} |x-y|^2 \,\mathrm{d}\pi + \varepsilon\,\mathscr{H}(\pi|\mu\otimes\nu) \\
&= M_2(\mu)+M_2(\nu) - 2\int_{\mathbb{R}^{2d}} x\cdot y\,\mathrm{d}\pi + \varepsilon\,\mathscr{H}(\pi|\mu\otimes\nu)\\
&= -\frac{d\varepsilon}{2}\log(1-e^{-2\kappa T}) + (1-e^{-\kappa T})\big(M_2(\mu)+M_2(\nu)\big)\\
&\qquad - \varepsilon\int_{\mathbb{R}^{2d}} \log \mathrm{p}_T(x,y)\,\mathrm{d}\pi + \varepsilon\,\mathscr{H}(\pi|\mu\otimes\nu)\\
&= -\frac{d\varepsilon}{2}\log(1-e^{-2\kappa T}) + (1-e^{-\kappa T})\big(M_2(\mu)+M_2(\nu)\big)\\
&\qquad - \varepsilon\int_{\mathbb{R}^{2d}} \log \frac{\mathrm{dR}}{\mathrm{d}(\mathfrak{m}\otimes\mathfrak{m})}\,\mathrm{d}\pi + \varepsilon\mathscr{H}(\pi|\mu\otimes\nu)\\
&= -\frac{d\varepsilon}{2}\log(1-e^{-2\kappa T}) + (1-e^{-\kappa T})\big(M_2(\mu)+M_2(\nu)\big) + \varepsilon\,\mathscr{H}(\pi|\mathrm{R}_{0,T})\\
&\qquad - \varepsilon\,\mathscr{H}(\mu|\mathfrak{m}) - \varepsilon\,\mathscr{H}(\nu|\mathfrak{m})\,.
\end{aligned} \tag{4.2.14}$$

Therefore the above EOT problem (4.2.9) has the same minimizer of SP with reference given by the above-chosen $\mathrm{R}_{0,T}$ and their values are linked according to the following identity

$$\begin{aligned}
\mathrm{EOT}_\varepsilon^{\mathrm{d}^2}(\mu,\nu) =&\ \varepsilon\,\mathcal{C}_T(\mu,\nu) - \varepsilon\,\mathscr{H}(\mu|\mathfrak{m}) - \varepsilon\,\mathscr{H}(\nu|\mathfrak{m}) - \frac{d\varepsilon}{2}\log(1-e^{-2\kappa T})\\
&\qquad + (1-e^{-\kappa T})\big(M_2(\mu)+M_2(\nu)\big)\\
=&\ \frac{\varepsilon}{T}\,\mathrm{EOT}_T^{c_T}(\mu,\nu) - \frac{d\varepsilon}{2}\log(1-e^{-2\kappa T}) + (1-e^{-\kappa T})\big(M_2(\mu)+M_2(\nu)\big)\,.
\end{aligned} \tag{4.2.15}$$

Since we are interested in stability results, take also $\bar\mu$, $\bar\nu \in \mathcal{P}_2(\mathbb{R}^d)$ such that

$$\mathrm{Ent}(\bar\mu),\ \mathrm{Ent}(\bar\nu) < +\infty \quad \text{and} \quad \mathscr{H}^{\mathrm{sym}}(\mu,\bar\mu),\ \mathscr{H}^{\mathrm{sym}}(\nu,\bar\nu) < +\infty\,.$$

We are now ready to apply Theorem 4.2.7 and get

$$
\begin{aligned}
|\mathrm{EOT}^{\mathsf{d}^2}_{\varepsilon}(\bar\mu,\bar\nu)-\mathrm{EOT}^{\mathsf{d}^2}_{\varepsilon}(\mu,\nu)| \leq{}& \frac{\varepsilon}{T}\left|\mathrm{EOT}^{c_T}_{T}(\bar\mu,\bar\nu)-\mathrm{EOT}^{c_T}_{T}(\mu,\nu)\right| \\
&+\left(1-e^{-\kappa T}\right)\left(|M_2(\bar\mu)-M_2(\mu)|+|M_2(\bar\nu)-M_2(\nu)|\right) \\
\leq{}& \varepsilon\left[\mathscr{H}^{\mathrm{sym}}(\mu,\bar\mu)\wedge\mathscr{H}^{\mathrm{sym}}(\nu,\bar\nu)\right] \\
&+\left(1-e^{-\kappa T}\right)\left(|M_2(\bar\mu)-M_2(\mu)|+|M_2(\bar\nu)-M_2(\nu)|\right) \\
&+\frac{\varepsilon}{\sqrt{E_{2\kappa}(T)}}\sqrt{\mathcal{C}_T(\mu,\nu)-\mathscr{H}(\mu|\mathfrak{m})}\,\|\mu-\bar\mu\|_{\dot{\mathrm{H}}^{-1}(\mu)} \\
&+\frac{\varepsilon}{\sqrt{E_{2\kappa}(T)}}\sqrt{\mathcal{C}_T(\mu,\nu)-\mathscr{H}(\nu|\mathfrak{m})}\,\|\nu-\bar\nu\|_{\dot{\mathrm{H}}^{-1}(\nu)} \\
&+\frac{\varepsilon}{\sqrt{E_{2\kappa}(T)}}\sqrt{\mathcal{C}_T(\bar\mu,\bar\nu)-\mathscr{H}(\bar\mu|\mathfrak{m})}\,\|\bar\mu-\mu\|_{\dot{\mathrm{H}}^{-1}(\bar\mu)} \\
&+\frac{\varepsilon}{\sqrt{E_{2\kappa}(T)}}\sqrt{\mathcal{C}_T(\bar\mu,\bar\nu)-\mathscr{H}(\bar\nu|\mathfrak{m})}\,\|\bar\nu-\nu\|_{\dot{\mathrm{H}}^{-1}(\bar\nu)}\,.
\end{aligned}
\tag{4.2.16}
$$

Since for any coupling $\pi\in\Pi(\bar\mu,\mu)$ we can write

$$
\begin{aligned}
|M_2(\bar\mu)-M_2(\mu)| &= \left|\int |x|^2-|y|^2\,\mathrm{d}\pi\right| \\
&\leq \int |x(x-y)|\,\mathrm{d}\pi+\int |y(x-y)|\,\mathrm{d}\pi \\
&\leq \left(\sqrt{M_2(\bar\mu)}+\sqrt{M_2(\mu)}\right)\left(\int |x-y|^2\,\mathrm{d}\pi\right)^{\frac{1}{2}}\,,
\end{aligned}
$$

by minimising the right-hand side over $\pi\in\Pi(\mu,\nu)$ we end up with

$$
|M_2(\bar\mu)-M_2(\mu)|\leq\left(\sqrt{M_2(\bar\mu)}+\sqrt{M_2(\mu)}\right)\mathbf{W}_2(\bar\mu,\mu)
$$

and (4.2.16) reads as

$$
\begin{aligned}
|\mathrm{EOT}^{\mathsf{d}^2}_\varepsilon(\bar\mu,\bar\nu)-\mathrm{EOT}^{\mathsf{d}^2}_\varepsilon(\mu,\nu)| \leq & \varepsilon\Big[\mathscr{H}^{\mathrm{sym}}(\mu,\bar\mu)\wedge\mathscr{H}^{\mathrm{sym}}(\nu,\bar\nu)\Big]\\
&+\left(1-e^{-\kappa T}\right)\left(\sqrt{M_2(\bar\mu)}+\sqrt{M_2(\mu)}\right)\mathbf{W}_2(\bar\mu,\mu)\\
&+\left(1-e^{-\kappa T}\right)\left(\sqrt{M_2(\bar\nu)}+\sqrt{M_2(\nu)}\right)\mathbf{W}_2(\bar\nu,\nu)\\
&+\frac{\varepsilon}{\sqrt{E_{2\kappa}(T)}}\sqrt{\mathcal{C}_T(\mu,\nu)-\mathscr{H}(\mu|\mathfrak{m})}\,\|\mu-\bar\mu\|_{\dot{\mathrm{H}}^{-1}(\mu)}\\
&+\frac{\varepsilon}{\sqrt{E_{2\kappa}(T)}}\sqrt{\mathcal{C}_T(\mu,\nu)-\mathscr{H}(\nu|\mathfrak{m})}\,\|\nu-\bar\nu\|_{\dot{\mathrm{H}}^{-1}(\nu)}\\
&+\frac{\varepsilon}{\sqrt{E_{2\kappa}(T)}}\sqrt{\mathcal{C}_T(\bar\mu,\bar\nu)-\mathscr{H}(\bar\mu|\mathfrak{m})}\,\|\bar\mu-\mu\|_{\dot{\mathrm{H}}^{-1}(\bar\mu)}\\
&+\frac{\varepsilon}{\sqrt{E_{2\kappa}(T)}}\sqrt{\mathcal{C}_T(\bar\mu,\bar\nu)-\mathscr{H}(\bar\nu|\mathfrak{m})}\,\|\bar\nu-\nu\|_{\dot{\mathrm{H}}^{-1}(\bar\nu)}\,.
\end{aligned}
\tag{4.2.17}
$$

Similarly, since $\pi^{\mu\to\nu,T}$ is also the optimizer of the EOT problem (4.2.9), we can translate Theorem 4.2.2 into a stability result between the optimal plans for (4.2.9), that is, between $\pi^{\mu,\nu}_\varepsilon$ and $\pi^{\bar\mu,\bar\nu}_\varepsilon$:

$$
\begin{aligned}
\varepsilon\,\mathscr{H}^{\mathrm{sym}}(\pi^{\mu,\nu}_\varepsilon,\pi^{\bar\mu,\bar\nu}_\varepsilon) &= \frac{\varepsilon}{T}\,T\,\mathscr{H}^{\mathrm{sym}}(\pi^{\mu\to\nu,T},\pi^{\bar\mu\to\bar\nu,T})\\
&\leq \varepsilon\,\mathscr{H}^{\mathrm{sym}}(\mu,\bar\mu)+\varepsilon\,\mathscr{H}^{\mathrm{sym}}(\nu,\bar\nu)\\
&+\frac{\varepsilon}{\sqrt{E_{2\kappa}(T)}}\sqrt{\mathcal{C}_T(\mu,\nu)-\mathscr{H}(\mu|\mathfrak{m})}\,\|\mu-\bar\mu\|_{\dot{\mathrm{H}}^{-1}(\mu)}\\
&+\frac{\varepsilon}{\sqrt{E_{2\kappa}(T)}}\sqrt{\mathcal{C}_T(\mu,\nu)-\mathscr{H}(\nu|\mathfrak{m})}\,\|\nu-\bar\nu\|_{\dot{\mathrm{H}}^{-1}(\nu)}\\
&+\frac{\varepsilon}{\sqrt{E_{2\kappa}(T)}}\sqrt{\mathcal{C}_T(\bar\mu,\bar\nu)-\mathscr{H}(\bar\mu|\mathfrak{m})}\,\|\bar\mu-\mu\|_{\dot{\mathrm{H}}^{-1}(\bar\mu)}\\
&+\frac{\varepsilon}{\sqrt{E_{2\kappa}(T)}}\sqrt{\mathcal{C}_T(\bar\mu,\bar\nu)-\mathscr{H}(\bar\nu|\mathfrak{m})}\,\|\bar\nu-\nu\|_{\dot{\mathrm{H}}^{-1}(\bar\nu)}\,.
\end{aligned}
\tag{4.2.18}
$$

Given the above bounds one can now get explicit estimates by choosing the curvature parameter $\kappa>0$.

Remark 4.2.10 (Small-noise limit)**.** *Let us point out that the above bounds are stable when $\varepsilon\to 0$. Indeed* (4.2.12) *guarantees us that the small-noise limit $\varepsilon\to 0$ is equivalent to the small-time limit $T\to 0$ and moreover the ratio $\frac{\varepsilon}{T}$ stays finite since*

$$
\frac{\varepsilon}{T}=\frac{\varepsilon\kappa}{\kappa T}=\frac{4\sinh(\kappa T)}{\kappa T}=2\,\frac{e^{\kappa T}-e^{-\kappa T}}{\kappa T}\;\to\;4\,,\quad \textit{as } T\to 0.
$$

Therefore in the small-noise limit, from (3.2.2) *we deduce for* (4.2.9) *the following sta-*

bility results

$$\begin{aligned}\limsup_{\varepsilon\to 0} |\mathrm{EOT}^{\mathsf{d}^2}_\varepsilon(\bar\mu,\bar\nu) - \mathrm{EOT}^{\mathsf{d}^2}_\varepsilon(\mu,\nu)| &\\ \le 2\mathbf{W}_2(\mu,\nu)&\left[\|\mu-\bar\mu\|_{\dot{\mathrm{H}}^{-1}(\mu)} + \|\nu-\bar\nu\|_{\dot{\mathrm{H}}^{-1}(\nu)}\right]\\ +2\mathbf{W}_2(\bar\mu,\bar\nu)&\left[\|\bar\mu-\mu\|_{\dot{\mathrm{H}}^{-1}(\bar\mu)} + \|\bar\nu-\nu\|_{\dot{\mathrm{H}}^{-1}(\bar\nu)}\right]\\ \limsup_{\varepsilon\to 0}\, \varepsilon\,\mathscr{H}^{\mathrm{sym}}(\boldsymbol{\pi}^{\mu,\nu}_\varepsilon, \boldsymbol{\pi}^{\bar\mu,\bar\nu}_\varepsilon) \le 2\mathbf{W}_2(\mu,\nu)&\left[\|\mu-\bar\mu\|_{\dot{\mathrm{H}}^{-1}(\mu)} + \|\nu-\bar\nu\|_{\dot{\mathrm{H}}^{-1}(\nu)}\right]\\ +2\mathbf{W}_2(\bar\mu,\bar\nu)&\left[\|\bar\mu-\mu\|_{\dot{\mathrm{H}}^{-1}(\bar\mu)} + \|\bar\nu-\nu\|_{\dot{\mathrm{H}}^{-1}(\bar\nu)}\right]\end{aligned} \tag{4.2.19}$$

which agree, up to a scaling constant, with the estimates obtained in (4.2.8) *for the Schrödinger setting.*

Notice that the above small-noise limit is independent from the choice of $\kappa > 0$. This suggests us to take the limit $\kappa \downarrow 0$ directly in (4.2.17) and (4.2.18). Indeed from (4.2.12) it follows $T \to \varepsilon/4$ as $\kappa \downarrow 0$. Furthermore, by rearranging (4.2.15) we notice that

$$\begin{aligned}T\,\mathcal{C}_T(\mu,\nu) - T\,\mathscr{H}(\mu|\mathfrak{m}) &= \frac{T}{\varepsilon}\,\mathrm{EOT}^{\mathsf{d}^2}_\varepsilon(\mu,\nu) - \frac{T}{\varepsilon}(1-e^{-\kappa T})\Big(M_2(\mu)+M_2(\nu)\Big)\\ &\qquad + T\,\mathscr{H}(\nu|\mathfrak{m}) + \frac{d\,T}{2}\log\left(1-e^{-2\kappa T}\right)\\ &= \frac{T}{\varepsilon}\,\mathrm{EOT}^{\mathsf{d}^2}_\varepsilon(\mu,\nu) - \frac{T}{\varepsilon}(1-e^{-\kappa T})\Big(M_2(\mu)+M_2(\nu)\Big) + T\,\mathrm{Ent}(\nu)\\ &\qquad + \frac{\kappa T}{2}M_2(\nu) + \frac{d\,T}{2}\log\left(4\pi T\,\frac{1-e^{-2\kappa T}}{2\kappa T}\right)\end{aligned}$$

and therefore we have

$$\lim_{\kappa\to 0}\, T\,\mathcal{C}_T(\mu,\nu) - T\,\mathscr{H}(\mu|\mathfrak{m}) = \frac{1}{4}\,\mathrm{EOT}^{\mathsf{d}^2}_\varepsilon(\mu,\nu) + \frac{\varepsilon}{4}\,\mathrm{Ent}(\mu) + \frac{1}{4}\,C_\varepsilon\,,$$

where $C_\varepsilon = \frac{d\varepsilon}{2}\log(4\pi\varepsilon)$. As a consequence, by taking the limit in the right-hand sides of (4.2.17) and (4.2.18), we deduce that

$$\begin{aligned}\left|\mathrm{EOT}_\varepsilon^{\mathsf{d}^2}(\bar\mu,\bar\nu)-\mathrm{EOT}_\varepsilon^{\mathsf{d}^2}(\mu,\nu)\right| &\le \varepsilon\Big[\mathscr{H}^{\mathrm{sym}}(\mu,\bar\mu)\wedge\mathscr{H}^{\mathrm{sym}}(\nu,\bar\nu)\Big]\\
&+2\sqrt{\mathrm{EOT}_\varepsilon^{\mathsf{d}^2}(\mu,\nu)+\varepsilon\,\mathrm{Ent}(\nu)+C_\varepsilon}\,\|\mu-\bar\mu\|_{\dot{\mathrm{H}}^{-1}(\mu)}\\
&+2\sqrt{\mathrm{EOT}_\varepsilon^{\mathsf{d}^2}(\mu,\nu)+\varepsilon\,\mathrm{Ent}(\mu)+C_\varepsilon}\,\|\nu-\bar\nu\|_{\dot{\mathrm{H}}^{-1}(\nu)}\\
&+2\sqrt{\mathrm{EOT}_\varepsilon^{\mathsf{d}^2}(\bar\mu,\bar\nu)+\varepsilon\,\mathrm{Ent}(\bar\nu)+C_\varepsilon}\,\|\bar\mu-\mu\|_{\dot{\mathrm{H}}^{-1}(\bar\mu)}\\
&+2\sqrt{\mathrm{EOT}_\varepsilon^{\mathsf{d}^2}(\bar\mu,\bar\nu)+\varepsilon\,\mathrm{Ent}(\bar\mu)+C_\varepsilon}\,\|\bar\nu-\nu\|_{\dot{\mathrm{H}}^{-1}(\bar\nu)}\end{aligned}$$

and

$$\begin{aligned}\varepsilon\,\mathscr{H}^{\mathrm{sym}}(\boldsymbol{\pi}_\varepsilon^{\mu,\nu},\boldsymbol{\pi}_\varepsilon^{\bar\mu,\bar\nu}) &\le \varepsilon\left[\mathscr{H}^{\mathrm{sym}}(\mu,\bar\mu)+\mathscr{H}^{\mathrm{sym}}(\nu,\bar\nu)\right]\\
&+2\sqrt{\mathrm{EOT}_\varepsilon^{\mathsf{d}^2}(\mu,\nu)+\varepsilon\,\mathrm{Ent}(\nu)+C_\varepsilon}\,\|\mu-\bar\mu\|_{\dot{\mathrm{H}}^{-1}(\mu)}\\
&+2\sqrt{\mathrm{EOT}_\varepsilon^{\mathsf{d}^2}(\mu,\nu)+\varepsilon\,\mathrm{Ent}(\mu)+C_\varepsilon}\,\|\nu-\bar\nu\|_{\dot{\mathrm{H}}^{-1}(\nu)}\\
&+2\sqrt{\mathrm{EOT}_\varepsilon^{\mathsf{d}^2}(\bar\mu,\bar\nu)+\varepsilon\,\mathrm{Ent}(\bar\nu)+C_\varepsilon}\,\|\bar\mu-\mu\|_{\dot{\mathrm{H}}^{-1}(\bar\mu)}\\
&+2\sqrt{\mathrm{EOT}_\varepsilon^{\mathsf{d}^2}(\bar\mu,\bar\nu)+\varepsilon\,\mathrm{Ent}(\bar\mu)+C_\varepsilon}\,\|\bar\nu-\nu\|_{\dot{\mathrm{H}}^{-1}(\bar\nu)}\ .\end{aligned}$$

Notice that from the above bounds, we get the validity of (4.2.19) in the small-noise limit $\varepsilon\downarrow 0$,.

Bibliographical Remarks

Recently, there has been an increasing interest in the quantitative stability for the EOT problem, which is strongly linked to SP. A first stability estimate for EOT in a general setting has been established in [GNB22]. There, the authors, without any integrability assumption, manage to prove a qualitative stability result by relying on a geometric notion inspired by the cyclical monotonicity property in Optimal Transport. To the best of our knowledge, the first quantitative stability result is due to Carlier and Laborde in [CL20] in the context of multi-marginal EOT. By considering bounded marginals equivalent to a common reference probability measure and a bounded cost, they show that the potentials are Lipschitz-continuous in L^2 and L^∞ w.r.t. the densities of the marginals. As far as concerns quantitative stability for the primal optimiser, the first result appeared in the work [DdBD24]. There the authors prove on compact metric spaces a quantitative uniform stability along Sinkhorn's algorithm which implies stability for EOT and more precisely an explicit $\mathbf{W}_1$-Lipschitzianity of the optimiser w.r.t. the marginals. More recently, in [EN22b] a quantitative stability result that holds on general metric spaces is shown, thus removing the compactness assumption at the cost of requiring some exponential integrability with respect to the marginals (condition met for instance when the marginals are sub-Gaussian). More precisely the authors prove a $\frac{1}{2p}$-Hölder continuity for $\mathbf{W}_p$ provided that the cost function satisfies an abstract condition, introduced there in order to bound the optimal values of EOT by means of the Wasserstein distance between the corresponding optimiser. Such condition is met for a wide enough class of cost functions such as $c(x,y) = |x-y|^p$ on the Euclidean space, with $p \in (1,\infty)$. Nevertheless, it is not easily verifiable in general, and in particular when considering the case $c = -T \log \mathsf{p}_T$ (which allows to translate results from EOT into results for SP). Lastly, it is worth mentioning that in the most recent [NW23] the authors study the qualitative stability of the Schrödinger potentials associated to EOT in a general setting assuming the cost function c satisfies $e^{\beta c} \in L^1(\mu \otimes \nu)$ for some $\beta > 0$, which leads to the convergence of Sinkhorn's algorithm. The (CD) condition does not have a natural counterpart in the EOT setting, where the results we have just discussed have been established; for this reason the stability results Theorem 4.2.2 and Theorem 4.2.7 are not implied by any of the stability bounds mentioned above.

The results presented in this chapter are based on [CCGT23].

This page was intentionally left blank.

Chapter 5

Exponential convergence of Sinkhorn's algorithm: perturbative approach

In this chapter we are going to study how Lipschitzianity propagates along Sinkhorn's algorithm and from that deduce a first exponential convergence result. The approach we employ here is based on Stochastic Optimal Control theory and Hamilton-Jacobi-Bellman equations (HJBs).

Hereafter we are going to consider a Schrödinger problem on $\mathbb{R}^d$ with reference measure induced by the Langevin SDE

$$\begin{cases} \mathrm{d}X_t = -\nabla U(X_t)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}B_t \\ X_0 \sim \mathfrak{m}\,, \end{cases} \tag{5.0.1}$$

with $\mathfrak{m}(\mathrm{d}x) \propto \exp(-U(x))\mathrm{d}x$. We will further assume that the potential $U \in \mathcal{C}^2(\mathbb{R}^d)$ is strongly convex, *i.e.*, that $(\mathbb{R}^d, \mathfrak{m})$ satisfies $\mathsf{CD}(\kappa, \infty)$ for some positive $\kappa > 0$. This condition can be further relaxed by considering potentials U that are asymptotically convex (*i.e.*, satisfying (6.0.10) in the next chapter), however for clarity purposes we stick here with the strong convexity assumption. We refer the reader to Remark 5.2.4.

Let us further mention that $\mathsf{CD}(\kappa, \infty)$ with $\kappa > 0$ guarantees that $\mathfrak{m} \in \mathcal{P}_p(\mathbb{R}^d)$ for any finite $p \geq 1$, since $\mathfrak{m}$ satisfies the Poincaré inequality with parameter κ^{-1} [BGL13, Proposition 4.8.1], and therefore it enjoys exponential integrability properties (namely $\exp(s|x|) \in \mathrm{L}^1(\mathfrak{m})$ for all $s < 2\sqrt{\kappa}$, see [BGL13, Proposition 4.4.2]).

The standing marginals' assumption in this chapter will be the following:

A4. *The two marginals $\mu, \nu \in \mathcal{P}_1(\mathbb{R}^d)$ admit continuously differentiable Lipschitz log-densities w.r.t. $\mathfrak{m}$, i.e., there exist two Lipschitz potentials U_μ, $U_\nu \in \mathcal{C}^1(\mathbb{R}^d)$ such* *that*

$$\mu(\mathrm{d}x) = \exp(-U_\mu(x))\mathfrak{m}(\mathrm{d}x)\,, \quad \nu(\mathrm{d}x) = \exp(-U_\nu(x))\mathfrak{m}(\mathrm{d}x)\,.$$

Clearly **A**4 implies **A**1 since for $\mathfrak{p} \in \{\mu, \nu\}$ we have

$$\mathscr{H}(\mathfrak{p}|\mathfrak{m}) = -\int U_{\mathfrak{p}}\,\mathrm{d}\mathfrak{p} \le -U_{\mathfrak{p}}(0) + \mathrm{Lip}(U_{\mathfrak{p}})\,M_1(\mathfrak{p}) < +\infty\,.$$

From this we may also deduce that $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$. Indeed, the probability measure $\mathfrak{m}$ under $\mathsf{CD}(\kappa,\infty)$ satisfies the Talagrand transportation inequality with parameter κ^{-1} [BGL13, Corollary 9.3.2] and therefore we can conclude that

$$M_2(\mathfrak{p}) \le 2\,M_2(\mathfrak{m}) + 2\,\mathbf{W}_2^2(\mathfrak{p},\mathfrak{m}) \le 2\,M_2(\mathfrak{m}) + \frac{4}{\kappa}\,\mathscr{H}(\mathfrak{p}|\mathfrak{m}) < +\infty\,.$$

Finally, let us point out that in this chapter and in the following one we will denote with $\varphi^\star$ and $\psi^\star$ the Schrödinger potentials defined in Theorem 2.2.1. The only purpose of this slightly different notation boils down to avoiding any possible confusion between Sinkhorn's iterates and potentials.

5.1 Lipschitz propagation along Sinkhorn's algorithm

The starting point of our discussion is considering the function

$$\mathcal{U}_t^{T,h} := -\log P_{T-t}\exp(-h)\,, \tag{5.1.1}$$

where $(P_t)_{t\in[0,T]}$ is the semigroup associated to (5.0.1) while h is a given Lipschitz function. In order to do that let us fix an underlying filtered probability space $(\Omega, (\mathcal{F}_s)_{s\in[0,T]}, \mathcal{F}, \mathbb{P})$ satisfying the usual conditions and endowed with the Brownian motion $(B_t)_{t\in[0,T]}$. Under some additional regularity assumption (namely, $h \in \mathcal{C}^3_{\mathrm{Lip}}(\mathbb{R}^d)$) it is known that $\mathcal{U}_t^{T,h}$ is a classical solution of the HJB equation

$$\begin{cases} \partial_t u_t + \Delta u_t - \nabla U \cdot \nabla u_t - |\nabla u_t|^2 = 0 \\ u_T = h\,, \end{cases} \tag{5.1.2}$$

and it coincides also with the the value function of the stochastic optimal control problem

$$\mathcal{J}_t^{T,h}(x) = \inf_{q_\cdot \in \mathcal{A}_{[t,T]}} \mathbb{E}\left[\int_t^T |q_s|^2 \mathrm{d}s + h(X_T^q)\right]$$
$$\text{where } \mathbb{P}\text{-a.s. it holds } \begin{cases} \mathrm{d}X_s^q = (-\nabla U(X_s^q) + 2\,q_s)\mathrm{d}s + \sqrt{2}\,\mathrm{d}B_s \\ X_t^q = x \end{cases} \tag{5.1.3}$$

where $\mathcal{A}_{[t,T]}$ denotes the set of admissible controls, *i.e.*, progressively measurable processes with finite moments on $(\Omega, (\mathcal{F}_s)_{s\in[0,T]}, \mathcal{F}, \mathbb{P})$. Moreover, the optimal control is a feedback-control process equal to $-\nabla\mathcal{U}_s^{T,h}(X_s^q)$. These are classic results in Stochastic Control theory and we refer the reader to [Con23, Proposition 3.1] for their validity.

Given the above, we can now show how Lipschitzianity backward propagates along HJB equations.

Lemma 5.1.1. *Let $h \in \mathrm{Lip}(\mathbb{R}^d)$ be a Lipschitz function and assume $\mathsf{CD}(\kappa,\infty)$ for $\kappa > 0$. Then for any $t \in [0,T]$ it holds*

$$\mathrm{Lip}(\mathcal{U}_t^{T,h}) \leq \exp(-\kappa\,(T-t))\,\mathrm{Lip}(h)\,. \tag{5.1.4}$$

Proof. We will firstly prove this result under the additional smoothness assumption that $h \in \mathcal{C}^3_{\mathrm{Lip}}(\mathbb{R}^d)$, so that the above stochastic optimal control (hereafter SOC) representation holds true.

Fix $t \in [0,T)$, $x, y \in \mathbb{R}^d$ and consider $\boldsymbol{q}_s := -\nabla\mathcal{U}_s^{T,h}(X_s^{\boldsymbol{q}})$ the optimal control process associated to the value function $\mathcal{U}_s^{T,h}(x)$ evaluated in x, via (5.1.3) and its corresponding controlled process

$$\begin{cases} \mathrm{d}X_s^{\boldsymbol{q}} = (-\nabla U(X_s^{\boldsymbol{q}}) + 2\,\boldsymbol{q}_s)\mathrm{d}s + \sqrt{2}\,\mathrm{d}B_s \\ X_t^{\boldsymbol{q}} = x\,. \end{cases} \tag{5.1.5}$$

Let us further consider the diffusion process

$$\begin{cases} \mathrm{d}Y_s = (-\nabla U(Y_s) + 2\,\boldsymbol{q}_s)\mathrm{d}s + \sqrt{2}\,\mathrm{d}B_s \\ Y_t = y \end{cases}$$

with $(B_s)_{s\in[0,T]}$ being the same Brownian motion as in (5.1.5), so that we consider a synchronous coupling between the diffusion processes $(X_s^{\boldsymbol{q}})_{s\in[t\,T]}$ and $(Y_s)_{s\in[t,T]}$.

Then, from the suboptimality of $(\boldsymbol{q}_s)_{s\in[t,T]}$ for the value function evaluated in y (that is when considering the diffusion (5.1.3) started at $X_t^q = y$) we know that

$$\begin{aligned} \mathcal{U}_t^{T,h}(y) - \mathcal{U}_t^{T,h}(x) &\leq \mathbb{E}\left[\int_t^T |\boldsymbol{q}_s|^2\mathrm{d}s + h(Y_T)\right] - \mathbb{E}\left[\int_t^T |\boldsymbol{q}_s|^2\mathrm{d}s + h(X_T^{\boldsymbol{q}})\right] \\ &= \mathbb{E}[h(Y_T) - h(X_T^{\boldsymbol{q}})] \leq \mathrm{Lip}(h)\,\mathbb{E}[|Y_T - X_T^{\boldsymbol{q}}|]\,. \end{aligned} \tag{5.1.6}$$

Therefore it is enough bounding the distance between $X_T^{\boldsymbol{q}}$ and Y_T which are two processes started respectively in $X_t^{\boldsymbol{q}} = x$ and $Y_t = y$ and that follow the same controlled SDE. Clearly their difference satisfies

$$\mathrm{d}(X_s^{\boldsymbol{q}} - Y_s) = -(\nabla U(X_s^{\boldsymbol{q}}) - \nabla U(Y_s))\mathrm{d}s\,,$$

from which it follows

$$\mathrm{d}|X_s^{\boldsymbol{q}} - Y_s|^2 = -2\langle X_s^{\boldsymbol{q}} - Y_s,\, \nabla U(X_s^{\boldsymbol{q}}) - \nabla U(Y_s)\rangle\,\mathrm{d}s \le -2\kappa\,|X_s^{\boldsymbol{q}} - Y_s|^2\,\mathrm{d}s\,,$$

where the last step follows from the κ-convexity of U. Gronwall Lemma finally yields to

$$|Y_T - X_T^{\boldsymbol{q}}| \le \exp(-\kappa(T-t))|y-x|\,.$$

By taking the expectation in the above estimate and combining it with (5.1.6), we conclude that for any $x, y \in \mathbb{R}^d$ it holds

$$|\mathcal{U}_t^{T,h}(y) - \mathcal{U}_t^{T,h}(x)| \le \exp(-\kappa(T-t))\,\mathrm{Lip(h)}\,|y-x|\,,$$

which is equivalent to our thesis.

Finally, this result can be relaxed to the general case $h \in \mathrm{Lip}(\mathbb{R}^d)$ by following a standard approximation technique, which is detailed in [Con23, Lemma 3.1]. □

The above estimate immediately allow to propagate Lipschitzianity along Sinkhorn's algorithm (cf. (2.2.18)) which we recall here to be defined as

$$\begin{cases}\varphi^{n+1} = U_\mu - \mathcal{U}_0^{T,\psi^n} \\ \psi^{n+1} = U_\nu - \mathcal{U}_0^{T,\varphi^{n+1}}\,,\end{cases} \tag{5.1.7}$$

with $\mathcal{U}_t^{T,h}$ as defined in (5.1.1). We further initialise the algorithm at ψ^0, being Lipschitz. Notice that the standard initialisation $\psi^0 = U_\nu$ is indeed Lipschitz under **A**4. In order to be consistent with the normalisation imposed on the Schrödinger potentials at (2.2.13), when dealing with the convergence of Sinkhorn's iterates we will actually refer to their normalised versions, namely the iterates

$$\varphi^{\diamond n} = \varphi^n - \left(\int \varphi^n \mathrm{d}\mu - \int \varphi^\star \mathrm{d}\mu\right), \qquad \psi^{\diamond n} = \psi^n - \left(\int \psi^n \mathrm{d}\nu - \int \psi^\star \mathrm{d}\nu\right), \tag{5.1.8}$$

so that

$$\int \varphi^{\diamond n}\mathrm{d}\mu = \int \varphi^\star \mathrm{d}\mu \quad\text{and}\quad \int \psi^{\diamond n}\mathrm{d}\nu = \int \psi^\star \mathrm{d}\nu\,. \tag{5.1.9}$$

Lemma 5.1.2. *Assume* $\mathsf{CD}(\kappa,\infty)$ *for some* $\kappa > 0$, *that the two marginals satisfy* ***A*** *4 and further assume the initialisation of Sinkhorn's algorithm* ψ^0 *to be a Lipschitz function. Then, for all* $n \ge 0$ *we have*

$$\begin{aligned}\mathrm{Lip}(\varphi^{n+1}) &\le \mathrm{Lip}(U_\mu) + \mathrm{e}^{-\kappa T}\mathrm{Lip}(\psi^n) \\ \mathrm{Lip}(\psi^{n+1}) &\le \mathrm{Lip}(U_\nu) + \mathrm{e}^{-\kappa T}\mathrm{Lip}(\varphi^{n+1})\,.\end{aligned} \tag{5.1.10}$$

Moreover, for all $n \ge 1$ *we have*

$$\begin{aligned}\mathrm{Lip}(\psi^n) &\le \frac{\mathrm{Lip}(U_\nu) + \exp(-\kappa T)\mathrm{Lip}(U_\mu)}{1-\exp(-2\kappa T)} \\ \mathrm{Lip}(\varphi^n) &\le \frac{\mathrm{Lip}(U_\mu) + \exp(-\kappa T)\mathrm{Lip}(U_\nu)}{1-\exp(-2\kappa T)}\,,\end{aligned} \tag{5.1.11}$$

and the same holds for the normalised $\varphi^{\diamond n}$, $\psi^{\diamond n}$.

Proof. As shown in Proposition Lemma 5.1.1, the Lipschitz-regularity backward propagates with rate $\kappa > 0$ along solutions of HJB equations (cf. (5.1.4)). The first claim (5.1.10) then follows from the triangular inequality of the $\mathrm{Lip}(\cdot)$ operator.

Concatenating the bounds in (5.1.10) yields to

$$\mathrm{Lip}(\psi^{n+1}) \leq \mathrm{Lip}(U_\nu) + \mathrm{e}^{-\kappa T}\mathrm{Lip}(U_\mu) + \mathrm{e}^{-2\kappa T}\mathrm{Lip}(\psi^n),$$

from which the first relation in (5.1.11) follows by induction. The second relation follows by symmetry. Finally the same statement holds true for $\varphi^{\diamond n}$, $\psi^{\diamond n}$ since they differ from the original iterates just by an additive constant. □

From the pointwise convergence of Sinkhorn's iterates φ^n, ψ^n towards the Schrödinger potentials [GN22, Corollary 4.8][1], the previous regularity result propagates to the potentials.

Corollary 5.1.3. *Assume* $\mathsf{CD}(\kappa,\infty)$ *for some* $\kappa > 0$*, that the two marginals satisfy* **A4** *and further assume the initialisation of Sinkhorn's algorithm* ψ^0 *to be a Lipschitz function. Then it holds*

$$\mathrm{Lip}(\psi^\star) \leq \frac{\mathrm{Lip}(U_\nu) + \exp(-\kappa T)\mathrm{Lip}(U_\mu)}{1-\exp(-2\kappa T)}$$
$$\mathrm{Lip}(\varphi^\star) \leq \frac{\mathrm{Lip}(U_\mu) + \exp(-\kappa T)\mathrm{Lip}(U_\nu)}{1-\exp(-2\kappa T)},$$

5.2 A first exponential convergence result

We are now ready to prove the key contraction estimates that will imply the exponential convergence of Sinkhorn's algorithm. Once again the main idea behind our proof is relying on a stochastic control problem where the Schrödinger potential contributes in the final cost while its gradient drives the controlled SDE. This allows to back-propagate along an HJB equation the Lipschitz regularity of the difference between the Sinkhorn iterates and the target Schrödinger potential. Indeed, if we denote with $\mathcal{D}^n_t := \mathcal{U}^{T,\psi^n}_t - \mathcal{U}^{T,\psi^\star}_t$ (the difference between the evolution along HJB of ψ^n and respectively the evolution of $\psi^\star$) from (5.1.2) we deduce that $\mathcal{D}^n_t$ solves

$$\begin{cases} \partial_t u_t + \Delta u_t + (-\nabla U - 2\,\nabla \mathcal{U}^{T,\psi^\star}_t)\cdot \nabla u_t - |\nabla u_t|^2 = 0 \\ u_T = \psi^n - \psi^\star\,, \end{cases} \tag{5.2.1}$$

[1] More precisely, [GN22, Corollary 4.8] implies μ-a.s. convergence of $\varphi^{\diamond n}$ towards $\varphi^\star$. Since the potentials are defined solely almost surely, we may tacitly assume this convergence to be pointwise. Similarly it holds for the convergence of $\psi^{\diamond n}$ towards $\psi^\star$. We also refer the interested reader to [Nut21, Theorem 6.15] for a similar result under stronger assumptions.

which can as well be represented as the value function of the stochastic control problem

$$\mathcal{D}^n_t(x) = \inf_{q_\cdot \in \mathcal{A}_{[t,T]}} \mathbb{E}\left[\int_t^T |q_s|^2 \mathrm{d}s + \psi^n(X^q_T) - \psi^\star(X^p_T)\right]$$
$$\text{where} \quad \begin{cases} \mathrm{d}X^q_s = (-\nabla U(X^q_s) - 2\,\nabla \mathcal{U}^{T,\psi^\star}_s(X^q_s) + 2\,q_s)\mathrm{d}s + \sqrt{2}\,\mathrm{d}B_s \\ X^q_t = x\,. \end{cases} \tag{5.2.2}$$

The connection between (5.2.1) and (5.2.2) is analogous to the one presented in the previous section and can be established as in [Con23, Proposition 3.1] (thanks to the uniform bound we provide below in Corollary 5.2.1).

Once the connection with the stochastic optimal control formulation is established, the proof boils down once again in studying how Lipschitz-regularity backward propagates along solutions of HJB equations. However here we may encounter two difficulties: the drift of the underlying control problem is time dependent and most importantly the drift is not strongly convex anymore (due to the presence of $-2\,\nabla\mathcal{U}^{T,\psi^\star}_s$). Nevertheless we will see that the Lipschitz estimates provided in Corollary 5.1.3 imply the asymptotic weak convexity of this new drift, which is enough in order to get Lipschitz backward propagation as in Lemma 5.1.1.

For this reason let us introduce the integrated convexity profile associated to a drift $(b_s)_{s\in[0,T]}$ as the function

$$\kappa_b(r) := \inf_{s\in[0,T]} \inf\left\{-\frac{\langle b_s(x) - b_s(y),\, x-y\rangle}{|x-y|^2}\ :\ |x-y| = r\right\}. \tag{5.2.3}$$

The function κ_b is often employed to quantify ergodicity of stochastic differential equations whose drift field is b, see [Ebe16] and our proof of Proposition 5.2.3 below. The term integrated convexity profile is motivated by the observation that if we consider a time-homogeneous drift induced by a potential (*i.e.*, $b = -\nabla U$), then $\kappa_U(r) \geq \alpha$ if and only if for any $x, v \in \mathbb{R}^d$, $|v| = 1$, $\int_0^r \langle \nabla^2 U(x+hv)v, v\rangle \mathrm{d}h \geq \alpha r$, which can be seen as an averaged convexity condition.

In this chapter we are going to consider the reference drift associated to the SOC problem (5.2.2), that is $b_s := -\nabla U - 2\,\nabla\mathcal{U}^{T,\psi^\star}_s$. Then, as a corollary of the discussion of the previous section we know that

Corollary 5.2.1. *Assume* $\mathsf{CD}(\kappa,\infty)$ *for some* $\kappa > 0$*, that the two marginals satisfy* **A**4 *and further assume the initialisation of Sinkhorn's algorithm* ψ^0 *to be a Lipschitz function. Then for any* $r > 0$ *it holds*

$$\kappa_b(r) \geq \kappa - \frac{8}{r}\,\frac{\mathrm{Lip}(U_\nu) \vee \mathrm{Lip}(U_\mu)}{1 - \exp(-2\kappa\,T)} \tag{5.2.4}$$

Proof. From $\mathsf{CD}(\kappa,\infty)$ we immediately deduce that for any $r>0$ it holds

$$\kappa_b(r) \geq \kappa + 2 \inf_{s\in[0,T]} \inf\left\{ \frac{\langle \nabla\mathcal{U}_s^{T,\psi^\star}(x) - \nabla\mathcal{U}_s^{T,\psi^\star}(y),\, x-y\rangle}{|x-y|^2} \,:\, |x-y|=r \right\}. \tag{5.2.5}$$

Then, Corollary 5.1.3 combined with Lemma 5.1.1 guarantees that uniformly in $x\in\mathbb{R}^d$ it holds

$$\|\nabla\mathcal{U}_s^{T,\psi^\star}(x)\|_2 \leq \mathrm{Lip}(\mathcal{U}_s^{T,\psi^\star}) \leq \exp(-\kappa(T-s))\frac{\mathrm{Lip}(U_\nu)+\exp(-\kappa\, T)\mathrm{Lip}(U_\mu)}{1-\exp(-2\kappa\, T)}\,,$$

which combined with (5.2.5) implies the thesis. □

For notation's sake from now on we let

$$L := 8\,\frac{\mathrm{Lip}(U_\nu)\vee\mathrm{Lip}(U_\mu)}{1-\exp(-2\kappa\, T)} \quad \text{and} \quad \bar\kappa(r) := \kappa - L/r\,. \tag{5.2.6}$$

As we have already mentioned the integrated convexity profile, more precisely its lower-bound $\bar\kappa$, plays a crucial role when proving the ergodicity of the (uncontrolled) SDE (5.2.2). Its core properties are that

$$\liminf_{r\to\infty}\bar\kappa(r) > 0 \quad \text{and} \quad \int_0^1 r\,\bar\kappa(r)^-\,\mathrm{d}r < +\infty\,, \tag{5.2.7}$$

where $\bar\kappa(r)^- := \max\{-\bar\kappa(r),0\}$. Indeed these guarantee the validity of the construction of a concave function ρ, which defines a distorted metric (yet equivalent to the standard Euclidean one) in which is possible obtaining Lipschitz estimates similar to Lemma 5.1.1 via the coupling by reflection technique [Wan94, Ebe16]. More precisely we have the following result.

Proposition 5.2.2 (Proposition 2.1 in [Con23])**.** *There exist a strictly increasing concave function $\rho\colon[0,+\infty)\to[0,+\infty)$, a positive rate λ and a positive constant C such that*

1. *it holds*

$$C\,r\leq\rho(r)\leq r \quad \textit{and} \quad C\leq\rho'(r)\leq 1 \quad \forall\, r\in[0,+\infty)$$

2. *for any $r>0$ the following differential inequality holds*

$$4\,\rho''(r) - r\,\rho'(r)\,\bar\kappa(r) \leq -\lambda\,\rho(r)\,. \tag{5.2.8}$$

We postpone the proof of this result to Section 5.A, where we also provide explicit estimates for the rate λ and the constant C. Proposition 5.2.2 holds for any integrated convexity profile $\bar\kappa$ satisfying (5.2.7), but in the following discussion we are just interested in the choice $\bar\kappa$ as in (5.2.6). Particularly, under this choice it holds

$$C := \exp(-L^2/_{8\kappa})/2 \tag{5.2.9}$$

and we show in Section 5.A.1 that we can always bound the rate λ with

$$\lambda \geq \frac{2\,\kappa^2}{L^2 + L\,\sqrt{8\,\kappa} + 4\,\kappa} e^{-L^2/8\kappa}\,. \tag{5.2.10}$$

The above proposition allows us to consider a distorted metrics induced by the concave function ρ. For this reason in the following result instead of considering the Lip-norm, we are going to consider Lipschitzianity according to this distorted metrics, namely the ρ-Lipschitz norm $\|\cdot\|_\rho$ defined as

$$\|\phi\|_\rho := \sup_{x \neq y \in \mathbb{R}^d} \frac{|\phi(x) - \phi(y)|}{\rho(|x-y|)}\,.$$

Let us further mention that the previous norm is equivalent to the usual Lipschitz norm Lip (*i.e.*, with ρ being the identity) since the function ρ built in the appendix is equivalent to the identity (cf. Proposition 5.2.2 and (5.2.9)):

$$\|\phi\|_{\mathrm{Lip}} \leq \|\phi\|_\rho \leq 2\, e^{L^2/8\kappa}\, \|\phi\|_{\mathrm{Lip}}\,. \tag{5.2.11}$$

We are now ready to show how Lipschitzianity backward propagates along the HJB equation

$$\begin{cases} \partial_t u_t + \Delta u_t + (-\nabla U - 2\,\nabla \mathcal{U}_t^{T,\psi^\star}) \cdot \nabla u_t - |\nabla u_t|^2 = 0 \\ u_T = h\,, \end{cases}$$

associated to the stochastic control problem

$$\begin{aligned} &\inf_{q_\cdot \in \mathcal{A}_{[t,T]}} \mathbb{E}\left[\int_t^T |q_s|^2 \mathrm{d}s + h(X_T^p)\right] \\ &\text{where} \quad \begin{cases} \mathrm{d}X_s^q = (b_s(X_s^q) + 2\,q_s)\mathrm{d}s + \sqrt{2}\,\mathrm{d}B_s \\ X_t^q = x\,, \end{cases} \end{aligned} \tag{5.2.12}$$

with $b_s := -\nabla U - 2\,\nabla \mathcal{U}_s^{T,\psi^\star}$. The following result is similar to Lemma 5.1.1 with the main difference being the use of coupling by reflection techniques instead of the synchronous coupling.

Proposition 5.2.3. *Let $h \in \mathrm{Lip}(\mathbb{R}^d)$ be a Lipschitz function and take (ρ, λ, C) as in Proposition 5.2.2. Then, if $\mathcal{D}_\cdot^{T,h}$ denotes the solution of* (5.2.12), *for any $t \in [0,T]$ it holds*

$$\|\mathcal{D}_t^{T,h}\|_\rho \leq \exp(-\lambda\,(T-t))\,\|h\|_\rho\,.$$

Proof. As we did for Lemma 5.1.1, we may further assume that $h \in \mathcal{C}^3_{\mathrm{Lip}}(\mathbb{R}^d)$ as the general case follows by approximation. Under this extra assumption

[Con23, Proposition 3.1] guarantees that $\mathcal{D}_t^{T,h}(x)$ coincides with the value function of (5.2.12) and that the optimal control is the feedback control process $\boldsymbol{q}_s := -\nabla\mathcal{D}_s^{T,h}(X_s^{\boldsymbol{q}})$ with $X^{\boldsymbol{q}}_{\cdot}$ strong solution of

$$\begin{cases} \mathrm{d}X_s^{\boldsymbol{q}} = (b_s(X_s^{\boldsymbol{q}}) + 2\,\boldsymbol{q}_s)\mathrm{d}s + \sqrt{2}\,\mathrm{d}B_s \\ X_t^{\boldsymbol{q}} = x\,. \end{cases}$$

Next consider the same control for the diffusion process

$$\begin{cases} \mathrm{d}Y_s = (b_s(Y_s) + 2\,\boldsymbol{q}_s)\mathrm{d}s + \sqrt{2}\,\mathrm{d}\hat{B}_s \quad \forall\, s \in [t,\tau) \text{ and } Y_s = X_s^{\boldsymbol{q}} \quad \forall\, s \in [\tau, T] \\ Y_t = y\,, \end{cases}$$

where $\tau := \inf\{s \geq t : Y_s = X_s^{\boldsymbol{q}}\} \wedge T$, and $(\hat{B}_s)_{s\geq t}$ is the reflected Brownian motion, defined as

$$\mathrm{d}\hat{B}_s := (\mathrm{I} - 2\, e_s\, e_s^{\mathsf{T}}\, \mathbf{1}_{\{s<\tau\}})\,\mathrm{d}B_s \qquad \text{where} \quad e_s := \begin{cases} \frac{Z_s}{|Z_s|} & \text{when } r_s > 0\,, \\ u & \text{when } r_s = 0\,. \end{cases}$$

where $Z_s := X_s^{\boldsymbol{q}} - Y_s$, $r_s := |Z_s|$ and $u \in \mathbb{R}^d$ is a fixed (arbitrary) unit-vector. By Lévy's characterisation, $(\hat{B}_s)_{s\geq t}$ is a d-dimensional Brownian motion. As a result of that and from the suboptimality of $\boldsymbol{q}_{\cdot}$ as control process for $\mathcal{D}_t^{T,h}(y)$ we deduce that

$$\mathcal{D}_t^{T,h}(y) - \mathcal{D}_t^{T,h}(x) \leq \mathbb{E}[h(Y_T) - h(X_T^{\boldsymbol{q}})] \leq \|h\|_\rho\, \mathbb{E}[\rho(r_T)]\,. \tag{5.2.13}$$

In addition to that, let us notice that $\mathrm{d}W_s := e_s^{\mathsf{T}}\mathrm{d}B_s$ is a one-dimensional Brownian motion and that for any $s < \tau$ it holds

$$\mathrm{d}Z_s = (b_s(X_s^{\boldsymbol{q}}) - b_s(Y_s))\,\mathrm{d}s + 2\sqrt{2}\, e_s\,\mathrm{d}W_s\,.$$

An application of Ito's formula proves then that for any $s < \tau$ it holds

$$\begin{aligned} \mathrm{d}r_s^2 &= 2\,\langle b_s(X_s^{\boldsymbol{q}}) - b_s(Y_s),\, Z_s\rangle\,\mathrm{d}s + 8\,\mathrm{d}s + 4\sqrt{2}\, r_s\,\mathrm{d}W_s \\ \mathrm{d}r_s &= \langle b_s(X_s^{\boldsymbol{q}}) - b_s(Y_s),\, e_s\rangle\,\mathrm{d}s + 2\sqrt{2}\,\mathrm{d}W_s\,. \end{aligned}$$

Therefore Ito's formula applied to the strictly increasing concave function ρ yields to

$$\begin{aligned} \mathrm{d}\rho(r_s) &= \rho'(r_s)\,\mathrm{d}r_s + 4\,\rho''(r_s)\,\mathrm{d}s \\ &= (4\,\rho''(r_s) + \rho'(r_s)\,\langle b_s(X_s^{\boldsymbol{q}}) - b_s(Y_s),\, e_s\rangle)\mathrm{d}s + 2\sqrt{2}\,\rho'(r_s)\,\mathrm{d}W_s \\ &\overset{(5.2.3)}{\leq} (4\,\rho''(r_s) - \rho'(r_s)\,\kappa_b(r_s)\, r_s)\mathrm{d}s + 2\sqrt{2}\,\rho'(r_s)\,\mathrm{d}W_s \\ &\overset{(5.2.4)}{\leq} (4\,\rho''(r_s) - r_s\,\rho'(r_s)\,\bar{\kappa}(r_s))\mathrm{d}s + 2\sqrt{2}\,\rho'(r_s)\,\mathrm{d}W_s \\ &\overset{(5.2.8)}{\leq} -\lambda\,\rho(r_s)\,\mathrm{d}s + 2\sqrt{2}\,\rho'(r_s)\,\mathrm{d}W_s\,. \end{aligned}$$

Since $\rho(r_s) = 0$ as soon as $s \geq \tau$, by taking expectation, integrating over $s \in [t, T]$, recalling that $r_t = |x-y|$, and applying Gronwall Lemma we finally deduce that

$$\mathbb{E}[\rho(r_T)] \leq \exp(-\lambda(T-t))\rho(|x-y|)\,,$$

which combined with (5.2.13) yields to

$$|\mathcal{D}_t^{T,h}(y) - \mathcal{D}_t^{T,h}(x)| \leq \exp(-\lambda(T-t))\rho(|x-y|)\|h\|_\rho\,.$$

Since $x \neq y \in \mathbb{R}^d$ were arbitrary, this concludes our proof. □

Remark 5.2.4. *In the previous proof we have shown that it is enough considering an asymptotically convex potential in the underlying SDE in order to backward propagate the Lipschitzianity. This means that the same coupling by reflection technique can by employed in the proof of Lemma 5.1.1, allowing us to consider there just an asymptotically convex potential U, instead of a potential satisfying* $\mathsf{CD}(\kappa, \infty)$ *with* $\kappa > 0$. *In order to rely on Proposition 5.2.2, we then need*

$$\kappa_U(r) := \inf\left\{ \frac{\langle \nabla U(x) - \nabla U(y),\, x-y \rangle}{|x-y|^2} \;:\; |x-y| = r \right\},$$

to satisfy (5.2.7). *This is the case if U is strongly convex outside a compact set, or also if U is a Lipschitz perturbation of a strongly convex potential (cf. Remark 6.0.1).*

Lemma 5.2.5. *Assume* $\mathsf{CD}(\kappa, \infty)$ *for some* $\kappa > 0$, *that the two marginals satisfy* ***A* 4** *and further assume the initialisation of Sinkhorn's algorithm* ψ^0 *to be a Lipschitz function. Take* $\bar{\kappa}$ *as in* (5.2.6) *and* (ρ, λ, C) *as in Proposition 5.2.2. Then it holds*

$$\begin{aligned} \|\psi^{n+1} - \psi^\star\|_\rho &\leq \exp(-\lambda\, T)\|\varphi^{n+1} - \varphi^\star\|_\rho \\ \|\varphi^{n+1} - \varphi^\star\|_\rho &\leq \exp(-\lambda\, T)\|\psi^{n} - \psi^\star\|_\rho\,. \end{aligned} \tag{5.2.14}$$

As a result

$$\begin{aligned} \|\psi^{n+1} - \psi^\star\|_\rho &\leq \exp(-2\lambda\, T)\|\psi^{n} - \psi^\star\|_\rho \\ \|\varphi^{n+1} - \varphi^\star\|_\rho &\leq \exp(-2\lambda\, T)\|\varphi^{n} - \varphi^\star\|_\rho\,. \end{aligned} \tag{5.2.15}$$

Proof. A first consequence of Proposition 5.2.3 is that $\mathcal{D}_t^n := \mathcal{U}_t^{T,\psi^n} - \mathcal{U}_t^{T,\psi^\star}$, that is the solution of (5.2.12) with $h = \psi^n - \psi^\star$, satisfies

$$\|\mathcal{D}_0^n\|_\rho \leq \exp(-\lambda T)\|\psi^n - \psi^\star\|_\rho\,.$$

By recalling Sinkhorn's iterates definition (5.1.7) and the Schrödinger system (2.2.17), we may then write that

$$\|\varphi^{n+1} - \varphi^\star\|_\rho = \|\mathcal{D}_0^n\|_\rho \leq \exp(-\lambda T)\|\psi^n - \psi^\star\|_\rho\,,$$

which is the latter bound in (5.2.14). The first bound appearing in (5.2.14) can be proven in the same way, by exchanging the role between iterates and potentials.

Finally, the estimates in (5.2.15) are just the two-step bounds given via the former ones. □

From the previous key propagation estimate we may finally deduce our first exponential convergence result.

Theorem 5.2.6. *Assume* $\mathsf{CD}(\kappa,\infty)$ *for some* $\kappa > 0$*, that the two marginals satisfy* ***A****4 and further assume the initialisation of Sinkhorn's algorithm* ψ^0 *to be a Lipschitz function. Let* (ρ,λ,C) *as in Proposition 5.2.2, associated to* $\bar\kappa$ *defined at* (5.2.6)*. Then for any* $n\in\mathbb{N}$ *it holds*

$$\begin{aligned}\mathrm{Lip}(\varphi^n-\varphi^\star)&\le\gamma^{2n-1}\,2\,e^{L^2/8\kappa}\,\mathrm{Lip}(\psi^0-\psi^\star)\\ \mathrm{Lip}(\psi^n-\psi^\star)&\le\gamma^{2n}\,2\,e^{L^2/8\kappa}\,\mathrm{Lip}(\psi^0-\psi^\star)\end{aligned}\tag{5.2.16}$$

where $\gamma=\exp(-\lambda T)$*. As a consequence, uniformly in* $x\in\mathbb{R}^d$ *it holds*

$$\begin{aligned}|\varphi^{\diamond n}-\varphi^\star|(x)&\le\gamma^{2n-1}\,(|x|+M_1(\mu))\,2\,e^{L^2/8\kappa}\,\mathrm{Lip}(\psi^0-\psi^\star)\\ |\psi^{\diamond n}-\psi^\star|(x)&\le\gamma^{2n}\,(|x|+M_1(\nu))\,2\,e^{L^2/8\kappa}\,\mathrm{Lip}(\psi^0-\psi^\star)\end{aligned}\tag{5.2.17}$$

which integrated further implies for $p\in\{1,2\}$

$$\begin{aligned}\|\varphi^{\diamond n}-\varphi^\star\|_{\mathrm{L}^p(\mu)}&\le\gamma^{2n-1}\,4\,e^{L^2/8\kappa}\,M_p(\mu)^{1/p}\,\mathrm{Lip}(\psi^0-\psi^\star)\\ \|\psi^{\diamond n}-\psi^\star\|_{\mathrm{L}^p(\nu)}&\le\gamma^{2n}\,4\,e^{L^2/8\kappa}\,M_p(\nu)^{1/p}\,\mathrm{Lip}(\psi^0-\psi^\star)\,.\end{aligned}\tag{5.2.18}$$

Finally, if we start Sinkhorn's algorithm at $\psi^0=U_\nu$*, then the last constant factor appearing in the above right hand sides can be further bounded as*

$$\mathrm{Lip}(\psi^0-\psi^\star)\le e^{-\kappa T}\,\frac{\mathrm{Lip}(U_\mu)+\exp(-\kappa\,T)\mathrm{Lip}(U_\nu)}{1-\exp(-2\kappa\,T)}\,.$$

Proof. The convergence estimates in (5.2.16) follow from Lemma 5.2.5 and (5.2.11).

Since $\int\varphi^{\diamond n}\mathrm{d}\mu=\int\varphi^\star\mathrm{d}\nu$ (see (5.1.9)), uniformly on $x\in\mathbb{R}^d$, it holds

$$\begin{aligned}|\varphi^{\diamond n}-\varphi^\star|(x)&=\left|\varphi^{\diamond n}(x)-\int\varphi^{\diamond n}\,\mathrm{d}\mu-\varphi^\star(x)+\int\varphi^\star\,\mathrm{d}\mu\right|\\ &=\left|\int\Big[(\varphi^n-\varphi^\star)(x)-(\varphi^n-\varphi^\star)(y)\Big]\mathrm{d}\mu(y)\right|\\ &\le\int\Big|(\varphi^n-\varphi^\star)(x)-(\varphi^n-\varphi^\star)(y)\Big|\mathrm{d}\mu(y)\\ &\le\mathrm{Lip}(\varphi^n-\varphi^\star)\int|x-y|\,\mathrm{d}\mu(y)\\ &\overset{(5.2.16)}{\le}\gamma^{2n-1}\,(|x|+M_1(\mu))\,2e^{\frac{L^2}{8\kappa}}\,\mathrm{Lip}(\psi^0-\psi^\star)\,.\end{aligned}$$

The second bound appearing in (5.2.17) can be obtained by reasoning in the same fashion, since $\int\psi^{\diamond n}\mathrm{d}\nu=\int\psi^\star\mathrm{d}\nu$.

The integrated bounds appearing in (5.2.18) are a direct consequence of (5.2.17).

Finally, if $\psi^0 = U_\nu$, then $\psi^0 - \psi^\star = \mathcal{U}_0^{T,\varphi^\star}$ and from Lemma 5.1.1 and Corollary 5.1.3 we may prove the last statement. □

From the previous result we may also deduce the exponential convergence rates for the primal Sinkhorn's iterates, *i.e.*, for Sinkhorn's plans $(\pi^{n,n})_{n\in\mathbb{N}}$ and $(\pi^{n+1,n})_{n\in\mathbb{N}}$.

Theorem 5.2.7. *Assume* $\mathsf{CD}(\kappa,\infty)$ *for some* $\kappa > 0$*, that the two marginals satisfy* **A**4 *and further assume the initialisation of Sinkhorn's algorithm* ψ^0 *to be a Lipschitz function. Let* (ρ,λ,C) *as in Proposition 5.2.2, associated to* $\bar\kappa$ *defined at* (5.2.6)*, and let* $\gamma = \exp(-\lambda T)$*. Then for any* $n \in \mathbb{N}$ *it holds*

$$\mathscr{H}^{\mathrm{sym}}(\pi^{n,n},\boldsymbol{\pi}^T) \leq \gamma^{2n-1}\,(4\,M_1(\mu) + C_1(\mu)\,\sqrt{\mathcal{H}(\mu^1|\mu)}\,)\,2\,e^{L^2/8\kappa}\,\mathrm{Lip}(\psi^0-\psi^\star),$$
$$\mathscr{H}^{\mathrm{sym}}(\pi^{n+1,n},\boldsymbol{\pi}^T) \leq \gamma^{2n}\,(4\,M_1(\nu) + C_1(\nu)\,\sqrt{\mathcal{H}(\nu^0|\nu)}\,)\,2\,e^{L^2/8\kappa}\,\mathrm{Lip}(\psi^0-\psi^\star).$$

for some non-negative constants $C_1(\mu)$, $C_1(\nu)$*. Finally, the the same bounds hold for*

$$\mathscr{H}^{\mathrm{sym}}(\pi^{n+1,n},\pi^{n,n}) = \mathscr{H}^{\mathrm{sym}}(\mu^n,\mu),\ \mathscr{H}^{\mathrm{sym}}(\pi^{n+1,n+1},\pi^{n+1,n}) = \mathscr{H}^{\mathrm{sym}}(\nu^n,\nu).$$

Proof. The proof is similar to the one we will give for Theorem 6.4.4 in the next chapter.

Firstly, by reasoning as in Corollary 6.4.2, relying on the weighted Csiszár-Kullback-Pinsker inequalities ([BV05, Theorem 2.1], see also Lemma 6.4.3[2], from (5.2.17) we deduce that it holds

$$\begin{aligned}
\|\varphi^{\diamond n}-\varphi^\star\|_{\mathrm{L}^1(\mu^n)} &\leq \gamma^{2n-1}\,(M_1(\mu^n)+M_1(\mu))\,2\,e^{L^2/8\kappa}\,\mathrm{Lip}(\psi^0-\psi^\star)\\
&\leq \gamma^{2n-1}\,(2\,M_1(\mu)+C_1(\mu)\,\sqrt{\mathcal{H}(\mu^n|\mu)})\,2\,e^{L^2/8\kappa}\,\mathrm{Lip}(\psi^0-\psi^\star)\\
\|\psi^{\diamond n}-\psi^\star\|_{\mathrm{L}^1(\nu^n)} &\leq \gamma^{2n}\,(M_1(\nu^n)+M_1(\nu))\,2\,e^{L^2/8\kappa}\,\mathrm{Lip}(\psi^0-\psi^\star)\\
&\leq \gamma^{2n}\,(2\,M_1(\nu)+C_1(\nu)\,\sqrt{\mathcal{H}(\nu^n|\nu)})\,2\,e^{L^2/8\kappa}\,\mathrm{Lip}(\psi^0-\psi^\star)\,,
\end{aligned}$$

for two positive constants $C_1(\mu)$, $C_1(\nu)$, independent from $n \in \mathbb{N}$. Since the sequences $(\mathscr{H}(\mu^n|\mu))_{n\in\mathbb{N}}$ and $(\mathscr{H}(\nu^n|\nu))_{n\in\mathbb{N}}$ are monotone decreasing along Sinkhorn's algorithm [Nut21, Proposition 6.10], we finally get

$$\begin{aligned}
\|\varphi^{\diamond n}-\varphi^\star\|_{\mathrm{L}^1(\mu^n)} &\leq \gamma^{2n-1}\,(2\,M_1(\mu)+C_1(\mu)\,\sqrt{\mathcal{H}(\mu^1|\mu)})\,2\,e^{L^2/8\kappa}\,\mathrm{Lip}(\psi^0-\psi^\star)\\
\|\psi^{\diamond n}-\psi^\star\|_{\mathrm{L}^1(\nu^n)} &\leq \gamma^{2n}\,(2\,M_1(\nu)+C_1(\nu)\,\sqrt{\mathcal{H}(\nu^0|\nu)})\,2\,e^{L^2/8\kappa}\,\mathrm{Lip}(\psi^0-\psi^\star)\,.
\end{aligned} \tag{5.2.19}$$

[2]Our standing assumption **A**4 states that our marginals are Lipschitz perturbations of a log-concave reference measure which satisfy **A**6, cf. Remark 6.0.1.

As a first consequence of the above bounds and (5.2.18), for any $n \geq 1$ it holds

$$\varphi^{\diamond n} - \varphi^\star \in \mathrm{L}^1(\mu) \cap \mathrm{L}^1(\mu^n) \quad \text{and} \quad \psi^{\diamond n} - \psi^\star \in \mathrm{L}^1(\nu) \cap \mathrm{L}^1(\nu^n)\,,$$

which will guarantee that the following integrals (and corresponding summations) are all well-defined.

Now, (2.2.3) and (2.2.19) imply that

$$\log \frac{\mathrm{d}\boldsymbol{\pi}^T}{\mathrm{d}\pi^{n,n}}(x,y) = \varphi^n(x) - \varphi^\star(x) + \psi^n(y) - \psi^\star(y)\,,$$

and hence the symmetric relative entropies can be rewritten as

$$\begin{aligned}
&\mathscr{H}(\pi^{n,n}|\boldsymbol{\pi}^T) + \mathscr{H}(\boldsymbol{\pi}^T|\pi^{n,n}) \\
&\quad = \int (\varphi^n - \varphi^\star) \oplus (\psi^n - \psi^\star)\, \mathrm{d}\boldsymbol{\pi}^T - \int (\varphi^n - \varphi^\star) \oplus (\psi^n - \psi^\star)\, \mathrm{d}\pi^{n,n} \\
&\quad = \int (\varphi^{\diamond n} - \varphi^\star) \oplus (\psi^{\diamond n} - \psi^\star)\, \mathrm{d}\boldsymbol{\pi}^T - \int (\varphi^{\diamond n} - \varphi^\star) \oplus (\psi^{\diamond n} - \psi^\star)\, \mathrm{d}\pi^{n,n} \\
&\quad = \int (\varphi^{\diamond n} - \varphi^\star)\, \mathrm{d}\mu^n - \int (\varphi^{\diamond n} - \varphi^\star)\, \mathrm{d}\mu\,.
\end{aligned}$$

By combining the above with (5.2.18) and (5.2.19) we conclude the proof of the first bound in our thesis. The second bound can be proven similarly.

Finally let us notice that from (2.2.19) we may also deduce that

$$\begin{aligned}
\mathscr{H}(\pi^{n+1,n}|\pi^{n,n}) = \mathscr{H}(\mu|\mu^n) \quad &\text{and} \quad \mathscr{H}(\pi^{n,n}|\pi^{n+1,n}) = \mathscr{H}(\mu^n|\mu)\,, \\
\mathscr{H}(\pi^{n+1,n+1}|\pi^{n+1,n}) = \mathscr{H}(\nu|\nu^n) \quad &\text{and} \quad \mathscr{H}(\pi^{n+1,n}|\pi^{n+1,n+1}) = \mathscr{H}(\nu^n|\nu)\,.
\end{aligned}$$

and hence the bound for the adjusted marginals is a consequence of the previous ones, and the trivial inequalities

$$\begin{aligned}
\mathscr{H}(\pi^{n,n}|\pi^{n+1,n}) &= \mathscr{H}(\mu^n|\mu) \leq \mathscr{H}(\pi^{n,n}|\boldsymbol{\pi}^T), \\
\mathscr{H}(\pi^{n+1,n}|\pi^{n,n}) &= \mathscr{H}(\mu|\mu^n) \leq \mathscr{H}(\boldsymbol{\pi}^T|\pi^{n,n}), \\
\mathscr{H}(\pi^{n+1,n}|\pi^{n+1,n+1}) &= \mathscr{H}(\nu^n|\nu) \leq \mathscr{H}(\pi^{n+1,n}|\boldsymbol{\pi}^T), \\
\mathscr{H}(\pi^{n+1,n+1}|\pi^{n+1,n}) &= \mathscr{H}(\nu|\nu^n) \leq \mathscr{H}(\boldsymbol{\pi}^T|\pi^{n+1,n}).
\end{aligned}$$

□

5.2.1 Application to the Entropic Optimal Transport problem with quadratic cost

Similarly to what we have done in Section 4.2.1, here we translate previous convergence result for Sinkhorn's algorithm to to the Euclidean EOT setting with quadratic cost

$$\mathrm{EOT}_\varepsilon^{\mathsf{d}^2}(\mu,\nu) := \inf_{\pi \in \Pi(\mu,\nu)} \int |x-y|^2\, \mathrm{d}\pi + \varepsilon\, \mathscr{H}(\pi|\mu\otimes\nu)\,, \qquad \varepsilon > 0\,.$$

We will manage to do that under sufficiently general conditions, so that our results will apply to any couple of marginals $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ whose densities are log-Lipschitz perturbation of an underlying log-concave measure. Namely we are going to assume that

A5. *There exists $\kappa > 0$ and Lipschitz potentials U_μ, $U_\nu \in \mathcal{C}^1(\mathbb{R}^d)$ such that*

$$\mu(\mathrm{d}x) = \exp\left(-U_\mu(x) - \frac{\kappa}{2}|x|^2\right)\mathrm{d}x \quad \textit{and} \quad \nu(\mathrm{d}x) = \exp\left(-U_\nu(x) - \frac{\kappa}{2}|x|^2\right)\mathrm{d}x\,.$$

Clearly **A**5 implies **A**4 when considering as reference equilibrium measure $\mathrm{d}\mathfrak{m} \sim e^{-\frac{\kappa}{2}|x|^2}$ (cf. (4.2.11)). Moreover by taking T as in (4.2.12), that is T such that $\frac{\varepsilon\kappa}{4} = \sinh(\kappa T)$ we know that (cf. (4.2.14))

$$\int_{\mathbb{R}^{2d}} |x-y|^2\,\mathrm{d}\pi + \varepsilon\,\mathscr{H}(\pi|\mu\otimes\nu) = \varepsilon\,\mathscr{H}(\pi|\mathrm{R}_{0,T}) + A\,,$$

where $\mathrm{R}_{0,T}$ is the joint law at time 0 and T of the Ornstein-Uhlenbeck diffusion reference process

$$\mathrm{d}X_t = -\kappa X_t \mathrm{d}t + \sqrt{2}\,\mathrm{d}B_t\,,$$

whereas the additive constant reads as

$$A := -\frac{d\varepsilon}{2}\log(1-e^{-2\kappa T}) + (1-e^{-\kappa T})\big(M_2(\mu)+M_2(\nu)\big) \\ -\varepsilon\,\mathscr{H}(\mu|\mathfrak{m}) - \varepsilon\,\mathscr{H}(\nu|\mathfrak{m})\,.$$

Therefore the unique minimiser for $\mathrm{EOT}_\varepsilon^{\mathsf{d}^2}(\mu,\nu)$ coincides with the Schrödinger plan $\boldsymbol{\pi}^T$, and if $\varphi^\star$, $\psi^\star$ denote the corresponding Schrödinger potentials clearly it holds

$$\begin{aligned}\boldsymbol{\pi}^T(\mathrm{d}x\mathrm{d}y) &= \exp(-\varphi^\star(x)-\psi^\star(y))\mathrm{R}_{0,T}(\mathrm{d}x\mathrm{d}y) \\ &\overset{(4.2.13)}{=} \frac{\exp(-\varphi^\star_\kappa(x)-\psi^\star_\kappa(y))}{(1-e^{-2\kappa T})^{\frac{d}{2}}}\exp\left\{-\frac{|x|^2-2e^{\kappa T}\,x\cdot y+|y|^2}{\frac{2}{\kappa}(e^{2\kappa T}-1)}\right\},\end{aligned}$$

with

$$\varphi^\star_\kappa(x) := \varphi^\star(x) + \frac{\kappa}{2}|x|^2 \quad \text{and} \quad \psi^\star_\kappa(y) = \psi^\star(y) + \frac{\kappa}{2}|y|^2\,.$$

Then, Theorem 5.2.6 guarantees the exponential convergence of Sinkhorn's iterates (5.1.7), towards $\varphi^\star$ and $\psi^\star$ from which we can explicitly recover $\mathrm{EOT}_\varepsilon^{\mathsf{d}^2}(\mu,\nu)$. Similarly, Theorem 5.2.7 provides explicit exponential convergence rates for Sinkhorn's plans $(\pi^{n,n})_{n\in\mathbb{N}}$ and $(\pi^{n+1,n})_{n\in\mathbb{N}}$ towards the entropic optimal transport plan $\boldsymbol{\pi}^T$.

Bibliographical Remarks

The results presented in this chapter are based on ideas developed in [GNCD23]. There we have developed our ideas on the compact torus, and we have adapted to that periodic setting the coupling by reflection technique. Here we have decided to work on the whole $\mathbb{R}^d$, and we have showed that the Stochastic Optimal Control approach based on Lipschitz propagation along HJB equations allows to treat the exponential convergence of Sinkhorn's algorithm, when considering log-Lipschitz perturbations of a fixed strongly (or weakly, cf. Remark 5.2.4) log-concave potential reference. The setback of this approach is that both marginals should be Lipschitz perturbation of the same underlying log-concave measure. In the next chapter we will find a different approach that circumvents this problem, allowing also for marginals being perturbations of two different log-concave measures.

We postpone the literature review on Sinkhorn's algorithm to the bibliographical remarks section in Chapter 6.

This page was intentionally left blank.

Appendix 5

5.A Explicit rates and construction of the concave function

Below we provide the explicit construction of the concave function ρ, for the specific choice of $\bar{\kappa}$ as in (5.2.6) and we prove Proposition 5.2.2. We follow the construction given in [Con23, Section 2]. In view of that, let us consider

$$\begin{aligned} R_0 &:= \inf\{R : \bar{\kappa}(r) \geq 0 , \quad \forall\, r \geq R\} = L/\kappa \\ R_1 &:= \inf\{R \geq R_0 : \bar{\kappa}(r) R(R-R_0) \geq 8 , \quad \forall\, r \geq R\} = L/\kappa + \sqrt{8/\kappa} . \end{aligned} \tag{5.A.1}$$

Next, define the auxiliary functions

$$\begin{aligned} \phi(r) &:= \exp\left(-\frac{1}{4}\int_0^r s\,\bar{\kappa}(s)^-\,\mathrm{d}s\right) = \begin{cases} \exp\left(-\frac{2\,L\,r-\kappa r^2}{8}\right) & \forall\, r \leq L/\kappa \\ \exp(-L^2/8\kappa) & \forall\, r \geq L/\kappa \end{cases} \\ \Phi(r) &:= \int_0^r \phi(s)\,\mathrm{d}s = \int_0^{r\wedge R_0} \exp\left(-\frac{2\,L\,r-\kappa r^2}{8}\right)\mathrm{d}s + (r-R_0)^+ \exp(-L^2/8\kappa) \\ h(r) &:= 1 - \frac{\int_0^{r\wedge R_1}(\Phi/\phi)(s)\,\mathrm{d}s}{2\int_0^{R_1}(\Phi/\phi)(s)\,\mathrm{d}s} , \end{aligned} \tag{5.A.2}$$

where $a^+ := \max\{a, 0\}$. Finally let us consider the concave function

$$\rho(r) := \int_0^r \phi(s)\,h(s)\,\mathrm{d}s ,$$

let us define the convergence rate

$$\lambda := 2\left(\int_0^{R_1} \Phi(r)/\phi(r)\,\mathrm{d}r\right)^{-1}$$

and consider the constant $C := \phi(R_0)/2 = \exp(-L^2/8\kappa)/2$.

Proof of Proposition 5.2.2. Let (ρ, λ, C) be defined as above and let us start by simply noticing that

$$\phi(R_0) \le \phi(s) \le 1\,, \quad \phi(R_0)\, r \le \Phi(r) \le r \quad \text{and} \quad \frac{1}{2} \le h(r) \le 1\,, \tag{5.A.3}$$

which immediately proves the bounds for $\rho'(r) = \phi(r)h(r)$ with $C = \phi(R_0)/2$. From the previous bound on h we immediately deduce also

$$\Phi(r)/2 \le \rho(r) \le \Phi(r)\,, \tag{5.A.4}$$

which combined with the above bound for Φ concludes the proof of the first item.

In order to prove the second item it is enough to compute $\rho'(r) = \phi(r)h(r)$ and

$$\begin{aligned}\rho''(r) &= \phi'(r)h(r) + \phi(r)h'(r) = -\frac{r}{4}\,\bar\kappa(r)^-\,\phi(r)h(r) + \phi(r)h'(r) \\ &= -\frac{\bar\kappa(r)^-}{4}\, r\,\rho'(r) + \phi(r)h'(r) \le \frac{\bar\kappa(r)}{4}\, r\,\rho'(r) + \phi(r)h'(r)\,.\end{aligned}$$

Indeed as a byproduct we get

$$\rho''(r) - \frac{\bar\kappa(r)}{4}\, r\,\rho'(r) \le \phi(r)h'(r)\,,$$

and since for any $r < R_1$ it holds $h'(r) = -\frac{\lambda}{4}\,\Phi(r)/\phi(r)$, we deduce

$$\rho''(r) - \frac{\bar\kappa(r)}{4}\, r\,\rho'(r) \le -\frac{\lambda}{4}\,\Phi(r) \overset{(5.A.4)}{\le} -\frac{\lambda}{4}\,\rho(r) \quad \forall\, r < R_1\,.$$

At the same time, for any $r \ge R_0$ we have $\phi(r) = \phi(R_0)$ which implies

$$\Phi(r) = \Phi(R_0) + (r - R_0)\phi(R_0) \quad \forall r \ge R_0\,. \tag{5.A.5}$$

Now if we introduce $\tilde\Phi(r) := \Phi(r)/r$, the above expression gives us

$$\tilde\Phi'(r) = -\frac{\Phi(R_0)}{r^2} + \frac{R_0}{r^2}\,\phi(R_0) = \frac{1}{r^2}\int_0^{R_0} (\phi(R_0) - \phi(s))\mathrm{d}s \le 0$$

which is non-positive since ϕ is a decreasing function. From this we may deduce that for any $r \ge R_1$ the ratio function $\tilde\Phi$ is decreasing and therefore that

$$\frac{\Phi(r)}{r} \le \frac{\Phi(R_1)}{R_1} \quad \forall\, r \ge R_1\,. \tag{5.A.6}$$

Given this premise, for any $r \ge R_1$ it holds

$$\phi(r) = \phi(R_0)\,, \quad h(r) = \frac{1}{2} \quad \text{and} \quad \bar\kappa(r)\, R_1(R_1 - R_0) \ge 8$$

(cf. Definition (5.A.1)), which implies that $\rho'(r)$ is constantly equal to $\frac{\phi(R_0)}{2}$ for any $r \geq R_1$. Therefore, for any $r \geq R_1$ we have

$$\rho''(r) - \frac{\bar\kappa(r)}{4}\, r\,\rho'(r) = -\frac{\bar\kappa(r)}{4}\, r\, \frac{\phi(R_0)}{2} = -\frac{\bar\kappa(r)}{8}\, r\,\phi(R_0)$$
$$\overset{(5.A.1)}{\leq} -\frac{r\,\phi(R_0)}{R_1(R_1-R_0)} \overset{(5.A.6)}{\leq} -\frac{\Phi(r)}{\Phi(R_1)}\frac{\phi(R_0)}{R_1-R_0} \overset{(\dagger)}{\leq} -\frac{\lambda}{4}\,\Phi(r) \overset{(5.A.4)}{\leq} -\frac{\lambda}{4}\, f(r)$$

where inequality (†) follows from the observation that $\phi(s) = \phi(R_0)$ for any $r \geq R_0$ and that

$$\begin{aligned}
2\,\lambda^{-1} &= \int_0^{R_1} \Phi(s)/\phi(s)\,\mathrm{d}s \geq \int_{R_0}^{R_1} \Phi(s)/\phi(s)\,\mathrm{d}s \\
&\overset{(5.A.5)}{=} \int_{R_0}^{R_1} \frac{\Phi(R_0)+(s-R_0)\phi(R_0)}{\phi(R_0)}\,\mathrm{d}s = \frac{\Phi(R_0)}{\phi(R_0)}(R_1-R_0) + \frac{(R_1-R_0)^2}{2} \\
&= \frac{(R_1-R_0)}{2\,\phi(R_0)}\,(2\Phi(R_0)+(R_1-R_0)\phi(R_0)) \\
&\overset{(5.A.5)}{=} \frac{(R_1-R_0)}{2\,\phi(R_0)}\,(\Phi(R_0)+\Phi(R_1)) \geq \frac{(R_1-R_0)}{2\,\phi(R_0)}\,\Phi(R_1)\,.
\end{aligned}$$

□

5.A.1 Explicit lower-bound for the rate of convergence

In this section we provide a lower-bound for the rate λ built in Proposition 5.2.2.

Proof of the lower-bound (5.2.10). By the definition of λ, from (5.A.5) we may immediately deduce that

$$\begin{aligned}
2\lambda^{-1} &= \int_0^{R_1} \Phi(r)/\phi(r)\,\mathrm{d}r \\
&= \int_0^{R_0} \Phi(r)/\phi(r)\,\mathrm{d}r + \int_{R_0}^{R_1} \frac{\Phi(R_0)+(r-R_0)\phi(R_0)}{\phi(R_0)}\mathrm{d}r \\
&= \int_0^{R_0} \Phi(r)/\phi(r)\,\mathrm{d}r + (R_1-R_0)\,\frac{\Phi(R_0)}{\phi(R_0)} + \frac{(R_1-R_0)^2}{2} \\
&\leq R_1\frac{\Phi(R_0)}{\phi(R_0)} + \frac{(R_1-R_0)^2}{2}\,,
\end{aligned}$$

where the last inequality follows from the monotonicity of $\Phi(r)/\phi(r)$ since a direct computation, combined with (5.A.2), shows that for any $r \leq R_0$

$$\frac{\mathrm{d}}{\mathrm{d}r}\left(\frac{\Phi(r)}{\phi(r)}\right) = 1 + \frac{L-\kappa r}{4}\,\frac{\Phi(r)}{\phi(r)} \geq 0\,.$$

Therefore from (5.A.3) and (5.A.1) we conclude that

$$2\lambda^{-1} \leq e^{L^2/8\kappa}\left(\frac{L^2}{\kappa^2} + \frac{L\sqrt{8}}{\kappa^{3/2}}\right) + \frac{4}{\kappa}$$

or equivalently that

$$\lambda \geq \frac{2\kappa}{\frac{L^2}{\kappa} + L\sqrt{8/\kappa} + 4\, e^{-L^2/8\kappa}} e^{-L^2/8\kappa} \,.$$

□

Chapter 6

Exponential convergence of Sinkhorn's algorithm: non-perturbative approach

In this chapter we are going to focus our attention on the classical Schrödinger problem (1.2.2), *i.e.*, when considering as a reference measure the Gaussian

$$\mathrm{R}_{0,T}(\mathrm{d}x,\mathrm{d}y) = (2\pi T)^{-d/2}\,\exp(-|x-y|^2/_{2T})\,\mathrm{d}x\mathrm{d}y\,,$$

which can be equivalently stated as the quadratic EOT problem

$$\inf_{\pi\in\Pi(\mu,\nu)}\int\frac{|x-y|^2}{2}\mathrm{d}\pi + T\,\mathscr{H}(\pi|\mu\otimes\nu)\,.$$

For notations' purposes and to avoid confusion with Sinkhorn's iterates, we will again denote with $\varphi^\star$ and $\psi^\star$ the Schrödinger potentials defined in Theorem 2.2.1, so that it holds

$$\boldsymbol{\pi}^T(\mathrm{d}x\mathrm{d}y) = (2\pi T)^{-d/2}\,\exp\Big(-\frac{|x-y|^2}{2T}-\varphi^\star(x)-\psi^\star(y)\Big)\mathrm{d}x\,\mathrm{d}y\in\Pi(\mu,\nu)\,. \tag{6.0.1}$$

The starting point of our discussion is the following well known result, see [PNW21, Proposition 2] or [CP23] for instance. Define $(x,A)\mapsto\pi_T^{x,h}(A)$ as the Markov kernel on $\mathbb{R}^d\times\mathcal{B}(\mathbb{R}^d)$ whose transition density w.r.t. Lebesgue measure is proportional to $(x,y)\mapsto\exp(-h(y)-|x-y|^2/(2T))$, *i.e.*, for any $x\in\mathbb{R}^d$, $\pi_T^{x,h}$ is defined through

$$\pi_T^{x,h}(\mathrm{d}y)\propto\exp\Big(-\frac{|y-x|^2}{2T}-h(y)\Big)\mathrm{d}y\,. \tag{6.0.2}$$

Then, as we already pointed out in (3.2.21), for any $n \in \mathbb{N}^*$, $h \in \{\psi^\star,\ \psi^n, \varphi^\star, \varphi^n\}$ it holds

$$\begin{aligned} \nabla \log P_T \mathrm{e}^{-h}(x) &= T^{-1} \int (y-x)\, \pi_T^{x,h}(\mathrm{d}y)\,, \\ \nabla^2 \log P_T \mathrm{e}^{-h}(x) &= -\,T^{-1}\,\mathrm{Id} + T^{-2}\,\mathrm{Cov}(\pi_T^{x,h})\,. \end{aligned} \tag{6.0.3}$$

The proof of this technical result is given for completeness in Proposition 6.A.2 in Section 6.A.

Let us record here two observations on this conditional probability measure. The first is that for any $x \in \mathbb{R}^d$ the conditional distribution, $\pi_T^{x,h}$ defined in (6.0.2) is the invariant probability measure for the SDE

$$\mathrm{d}Y_t = -\left(\frac{Y_t - x}{2T} + \frac{1}{2}\nabla h(Y_t)\right)\mathrm{d}t + \mathrm{d}B_t\,. \tag{6.0.4}$$

The second one is that, defining the adjusted marginals produced along Sinkhorn as in (2.2.21), that is as the probability measures

$$\mu^n := (\mathrm{proj}_x)_\sharp \pi^{n,n} \quad \text{and} \quad \nu^n := (\mathrm{proj}_y)_\sharp \pi^{n+1,n}\,,$$

using (6.0.1)-(2.2.19), $\boldsymbol{\pi}^T \in \Pi(\mu,\nu)$ and $\pi^{n,n} \in \Pi(\star,\nu)$, the Schrödinger and Sinkhorn's plans can be written as

$$\boldsymbol{\pi}^T(\mathrm{d}x\mathrm{d}y) = \mu(\mathrm{d}x) \otimes \pi_T^{x,\psi^\star}(\mathrm{d}y) \quad \text{and} \quad \pi^{n,n}(\mathrm{d}x\mathrm{d}y) = \mu^n(\mathrm{d}x) \otimes \pi_T^{x,\psi^n}(\mathrm{d}y)\,, \tag{6.0.5}$$

and a direct computation shows that

$$\begin{aligned} &\int \pi_T^{x,\psi^\star}(\mathrm{d}y)\, \mu(\mathrm{d}x) = \nu(\mathrm{d}y)\,, \qquad &&\int \pi_T^{x,\psi^n}(\mathrm{d}y)\, \mu^n(\mathrm{d}x) = \nu(\mathrm{d}y)\,, \\ &\int \pi_T^{y,\varphi^\star}(\mathrm{d}x)\, \nu(\mathrm{d}y) = \mu(\mathrm{d}x)\,, \qquad &&\int \pi_T^{y,\varphi^n}(\mathrm{d}x)\, \nu^n(\mathrm{d}y) = \mu(\mathrm{d}x)\,. \end{aligned} \tag{6.0.6}$$

The identities in (6.0.3) show that the convergence of the gradient and Hessian along Sinkhorn's iterates is tightly linked to the conditional measures and their ergodicity, or equivalently to their concavity profile. Indeed, from the definition of Sinkhorn's algorithm (2.2.18) and from these formulas (with $h = \psi^n,\ \psi^\star$) we deduce the upper bound

$$\int |\nabla \varphi^{n+1} - \nabla \varphi^\star|(x)\, \mu(\mathrm{d}x) \le T^{-1} \int \mathbf{W}_1(\pi_T^{x,\psi^n}, \pi_T^{x,\psi^\star})\, \mu(\mathrm{d}x)\,,$$

where $\mathbf{W}_1(\cdot,\cdot)$ denotes the Wasserstein distance of order one. Combining together the lower bounds on the integrated convexity profiles κ_{ψ^n} (obtained in Section 6.1), the representation of $\pi_T^{x,\psi^n}, \pi_T^{x,\psi^\star}$ as invariant measures for (6.0.4) and coupling techniques (see Section 6.2), we obtain in Corollary 6.2.3 that the key estimate

$$\mathbf{W}_1(\pi_T^{x,\psi^n}, \pi_T^{x,\psi^\star}) \le \gamma_n^\nu \int |\nabla \psi^n - \nabla \psi^\star|(y)\, \pi_T^{x,\psi^\star}(\mathrm{d}y) \tag{6.0.7}$$

holds uniformly in $x \in \mathbb{R}^d$ for some $\gamma_n^\nu > 0$. Integrating the key estimate (6.0.7) w.r.t. μ and invoking (6.0.6) yields

$$\int |\nabla \varphi^{n+1} - \nabla \varphi^\star|(x)\, \mu(\mathrm{d}x) \le \frac{\gamma_n^\nu}{T} \int |\nabla \psi^n - \nabla \psi^\star|(y)\nu(\mathrm{d}y)\,. \tag{6.0.8}$$

Repeating the same argument but exchanging the roles of $\psi^n, \psi^\star$ and $\varphi^n, \varphi^\star$ we obtain

$$\int |\nabla \psi^n - \nabla \psi^\star|(y)\, \nu(\mathrm{d}y) \le \frac{\gamma_{n-1}^\mu}{T} \int |\nabla \varphi^n - \nabla \varphi^\star|(x)\mu(\mathrm{d}x) \tag{6.0.9}$$

for some $\gamma_{n-1}^\mu > 0$. Combining (6.0.9) with (6.0.8) allows to establish exponential convergence provided $T^{-2}\gamma_n^\nu \gamma_{n-1}^\mu < 1$ for n large enough.

In the rest of this chapter we are going to properly justify the proof strategy explained above.

Assumptions

In this section we provide a rigorous statement of the main assumptions we impose on the marginals μ, ν. In view of this, it is convenient to introduce some notation and terminology. A crucial role in this chapter is played by the integrated convexity profile $\kappa_U : \mathbb{R}_+^* \to \mathbb{R}$, which for any differentiable function $U : \mathbb{R}^d \to \mathbb{R}$ is defined as the function

$$\kappa_U(r) := \inf\left\{ \frac{\langle \nabla U(x) - \nabla U(y),\, x - y\rangle}{|x-y|^2} \quad : \quad |x-y| = r \right\}.$$

Likewise, for a distribution $\zeta(\mathrm{d}x) \propto \exp(-U)\mathrm{d}x$, we simply write κ_ζ for κ_U and referred to this function as to the integrated log-concavity profile of ζ. The function κ_U is often employed to quantify ergodicity of stochastic differential equations whose drift field is $-\nabla U$, see [Ebe16] and the discussion concerning (5.2.3) in the previous chapter. The integrated concavity profile of U is defined in a similar way as $\ell_U = -\kappa_{-U}$, and for ζ of the form $\zeta(\mathrm{d}x) \propto \exp(-U)\mathrm{d}x$ we set $\ell_\zeta = \ell_U$. Our main results apply when the marginals μ, ν satisfy the following property for $\zeta \in \{\mu, \nu\}$,

$$\liminf_{r \to +\infty} \kappa_\zeta(r) > 0\,, \quad \liminf_{r \to 0} r\kappa_\zeta(r) = 0\,.$$

Below we are going to give a more precise and detailed assumption on marginals. In view of that, let us consider two sets of functions that will appear in our conditions: $\mathcal{G} := \{g \in \mathcal{C}^2((0,+\infty), \mathbb{R}_+) : g \text{ satisfies } (\mathbf{H}_\mathcal{G})\}$ with

$$(r \mapsto r^{1/2}\, g(r^{1/2})) \text{ is non-decreasing and concave,} \quad \lim_{r\downarrow 0} r\, g(r) = 0 \tag{$\mathbf{H}_\mathcal{G}$}$$

and its subset

$$\tilde{\mathcal{G}} := \left\{ g \in \mathcal{G} \text{ bounded and s.t. } \lim_{r \downarrow 0} g(r) = 0\,,\ g' \geq 0 \text{ and } 2g'' + g\,g' \leq 0 \right\}.$$

The above classes of functions are non-empty and in particular $\tilde{\mathcal{G}}$ contains $r \mapsto 2\tanh(r/2)$. Though it may not appear as the most natural at first sight, these choices will become clear in light of our proofs. Indeed, the sets $\mathcal{G}$ and $\tilde{\mathcal{G}}$ enjoy special invariance properties (see Theorem 6.1.1 and Lemma 6.1.8) under the mapping

$$g \mapsto -\log P_T \exp(-g)\,,$$

upon which the proof of the lower bounds on the integrated convexity profiles of potentials (see Section 6.1) are built.

Note that the properties prescribed in the definition of $\tilde{\mathcal{G}}$ in particular imply that its elements are sublinear (*i.e.*, $\sup_{r>0} g(r)/r < +\infty$) concave functions on $(0, +\infty)$. Indeed, any $g \in \tilde{\mathcal{G}}$ is clearly non-negative and non-decreasing, which combined with the differential inequality, implies its concavity. Finally, since $g(0^+) = 0$, the sublinearity of any $g \in \tilde{\mathcal{G}}$ follows. We say that a potential $U : \mathbb{R}^d \to \mathbb{R}$ is asymptotically strongly convex if there exist $\alpha_U \in \mathbb{R}^*_+$ and $\tilde{g}_U \in \tilde{\mathcal{G}}$ such that

$$\kappa_U(r) \geq \alpha_U - r^{-1}\,\tilde{g}_U(r) \tag{6.0.10}$$

holds for all $r \geq 0$. We consider the set of asymptotically strongly log-concave probability measures

$$\mathcal{P}_{\mathrm{alc}}(\mathbb{R}^d) = \{\zeta(\mathrm{d}x) = e^{-U}\mathrm{d}x : \ U \in \mathcal{C}^2(\mathbb{R}^d),\ U \text{ asymptotically strongly convex}\} \tag{6.0.11}$$

Note that as soon as $\zeta(\mathrm{d}x) = \exp(-U(x))\mathrm{d}x$ there exist $\beta_U \in (0, +\infty]$ and $g \in \mathcal{G}$ such that

$$\ell_U(r) \leq \beta_U + r^{-1}\,g_U(r) \tag{6.0.12}$$

holds for all $r \geq 0$. This is trivially true as we can choose $\beta_U = +\infty$. However, if the above holds for some $\beta_U < +\infty$, we obtain better lower bounds on the integrated convexity profile of Sinkhorn's potentials and all contraction rates appearing in our main results are better than those obtained for $\beta_U = +\infty$.

Remark 6.0.1. *Let us remark that the class of probability measures $\mathcal{P}_{\mathrm{alc}}(\mathbb{R}^d)$ contains in particular probability measures ζ associated with potentials U satisfying*

$$\kappa_U(r) \geq \begin{cases} \alpha & \text{if } r > R \\ \alpha - C_U & \text{if } r \leq R\,, \end{cases} \tag{6.0.13}$$

for $\alpha, C_U, R > 0$. Indeed, [Con24, Proposition 5.1] implies that $\kappa_\zeta(r) \geq \alpha - r^{-1}\,\tilde{g}_L(r)$ where $\tilde{g}_L \in \tilde{\mathcal{G}}$ is given by

$$\tilde{g}_L(r) := 2\,(L)^{1/2}\,\tanh(r\,L^{1/2}/2) \ \text{ with } \ L := \inf\{\bar{L} :\ R^{-1}\,\tilde{g}_{\bar{L}}(R) \geq C_U\}\,. \tag{6.0.14}$$

Finally, it is worth mentioning that (6.0.13) *holds if U can be expressed as the sum of a strongly convex function and a Lipschitz function with second derivative bounded from below.*

We have now all the concepts and notations to introduce our assumptions.

A6. *The marginals $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ with log-densities U_μ, U_ν, that is*

$$\mu(\mathrm{d}x) = \exp(-U_\mu(x))\mathrm{d}x\,, \qquad \nu(\mathrm{d}x) = \exp(-U_\nu(x))\mathrm{d}x\,,$$

belong to the set $\mathcal{P}_{\mathrm{alc}}(\mathbb{R}^d)$ defined at (6.0.11) *and have finite relative entropy with respect to the Lebesgue measure* Leb, $\mathrm{Ent}(\mu)$, $\mathrm{Ent}(\nu) < +\infty$.

Under **A**6, we denote by α_μ, β_μ, $\tilde{g}_\mu$ and g_μ (resp. α_ν, β_ν, $\tilde{g}_\nu$ and g_ν) the constants and functions associated with μ (resp. ν) such that (6.0.10) and (6.0.12) hold for U_μ (resp. U_ν).

A special case of **A**6 corresponds to the strongly log-concave case, that is when assuming the marginals μ, ν satisfying

A7. *There exist α_μ, $\alpha_\nu \in (0, +\infty)$ and β_μ, $\beta_\nu \in (0, +\infty]$ such that*

$$\alpha_\mu \leq \nabla^2 U_\mu \leq \beta_\mu \quad \textit{and} \quad \alpha_\nu \leq \nabla^2 U_\nu \leq \beta_\nu\,.$$

Indeed this clearly implies the validity of **A**6 with $\tilde{g}_\mu$, g_μ, $\tilde{g}_\nu$ and g_ν all null. We will specify our results to the strongly log-concave case as in **A**7 in Section 6.6 where the rate of convergence get a simple and explicit expression.

6.1 Integrated convexity profile propagation along Sinkhorn's algorithm

In this section we establish lower bounds on the integrated convexity profile of Sinkhorn's potentials. Before proceeding further, let us point out here that the results of this section hold under a weaker assumption than **A**6. Namely, let us consider the followings

A'1. *The two distributions $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ specified by* (2.2.16) *have finite relative entropy with respect to the Lebesgue measure* Leb, $\mathrm{Ent}(\mu)$, $\mathrm{Ent}(\nu) < +\infty$;

A'2.

(i) There exist $\alpha_\nu \in (0, +\infty)$ and $\beta_\mu \in (0, +\infty]$ such that

$$\kappa_{U_\nu}(r) \geq \alpha_\nu - r^{-1}\, \hat{g}_\nu(r) \quad \textit{and} \quad \ell_{U_\mu}(r) \leq \beta_\mu + r^{-1}\, g_\mu(r)\,,$$

with $g_\mu \in \mathcal{G}$ and $\hat{g}_\nu \in \hat{\mathcal{G}}$;

(ii) There exist $\alpha_\mu \in (0, +\infty)$ *and* $\beta_\nu \in (0, +\infty]$ *such that*

$$\kappa_{U_\mu}(r) \geq \alpha_\mu - r^{-1}\,\hat{g}_\mu(r) \quad \text{and} \quad \ell_{U_\nu}(r) \leq \beta_\nu + r^{-1}\,g_\nu(r)\,,$$

with $g_\nu \in \mathcal{G}$ *and* $\hat{g}_\mu \in \hat{\mathcal{G}}$,

where

$$\hat{\mathcal{G}} := \left\{ g \in \mathcal{G} \quad \text{bounded and s.t.} \quad 2g'' + g\,g' \leq 0 \right\} \subseteq \mathcal{G}\,.$$

Notice that the above assumption allows for concave functions having non-null limit value in the origin (whereas the elements of $\tilde{\mathcal{G}}$ are sublinear in a neighborhood of the origin), that $\tilde{\mathcal{G}} \subseteq \hat{\mathcal{G}}$ and hence that **A**6 implies **A′**1 and **A′**2.

Let us introduce for any fixed $\beta > 0$ and any $g \in \mathcal{G}$ and $\hat{g} \in \hat{\mathcal{G}}$, the following functions for α, s, $u \geq 0$

$$\begin{aligned} F_\beta^{g,\hat{g}}(\alpha, s) &= \beta\, s + \frac{s}{T(1+T\alpha)} + s^{1/2}\,g(s^{1/2}) + \frac{s^{1/2}\,\hat{g}(s^{1/2})}{(1+T\alpha)^2}\,, \\ G_\beta^{g,\hat{g}}(\alpha, u) &= \inf\{s \geq 0 :\ F_\beta^{g,\hat{g}}(\alpha, s) \geq u\}\,, \end{aligned} \tag{6.1.1}$$

with the convention $G_\beta^{g,\hat{g}}(\alpha, u) \equiv 0$ whenever $\beta = +\infty$.

Then, the main result of this section can be stated as follows.

Theorem 6.1.1. *Assume* **A′**1. *If* **A′**2-(i) *holds and if*

$$\kappa_{\psi^0}(r) \geq \alpha_\nu - T^{-1} - r^{-1}\hat{g}_\nu(r)\,, \tag{6.1.2}$$

then there exists a monotone increasing sequence $(\alpha_{\nu,n})_{n\in\mathbb{N}} \subseteq (\alpha_\nu - T^{-1}, \alpha_\nu - T^{-1} + (\beta_\mu\, T^2)^{-1}]$ *such that for any* $n \geq 1$ *and* $r > 0$ *it holds*

$$\ell_{\varphi^n}(r) \leq r^{-2}\,F_{\beta_\mu}^{g_\mu,\hat{g}_\nu}(\alpha_{\nu,n}, r^2) - T^{-1} \quad \text{and} \quad \kappa_{\psi^n}(r) \geq \alpha_{\nu,n} - r^{-1}\hat{g}_\nu(r)\,, \tag{6.1.3}$$

with $F_{\beta_\mu}^{g_\mu,\hat{g}_\nu}$ *defined as in* (6.1.1). *Moreover, the sequence can be explicitly built by setting*

$$\begin{cases} \alpha_{\nu,0} := \alpha_\nu - T^{-1}\,, \\ \alpha_{\nu,n+1} := \alpha_\nu - T^{-1} + \dfrac{G_{\beta_\mu}^{g_\mu,\hat{g}_\nu}(\alpha_{\nu,n}, 2)}{2\,T^2}\,, \quad n \in \mathbb{N}\,, \end{cases} \tag{6.1.4}$$

$G_{\beta_\mu}^{g_\mu,\hat{g}_\nu}$ *given in* (6.1.1). *Finally,* $(\alpha_{\nu,n})_{n\in\mathbb{N}}$ *converges to* $\alpha_{\psi^\star} \in (\alpha_\nu - T^{-1}, \alpha_\nu - T^{-1} + (\beta_\mu\, T^2)^{-1}]$, *fixed point solutions of* (6.1.4) *and for any* $r > 0$,

$$\ell_{\varphi^\star}(r) \leq r^{-2}\,F_{\beta_\mu}^{g_\mu,\hat{g}_\nu}(\alpha_{\psi^\star}, r^2) - T^{-1} \quad \text{and} \quad \kappa_{\psi^\star}(r) > \alpha_{\psi^\star} - r^{-1}\hat{g}_\nu(r)\,, \tag{6.1.5}$$

where $\varphi^\star$ *and* $\psi^\star$ *are the Schrödinger potentials introduced in* (6.0.1).

*Similarly, under **A'**2-(ii) there exists a monotone increasing sequence* $(\alpha_{\mu,n})_{n\in\mathbb{N}} \subseteq (\alpha_\mu - T^{-1}, \alpha_\mu - T^{-1} + (\beta_\nu T^2)^{-1}]$ *such that for any* $n \geq 1$ *and* $r > 0$ *it holds*

$$\ell_{\psi^n}(r) \leq r^{-2} F^{g_\nu,\hat{g}_\mu}_{\beta_\nu}(\alpha_{\mu,n}, r^2) - T^{-1} \quad \text{and} \quad \kappa_{\varphi^n}(r) \geq \alpha_{\mu,n} - r^{-1}\hat{g}_\mu(r)\,, \tag{6.1.6}$$

with $F^{g_\nu,\hat{g}_\mu}_{\beta_\nu}$ *defined as in* (6.1.1) *and*

$$\begin{cases} \alpha_{\mu,1} := \alpha_\mu - T^{-1}\,, \\ \alpha_{\mu,n+1} := \alpha_\mu - T^{-1} + \frac{G^{g_\nu,\hat{g}_\mu}_{\beta_\nu}(\alpha_{\mu,n},2)}{2\,T^2}\,, \quad n \in \mathbb{N}\,, \end{cases} \tag{6.1.7}$$

with $G^{g_\nu,\hat{g}_\mu}_{\beta_\nu}$ *defined as in* (6.1.1)*. Finally,* $(\alpha_{\mu,n})_{n\in\mathbb{N}}$ *converges to* $\alpha_{\varphi^\star} \in (\alpha_\mu - T^{-1}, \alpha_\mu - T^{-1} + (\beta_\nu T^2)^{-1}]$*, fixed point solutions of* (6.1.7) *and for any* $r > 0$*,*

$$\ell_{\psi^\star}(r) \leq r^{-2} F^{g_\nu,\hat{g}_\mu}_{\beta_\nu}(\alpha_{\varphi^\star}, r^2) - T^{-1} \quad \text{and} \quad \kappa_{\varphi^\star}(r) > \alpha_{\varphi^\star} - r^{-1}\hat{g}_\mu(r)\,. \tag{6.1.8}$$

Remark 6.1.2. *The above result is an extension of [Con24, Theorem 1.2] where the author just provides the limit-bounds* (6.1.5) *and* (6.1.8) *in the case when* $g^\mu = g^\nu \equiv 0$ *and* $\hat{g}_\mu, \hat{g}_\nu$ *take the form* (6.0.14)*. In the above result we show that the iterative proof given there can be actually employed when proving the estimates* (6.1.3) *and* (6.1.6) *along Sinkhorn's algorithm.*

Let us also mention that our result encompasses [CP23, Theorem 4] when considering $\hat{g}_\nu \equiv 0$ *and* $g_\mu \equiv 0$ *in **A'**2-(i).*

We provide a proof of the above theorem at the end of this section. Let us mention here that we will show in Remark 6.1.9 that Assumption (6.1.2) on ψ^0 can be essentially dropped; here we simply observe that it is met for a regular enough initial condition, e.g., for $\psi^0 = U_\nu$.

Let us also point out that the above theorem guarantees the existence and uniqueness for the strong solution of the SDE

$$\mathrm{d}Y_t = -\left(\frac{Y_t - x}{2T} + \frac{1}{2}\nabla h(Y_t)\right)\mathrm{d}t + \mathrm{d}B_t\,. \tag{6.1.9}$$

for any choice of $h = \psi^\star$, ψ^n, $\varphi^\star$, φ^n. Indeed from Theorem 6.1.1 we immediately deduce

Corollary 6.1.3. *Under the assumptions of Theorem 6.1.1 it holds for any* $y \in \mathbb{R}^d$

$$\begin{aligned} \left\langle \nabla\psi^\star(y), \frac{y}{|y|}\right\rangle &\geq \alpha_{\psi^\star}\,|y| - \hat{g}_\nu(|y|) - |\nabla\psi^\star(0)|\,, \\ \left\langle \nabla\psi^n(y), \frac{y}{|y|}\right\rangle &\geq \alpha_{\nu,n}\,|y| - \hat{g}_\nu(|y|) - |\nabla\psi^n(0)|\,, \\ \left\langle \nabla\varphi^\star(y), \frac{y}{|y|}\right\rangle &\geq \alpha_{\varphi^\star}\,|y| - \hat{g}_\mu(|y|) - |\nabla\varphi^\star(0)|\,, \\ \left\langle \nabla\varphi^n(y), \frac{y}{|y|}\right\rangle &\geq \alpha_{\mu,n}\,|y| - \hat{g}_\mu(|y|) - |\nabla\varphi^n(0)|\,. \end{aligned}$$

As a consequence for any even $p \geq 2$ the potential $V_p(y) = 1 + |y|^p$ is a Lyapunov function for (6.1.9) *with $h \in \{\psi^\star, \psi^n, \varphi^\star, \varphi^n\}$. As a consequence, existence and uniqueness of strong solutions hold for these SDEs.*

Proof. The lower-bounds displayed above are a direct consequence of Theorem 6.1.1. Let $x \in \mathbb{R}^d$ be fixed. We will only consider now the case $h = \psi^\star$. The other cases follow the same lines. Since $\hat{g}_\nu \in \hat{\mathcal{G}}$ is bounded, it holds for any $y \in \mathbb{R}^d$

$$-\frac{1}{2}\left\langle T^{-1}(y-x) + \nabla\psi^\star(y), y\right\rangle \leq -\frac{\alpha_{\psi^\star} + T^{-1}}{2}|y|^2 + \frac{\|\hat{g}_\nu\|_\infty + |\nabla h(0)| + T^{-1}|x|}{2}|y|\,,$$

and hence there exist $\gamma > 0$ and $R > 0$ such that

$$-\frac{1}{2}\left\langle T^{-1}(y-x) + \nabla\psi^\star(y), y\right\rangle \leq -\gamma\,|y|^2 \quad \forall\,|y| \geq R\,.$$

At this stage, [MSH02, Lemma 4.2] guarantees that for any even $p \geq 2$ the potential $V_p(y) = 1 + |y|^p$ is a Lyapunov function for the diffusion (6.1.9).

More precisely it holds a geometric drift condition, *i.e.*, for any $A_{\psi^\star} \in (0, p\gamma)$ there exists a finite constant $B_{\psi^\star} = B_{\psi^\star}(A_{\psi^\star}, p)$ such that for any $y \in \mathbb{R}^d$

$$\mathcal{L}_{\psi^\star} V_p(y) \leq -A_{\psi^\star} V_p(y) + B_{\psi^\star}\,, \tag{6.1.10}$$

where above $\mathcal{L}_{\psi^\star} := \Delta/2 - \frac{1}{2}\langle T^{-1}(y-x) + \nabla\psi^\star(y), \nabla\rangle$ denotes the generator associated to the SDE (6.1.9). Finally, existence and uniqueness of strong solutions of (6.1.9) follows from [RT96, Theorem 2.1] (see also [MT93, Section 2]). ☐

The proof of Theorem 6.1.1 will be based on a propagation of integrated-convexity along Hamilton-Jacobi-Bellman (HJB) equations observed in [Con24], based on coupling by reflection techniques, which reads as follows

Theorem 6.1.4 (Theorem 2.1 in [Con24])**.** *For any fixed function $\hat{g} \in \hat{\mathcal{G}}$, consider the class of functions*

$$\mathcal{F}_{\hat{g}} := \{h \in \mathcal{C}^1(\mathbb{R}^d) \,:\, \kappa_h(r) \geq -r^{-1}\,\hat{g}(r) \quad \forall\, r > 0\}\,.$$

Then, the class $\mathcal{F}_{\hat{g}}$ is stable under the action of the HJB flow, i.e.,

$$h \in \mathcal{F}_{\hat{g}} \;\Rightarrow\; -\log P_{T-t}\exp(-h) \in \mathcal{F}_{\hat{g}} \quad \forall\, 0 \leq t \leq T\,.$$

We omit the proof of the above result since it runs exactly as stated in [Con24]. There it is proven when $\hat{g}$ is of the form (6.0.14). However the same proof allows to reach the conclusion for any function $\hat{g} \in \hat{\mathcal{G}}$ since it only requires $\hat{g}$ to satisfy the differential inequality

$$2(\hat{g})'' + \hat{g}\,(\hat{g})' \leq 0\,,$$

which is an equality in the special case considered there $\hat{g}(r) = \tanh(r)$.

As a first consequence of the previous theorem we may immediately deduce the following integrated propagation *convexity-to-concavity* result.

Lemma 6.1.5. *Assume **A′**1. If **A′**2-(i) holds true, $\alpha_{\nu,n} > -T^{-1}$ and if for any $r > 0$*

$$\kappa_{\psi^n}(r) \geq \alpha_{\nu,n} - r^{-1}\,\hat{g}_\nu(r)\,, \tag{6.1.11}$$

then

$$\ell_{\varphi^{n+1}}(r) \leq \beta_\mu + g_\mu(r) - \frac{\alpha_{\nu,n}}{1+T\alpha_{\nu,n}} + \frac{r^{-1}\,\hat{g}_\nu(r)}{(1+T\alpha_{\nu,n})^2} = -T^{-1} + r^{-2}\,F^{g_\mu,\hat{g}_\nu}_{\beta_\mu}(\alpha_{\nu,n}, r^2).$$

*Similarly if **A′**2-(ii) holds, $\alpha_{\mu,n} > -T^{-1}$ and if for any $r > 0$*

$$\kappa_{\varphi^n}(r) \geq \alpha_{\mu,n} - r^{-1}\,\hat{g}_\mu(r)\,,$$

then

$$\ell_{\psi^n}(r) \leq \beta_\nu + r^{-1}\,g_\nu(r) - \frac{\alpha_{\mu,n}}{1+T\alpha_{\mu,n}} + \frac{r^{-1}\,\hat{g}_\mu(r)}{(1+T\alpha_{\mu,n})^2} = -T^{-1} + r^{-2}\,F^{g_\nu,\hat{g}_\mu}_{\beta_\nu}(\alpha_{\mu,n}, r^2).$$

Proof. Let us firstly notice that our assumption (6.1.11) is equivalent to stating that

$$\bar{\psi}^n := \psi^n - \frac{\alpha_{\nu,n}}{2}|\cdot|^2 \in \mathcal{F}_{\hat{g}_\nu}\,,$$

and therefore Theorem 6.1.4 implies that

$$-\log P_T \exp(-\bar{\psi}^n) \in \mathcal{F}_{\hat{g}_\nu}\,. \tag{6.1.12}$$

By recalling that φ^{n+1} is defined via (2.2.18), in order to conclude it is enough noticing now that

$$\begin{aligned}
&-\log P_T \exp(-\psi^n)(x) - \frac{d}{2}\log(2\pi T)\\
&= -\log \int \exp\Big(-\frac{|x-y|^2}{2T} - \frac{\alpha_{\nu,n}}{2}|y|^2 - \bar{\psi}^n(y)\Big)\,\mathrm{d}y\\
&= \frac{\alpha_{\nu,n}\,|x|^2}{2(1+T\alpha_{\nu,n})} - \log\int \exp\Big(-\frac{1+T\alpha_{\nu,n}}{2T}\,|y-(1+T\alpha_{\nu,n})^{-1}x|^2 - \bar{\psi}^n(y)\Big)\,\mathrm{d}y\\
&= \frac{\alpha_{\nu,n}\,|x|^2}{2(1+T\alpha_{\nu,n})} - \log P_{T/(1+T\alpha_{\nu,n})}\exp(-\bar{\psi}^n)((1+T\alpha_{\nu,n})^{-1}x) - \frac{d}{2}\log\frac{2\pi T}{1+T\alpha_{\nu,n}}
\end{aligned}$$

and combining it with (6.1.12) and **A′**2-(i).

The second part of the statement follows the same lines. □

Lemma 6.1.6. *Fix $\beta \in (0, +\infty]$ and two functions $g \in \mathcal{G}$ and $\hat{g} \in \hat{\mathcal{G}}$. Then the following properties hold true.*

1. *For any $\alpha > -T^{-1}$ the function $s \mapsto F_\beta^{g,\hat{g}}(\alpha, s)$ is concave and increasing on $[0, +\infty)$.*

2. *The function $\alpha \mapsto G_\beta^{g,\hat{g}}(\alpha, 2)$ is positive and non-decreasing over $(-T^{-1}, +\infty)$.*

3. *For any given $a_0 > 0$, the fixed-point problem*

$$\alpha = a_0 - T^{-1} + \frac{G_\beta^{g,\hat{g}}(\alpha, 2)}{2\, T^2} \tag{6.1.13}$$

admits at least one solution on $(a_0 - T^{-1}, a_0 - T^{-1} + (\beta\, T^2)^{-1}]$ and, as soon as $\beta < +\infty$, $a_0 - T^{-1}$ does not belong to the closure of the set of solutions of (6.1.13).

Proof.

1. Since $r \mapsto r\, \hat{g}(r)$ and $r \mapsto r\, g(r)$ are non-decreasing and $\alpha > -T^{-1}$, an explicit computation shows that $s \mapsto F_\beta^{g,\hat{g}}(\alpha, s)$ is an increasing function. The concavity of the latter function follows from the properties of g, $\hat{g} \in \mathcal{G}$, since for $h = g$, $\hat{g}$ it holds

$$\frac{\mathrm{d}^2}{\mathrm{d}u^2} \left(u^{1/2}\, h(u^{1/2}) \right) \Big|_{u=s} = \frac{s^{-1/2}}{4} \left[h''(s^{1/2}) + s^{-1/2}\, h'(s^{1/2}) - s^{-1}\, h(s^{1/2}) \right] \leq 0\,.$$

2. The proof is by contradiction. Notice that $G_\beta^{g,\hat{g}}(\cdot, 2)$ is a continuous function on $(-T^{-1}, +\infty)$ and assume that is not a positive function, which implies that there exists some $\alpha > -T^{-1}$ such that $G_\beta^{g,\hat{g}}(\alpha, 2) = 0$ and hence by definition that there exists a sequence $(s_n)_{n \in \mathbb{N}}$ converging to zero and such that $F_\beta^{g,\hat{g}}(\alpha, s) \geq 2$, which is clearly impossible since we have $\lim_{s \downarrow 0} F_\beta^{g,\hat{g}}(\alpha, s) = 0$. Hence $G_\beta^{g,\hat{g}}(\cdot, 2)$ is a positive function. The monotonicity of $G_{\beta_\mu}^{g_\mu,\hat{g}_\nu}(\cdot, 2)$ follows from the fact that $F_\beta^{g,\hat{g}}(\alpha, s)$ is increasing in s and decreasing in $\alpha \in (-T^{-1}, +\infty)$, which implies for any $\alpha' \geq \alpha$ and $u \geq 0$

$$\{s : F_\beta^{g,\hat{g}}(\alpha', s) \geq u\} \subseteq \{s : F_\beta^{g,\hat{g}}(\alpha, s) \geq u\}\,.$$

3. Consider the map associated to the fixed-point problem (6.1.4), *i.e.*, the continuous function $H \colon [a_0 - T^{-1}, +\infty) \to \mathbb{R}$ defined for $\alpha \in (-T^{-1}, +\infty)$ as

$$H(\alpha) := \alpha - a_0 + T^{-1} - \frac{G_\beta^{g,\hat{g}}(\alpha, 2)}{2\, T^2}\,.$$

Let us now prove that

$$H(a_0 - T^{-1}) < 0 \quad \text{and} \quad \lim_{\alpha\to+\infty} H(\alpha) = +\infty\,. \tag{6.1.14}$$

The first inequality follows from a direct computation, showing that it holds $G^{g,\hat{g}}_{\beta}(a_0 - T^{-1}, 2) > 0$. In order to prove the second statement it is enough noticing that $G^{g,\hat{g}}_{\beta}(\cdot, 2)$ is bounded, which is immediate since $g, \hat{g} \geq 0$ implies for any $\alpha > -T^{-1}$ and $s > 0$ that $F^{g,\hat{g}}_{\beta}(\alpha, s) \geq \beta s$ and hence that

$$G^{g,\hat{g}}_{\beta}(\alpha, 2) \leq 2/\beta\,. \tag{6.1.15}$$

From (6.1.14) and the continuity of $G^{g,\hat{g}}_{\beta}(\cdot, 2)$ we finally deduce the existence of some $\bar{\alpha} \in (a_0 - T^{-1}, +\infty)$ such that $H(\bar{\alpha}) = 0$, *i.e.*, a fixed point for (6.1.13). As a consequence (6.1.15) further implies $\bar{\alpha} \leq a_0 - T^{-1} + (\beta T^2)^{-1}$. Finally $a_0 - T^{-1}$ does not belong to the closure of the set of fixed-points solutions, because if this was the case then the continuity of $G^{g,\hat{g}}_{\beta}(\cdot, 2)$ would have implied $H(a_0 - T^{-1}) = 0$, clearly in contrast with (6.1.14).

□

As a corollary of the previous lemma we have already proven the following

Corollary 6.1.7. *Assume **A′1**. If **A′**2-(i) holds true, then there exists at least one solution $\alpha^\star_\nu$ on $(\alpha_\nu - T^{-1}, \alpha_\nu - T^{-1} + (\beta_\mu T^2)^{-1}]$ to the fixed point associated to* (6.1.4). *Moreover, if β_μ is finite then $\alpha_\nu - T^{-1}$ is not an accumulation point for the set of solutions.*

*Similarly if **A′**2-(ii) holds true, there exists at least one solution $\alpha^\star_\mu$ on $(\alpha_\mu - T^{-1}, \alpha_\mu - T^{-1} + (\beta_\nu T^2)^{-1}]$ to the fixed point associated to* (6.1.7). *Moreover, if β_ν is finite then $\alpha_\mu - T^{-1}$ is not an accumulation point for the set of solutions.*

The next result is the counterpart to Lemma 6.1.5 and it shows that we do also have an integrated propagation *concavity-to-convexity*.

Lemma 6.1.8. *Assume **A′1**. If **A′**2-(i) holds, $\alpha_{\nu,n} > -T^{-1}$ and if*

$$\ell_{\varphi^{n+1}}(r) \leq -T^{-1} + r^{-2}\, F^{g_\mu, \hat{g}_\nu}_{\beta_\mu}(\alpha_{\nu,n}, r^2)\,, \tag{6.1.16}$$

then $\alpha_{\nu,n+1} > -T^{-1}$ and for any $r > 0$

$$\kappa_{\psi^{n+1}}(r) \geq \alpha_{\nu,n+1} - r^{-1}\, \hat{g}_\nu(r)\,.$$

*Similarly if **A′**2-(ii) holds, $\alpha_{\mu,n} > -T^{-1}$ and if*

$$\ell_{\psi^{n}}(r) \leq -T^{-1} + r^{-2}\, F^{g_\nu, \hat{g}_\mu}_{\beta_\nu}(\alpha_{\mu,n}, r^2)\,,$$

then $\alpha_{\mu,n+1} > -T^{-1}$ *and for any* $r > 0$

$$\kappa_{\varphi^{n+1}}(r) \geq \alpha_{\mu,n+1} - r^{-1}\,\hat{g}_\mu(r)\,.$$

Proof. We are going to prove only the first part of the lemma since the second can be proven by an analogous reasoning. Firstly, let us consider the function

$$\hat{\psi}^{n+1}(y) := T\psi^{n+1}(y) - TU_\nu(y) + \frac{|y|^2}{2}\,. \tag{6.1.17}$$

Then (6.0.3) implies that its Hessian is given for any $y \in \mathbb{R}^d$ by

$$\nabla^2\hat{\psi}^{n+1}(y) = \frac{1}{T}\,\mathrm{Cov}(\pi_T^{y,\varphi^{n+1}})\,, \tag{6.1.18}$$

where we recall from the definition of $\pi_T^{y,\varphi^{n+1}}$ (6.0.2) that

$$\pi_T^{y,\varphi^{n+1}}(\mathrm{d}x) \propto \exp\Big(-\frac{|x-y|^2}{2T} - \varphi^{n+1}(x)\Big)\,\mathrm{d}x\,.$$

Moreover, if we set for notations' sake $V^{y,n+1} := -\log(\mathrm{d}\pi_T^{y,\varphi^{n+1}}/\mathrm{dLeb})$, then our assumption implies

$$\ell_{V^{y,n+1}}(r) \leq r^{-2}\,F_{\beta_\mu}^{g_\mu,\hat{g}_\nu}(\alpha_{\nu,n}, r^2) \quad \forall\, r > 0\,. \tag{6.1.19}$$

In order to prove the desired bound for $\kappa_{\psi^{n+1}}$, we will first establish a lower bound for the Hessian $\nabla^2\hat{\psi}^{n+1}$, *i.e.*, a lower bound for the covariance matrix (6.1.18). In view of that, let us consider for any fixed $y \in \mathbb{R}^d$, the variance

$$\mathrm{Var}_{X\sim\pi_T^{y,\varphi^{n+1}}}(X_1)$$

where X_i denotes the i^{th} scalar component of the random vector $X \sim \pi_T^{y,\varphi^{n+1}}$. Next, observe that

$$\mathrm{Var}_{X\sim\pi_T^{y,\varphi^{n+1}}}(X_1) \geq \mathbb{E}_{X\sim\pi_T^{y,\varphi^{n+1}}}[\mathrm{Var}_{X\sim\pi_T^{y,\varphi^{n+1}}}(X_1|X_2,\,\ldots,\,X_d)]\,, \tag{6.1.20}$$

and notice that for any given $z = (z_2,\,\ldots,\,z_d) \in \mathbb{R}^{d-1}$ it holds

$$\begin{aligned}&\mathrm{Var}_{X\sim\pi_T^{y,\varphi^{n+1}}}(X_1|X_2 = z_2,\ldots,X_d = z_d)\\ &\qquad= \frac{1}{2}\int_{\mathbb{R}^2}|x-\hat{x}|^2\,\pi_T^{y,\varphi^{n+1}}(\mathrm{d}x|z)\,\pi_T^{y,\varphi^{n+1}}(\mathrm{d}\hat{x}|z)\end{aligned}$$

where $(y,z,A) \mapsto \pi_T^{y,h}(A|z)$ is the Markov kernel on $\mathbb{R}^d \times \mathbb{R}^{d-1} \times \mathcal{B}(\mathbb{R})$ whose transition density w.r.t. Lebesgue measure is proportional to $\exp(-V^{y,n+1}(x,z))$.

If we set $V^{y,z}(x) := V^{y,n+1}(x,z)$ we then have, uniformly in $z \in \mathbb{R}^{d-1}$,

$$\begin{aligned}
1 =& \frac{1}{2}\int (\partial_x V^{y,z}(x) - \partial_x V^{y,z}(\hat{x}))(x-\hat{x})\ \pi_T^{y,\varphi^{n+1}}(\mathrm{d}x|z)\,\pi_T^{y,\varphi^{n+1}}(\mathrm{d}\hat{x}|z)\\
=& \frac{1}{2}\int \langle \nabla V^{y,n+1}(x,z) - \nabla V^{y,n+1}(\hat{x}),\ (x,z)-(\hat{x},z)\rangle\ \pi_T^{y,\varphi^{n+1}}(\mathrm{d}x|z)\,\pi_T^{y,\varphi^{n+1}}(\mathrm{d}\hat{x}|z)\\
\overset{(6.1.19)}{\leq}& \frac{1}{2}\int F^{g_\mu,\hat{g}_\nu}_{\beta_\mu}(\alpha_{\nu,n}, |x-\hat{x}|^2)\ \pi_T^{y,\varphi^{n+1}}(\mathrm{d}x|z)\,\pi_T^{y,\varphi^{n+1}}(\mathrm{d}\hat{x}|z)\\
\leq& F^{g_\mu,\hat{g}_\nu}_{\beta_\mu}\left(\alpha_{\nu,n}, 2\,\mathrm{Var}_{X\sim\pi_T^{y,\varphi^{n+1}}}(X_1|X_2=z_d,\ldots,X_d=z_d)\right),
\end{aligned}$$

where the last step follows from the concavity of $s \mapsto F^{g_\mu,\hat{g}_\nu}_{\beta_\mu}(\alpha_{\nu,n}, s)$ (point (i) in Lemma 6.1.6) and Jensen's inequality. By combining the above with (6.1.20), since $\alpha_{\nu,n} > -T^{-1}$ and $F^{g_\mu,\hat{g}_\nu}_{\beta_\mu}(\alpha_{\nu,n},\cdot)$ is increasing (cf. Lemma 6.1.6), we deduce by definition of $G^{g_\mu,\hat{g}_\nu}_{\beta_\mu}$ that

$$\mathrm{Var}_{X\sim\pi_T^{y,\varphi^{n+1}}}(X_1) \geq \frac{1}{2}\,G^{g_\mu,\hat{g}_\nu}_{\beta_\mu}(\alpha_{\nu,n},2)\,.$$

Since the definition ℓ_U is invariant under orthonormal transformation, for any orthonormal matrix O the functions $\varphi^{n+1}(O\,\cdot)$ satisfy the condition (6.1.16) too. The previous bound and this observation leads to

$$\mathrm{Var}_{X\sim\pi_T^{y,\varphi^{n+1}}}(\langle v,\,X\rangle) \geq \frac{1}{2}\,G^{g_\mu,\hat{g}_\nu}_{\beta_\mu}(\alpha_{\nu,n},2) \quad \forall\, y,\, v\in\mathbb{R}^d \text{ s.t. } |v|=1\,,$$

and hence from (6.1.18) we finally deduce

$$\langle v,\,\nabla^2\hat{\psi}^{n+1}(y)\,v\rangle \geq G^{g_\mu,\hat{g}_\nu}_{\beta_\mu}(\alpha_{\nu,n},2)\,\frac{|v|^2}{2T} \quad \forall\, y,\, v\in\mathbb{R}^d\,.$$

By recalling (6.1.17) and (6.1.4), the above bound concludes our proof. □

Remark 6.1.9 (A first trivial lower bound)**.** *Under **A**'1 and **A**'2, the above discussion already provides a first trivial lower bound for κ_{ψ^n}. Indeed* (6.1.18) *tells us that $\hat{\psi}^{n+1}$ is a convex function for any $n \geq 0$, which combined with* (6.1.17) *yields to for $n \geq 0$*

$$\kappa_{\psi^{n+1}}(r) \geq \alpha_\nu - T^{-1} - r^{-1}\hat{g}_\nu(r)\,.$$

Therefore in Theorem 6.1.1 we could consider any initialisation ψ^0 without prescriptions on its behaviour and run an iteration of Sinkhorn in order to get $\alpha_{\nu,1} \geq \alpha_\nu - T^{-1}$. At this point once can proceed again with the proof of Theorem 6.1.1 with the same (but shifted by -1) sequence of parameters $(\alpha_{\nu,n})_{n\in\mathbb{N}}$.

Let us notice that the same discussion holds for the sequence of φ^n, which yields to for $n \geq 0$

$$\kappa_{\varphi^{n+1}} \geq \alpha_\mu - T^{-1} - r^{-1}\hat{g}_\mu(r)\,,$$

which for $n = 0$ *gives for free the base step of the iteration* (6.1.7)*.*

Finally, let us remark here that since the potentials couple $(\varphi^\star, \psi^\star)$ *can be thought as a constant sequence of Sinkhorn's iterates, the above discussion proves also that*

$$\kappa_{\psi^\star}(r) \geq \alpha_\nu - T^{-1} - r^{-1}\hat{g}_\nu(r) \quad \textit{and} \quad \kappa_{\varphi^\star} \geq \alpha_\mu - T^{-1} - r^{-1}\hat{g}_\mu(r)\ .$$

Given the above lemmata, we are finally able to prove how lower-bounds for integrated convexity profiles propagate along Sinkhorn's algorithm.

Proof of Theorem 6.1.1. Let us start showing the first statement. Consider the sequence $(\alpha_{\nu,n})_{n\in\mathbb{N}}$ defined in (6.1.4). We will prove our statement by induction. The case $n = 0$ is met under the assumption $\kappa_{\psi^0}(r) \geq \alpha_\nu - T^1 - r^{-1}\hat{g}_\nu(r)$ The inductive step follows by applying consecutively Lemma 6.1.8 and Lemma 6.1.5. As a direct consequence of item (ii) in Lemma 6.1.6 we deduce that the sequence $(\alpha_{\nu,n})_{n\in\mathbb{N}}$ is non-decreasing and hence $\alpha_{\nu,n} \geq \alpha_{\nu,0} = \alpha_\nu - T^{-1}$. Since $G^{g_\mu,\hat{g}_\nu}_{\beta_\mu}$ is continuous and $\alpha_\nu - T^{-1}$ is not an accumulation point for the set of solutions of (6.1.13) (cf. item (iii) in Lemma 6.1.6), we deduce that $\alpha_{\nu,n} > \alpha_{\nu,0} = \alpha_\nu - T^{-1}$ for $n \geq 1$ and that the same holds for its limit $\alpha_{\psi^\star}$. The upper bound on $\alpha_{\nu,n}$ comes for free from (6.1.4) and the upper bound (6.1.15). The proof of (6.1.5) is obtained in the same way by considering the (constant) Sinkhorn's iterates $(\varphi^\star, \psi^\star)$ with the same sequence of $(\alpha_{\nu,n})_{n\in\mathbb{N}}$.

The proof of the second statement is completely analogous and for this reason we omit it. The only difference here relies in proving that the base case $n = 1$ holds true, but this has been already proven in the discussion of Remark 6.1.9. □

6.2 Wasserstein distance w.r.t a measure with asymptotically log-concave profile

In this section we consider two probability measures $\mathfrak{p}$, $\mathfrak{q} \in \mathcal{P}(\mathbb{R}^d)$ that can be again written with log-densities as

$$\mathfrak{p}(\mathrm{d}x) = \exp(-U_\mathfrak{p}(x))\mathrm{d}x\,, \qquad \mathfrak{q}(\mathrm{d}x) = \exp(-U_\mathfrak{q}(x))\mathrm{d}x\ .$$

AO1. *Assume that* $U_\mathfrak{p}$, $U_\mathfrak{q} \in \mathcal{C}^1(\mathbb{R}^d)$ *and that*

1. $U_\mathfrak{q}$ *is coercive, i.e., there exist* $\gamma_\mathfrak{q} > 0$ *and* $R_\mathfrak{q} \geq 0$ *such that*
$$-\frac{1}{2}\langle \nabla U_\mathfrak{q}(x), x\rangle \leq -\gamma_\mathfrak{q}\,|x|^2 \quad \forall\,|x| \geq R_\mathfrak{q}\ .$$
2. $U_\mathfrak{p}$ *has an integrated convex profile, i.e., there exist some* $\alpha_\mathfrak{p} > 0$ *and* $\tilde{g}_\mathfrak{p} \in \tilde{\mathcal{G}}$ *such that*
$$\kappa_{U_\mathfrak{p}}(r) \geq \alpha_\mathfrak{p} - r^{-1}\,\tilde{g}_\mathfrak{p}(r) \quad \forall r > 0\ .$$

Let us also emphasize here that the convexity of integrated profile assumption is stronger than the coercivity, since the former implies

$$-\frac{1}{2}\langle \nabla U_{\mathfrak{p}}(x), x\rangle \leq -\frac{\alpha_{\mathfrak{p}}}{2}|x|^2 + \frac{|\nabla U_{\mathfrak{p}}(0)| + \|\tilde{g}_{\mathfrak{p}}\|_\infty}{2}|x| ,$$

and hence that the coercive condition holds

$$-\frac{1}{2}\langle \nabla U_{\mathfrak{p}}(x), x\rangle \leq -\gamma_{\mathfrak{p}}|x|^2 \quad \forall |x| \geq R_{\mathfrak{p}}$$

for some $\gamma_{\mathfrak{p}} > 0$ and $R_{\mathfrak{p}} > 0$. Notice that $\mathfrak{p}$ and $\mathfrak{q}$ can be seen as invariant measures of the corresponding SDEs

$$\begin{cases} \mathrm{d}X_t = -\frac{1}{2}\nabla U_{\mathfrak{p}}(X_t)\,\mathrm{d}t + \mathrm{d}B_t , \\ \mathrm{d}Y_t = -\frac{1}{2}\nabla U_{\mathfrak{q}}(Y_t)\,\mathrm{d}t + \mathrm{d}B_t , \end{cases} \tag{6.2.1}$$

which admit unique strong solutions, in view of the coercivity of the corresponding drifts and owing to [RT96, Theorem 2.1]. Finally let $(P_t^{\mathfrak{p}})_{t\geq 0}$ and $(P_t^{\mathfrak{q}})_{t\geq 0}$ denote the corresponding Markov semigroups associated to the above SDEs. Since $\mathfrak{p}$ and $\mathfrak{q}$ are invariant measures we clearly have $\mathfrak{p}P_t^{\mathfrak{p}} = \mathfrak{p}$ and $\mathfrak{q}P_t^{\mathfrak{q}} = \mathfrak{q}$ for any $t \geq 0$.

The main result of this section is showing how the Wasserstein distance between $\mathfrak{p}$ and $\mathfrak{q}$ can be bounded w.r.t. the integrated difference of the drifts appearing in (6.2.1).

Theorem 6.2.1. *Assume **AO1**. Then it holds*

$$\begin{aligned} \mathbf{W}_1(\mathfrak{p},\ \mathfrak{q}) \leq& \left(\frac{\tilde{g}_{\mathfrak{p}}{}'(0)}{\tilde{g}_{\mathfrak{p}}{}'(R)}\right)^2 \frac{1}{\alpha_{\mathfrak{p}} + \tilde{g}_{\mathfrak{p}}{}'(0)} \int |\nabla U_{\mathfrak{p}} - \nabla U_{\mathfrak{q}}|\mathrm{d}\mathfrak{q} \\ =& \left(\frac{\tilde{g}_{\mathfrak{p}}{}'(0)}{\tilde{g}_{\mathfrak{p}}{}'(R)}\right)^2 \frac{1}{\alpha_{\mathfrak{p}} + \tilde{g}_{\mathfrak{p}}{}'(0)} \int \left|\nabla \log \frac{\mathrm{d}\mathfrak{p}}{\mathrm{d}\mathfrak{q}}\right| \mathrm{d}\mathfrak{q} , \end{aligned}$$

with $R := \|\tilde{g}_{\mathfrak{p}}\|_\infty \left((\tilde{g}_{\mathfrak{p}}{}'(0))^{-1} + 2/\alpha_{\mathfrak{p}}\right)$.

Proof. Firstly, let us consider the function

$$f_{\mathfrak{p}}(r) := \begin{cases} \tilde{g}_{\mathfrak{p}}(r) & \text{if } r \leq R , \\ \tilde{g}_{\mathfrak{p}}(R) + \tilde{g}_{\mathfrak{p}}{}'(R)\,(r-R) & \text{otherwise.} \end{cases}$$

Notice that $f(0) = 0$ and that $f_{\mathfrak{p}} \in \mathcal{C}^1((0,+\infty), \mathbb{R}_+) \cap \mathcal{C}^2((0,R) \cup (R,+\infty), \mathbb{R}_+)$ with $f_{\mathfrak{p}}{}''$ having a jump discontinuity in $r = R$. Moreover, $f_{\mathfrak{p}}$ is non-decreasing and concave and equivalent to the identity *i.e.*, for any $r > 0$ it holds

$$\tilde{g}_{\mathfrak{p}}{}'(R)\,r \leq f_{\mathfrak{p}}(r) \leq \tilde{g}_{\mathfrak{p}}{}'(0)\,r , \quad \text{since } \tilde{g}_{\mathfrak{p}}{}'(R) \leq f_{\mathfrak{p}}{}'(r) \leq \tilde{g}_{\mathfrak{p}}{}'(0) . \tag{6.2.2}$$

Notice that, since $\tilde{g}_{\mathfrak{p}} \in \tilde{\mathcal{G}}$, for any $r \in (0, R)$ it holds

$$\begin{aligned} 2\, f''_{\mathfrak{p}}(r) - \frac{r_t\, f'_{\mathfrak{p}}(r)}{2} \kappa_{U_{\mathfrak{p}}}(r) &\le 2\, \tilde{g}''_{\mathfrak{p}}(r) - \frac{\alpha_{\mathfrak{p}}}{2}\, \tilde{g}_{\mathfrak{p}}{}'(r)\, r + \frac{1}{2} \tilde{g}_{\mathfrak{p}}(r) \tilde{g}_{\mathfrak{p}}{}'(r) \\ &\le -\, \frac{\tilde{g}_{\mathfrak{p}}{}'(r)}{2} \left(\alpha_{\mathfrak{p}}\, r + \tilde{g}_{\mathfrak{p}}(r)\right) \\ &\le -\, \frac{\tilde{g}_{\mathfrak{p}}{}'(R)}{2} \left(\frac{\alpha_{\mathfrak{p}}}{\tilde{g}_{\mathfrak{p}}{}'(0)} + 1\right) \tilde{g}_{\mathfrak{p}}(r) \\ &= -\, \frac{\tilde{g}_{\mathfrak{p}}{}'(R)}{2} \left(\frac{\alpha_{\mathfrak{p}}}{\tilde{g}_{\mathfrak{p}}{}'(0)} + 1\right) f_{\mathfrak{p}}(r)\,; \end{aligned}$$

whereas for any $r > R$ it holds

$$\begin{aligned} 2\, f''_{\mathfrak{p}}(r) - \frac{r_t\, f'_{\mathfrak{p}}(r)}{2} \kappa_{U_{\mathfrak{p}}}(r) &= -\, \frac{\tilde{g}_{\mathfrak{p}}{}'(R)}{2}\, r\, \kappa_{U_{\mathfrak{p}}}(r) \le -\frac{\tilde{g}_{\mathfrak{p}}{}'(R)}{2} \left(\alpha_{\mathfrak{p}}\, r - \tilde{g}_{\mathfrak{p}}(r)\right) \\ &\le -\, \frac{\tilde{g}_{\mathfrak{p}}{}'(R)}{2}\, \alpha_{\mathfrak{p}}\, \frac{R}{\tilde{g}_{\mathfrak{p}}(R)}\, f_{\mathfrak{p}}(r) + \frac{\tilde{g}_{\mathfrak{p}}{}'(R)}{2}\, f_{\mathfrak{p}}(r)\,, \end{aligned}$$

where the last step follows from the concavity of $\tilde{g}_{\mathfrak{p}}$, which implies for any $r \ge R$ that $\tilde{g}_{\mathfrak{p}}(r) \le f_{\mathfrak{p}}(r)$ and the monotonicity of $F(r) := r/f(r)$, since for any $r \ge R$ it holds

$$F'(r) = \frac{f_{\mathfrak{p}}(r) - r\, f_{\mathfrak{p}}{}'(r)}{f_{\mathfrak{p}}(r)^2} = \frac{\tilde{g}_{\mathfrak{p}}(R) - R\, \tilde{g}_{\mathfrak{p}}{}'(R)}{f_{\mathfrak{p}}(r)^2} \ge 0\,.$$

Therefore, by recalling the definition of R, for any $r \in (0, R) \cup (R, +\infty)$ we have shown that

$$2\, f''_{\mathfrak{p}}(r) - \frac{r_t\, f'_{\mathfrak{p}}(r)}{2} \kappa_{U_{\mathfrak{p}}}(r) \le -\lambda_{\mathfrak{p}}\, f_{\mathfrak{p}}(r) \quad \text{with } \lambda_{\mathfrak{p}} := \frac{\tilde{g}_{\mathfrak{p}}{}'(R)}{2} \left(\frac{\alpha_{\mathfrak{p}}}{\tilde{g}_{\mathfrak{p}}{}'(0)} + 1\right). \tag{6.2.3}$$

By relying on the above construction, we will prove our thesis considering the Wasserstein distance

$$\mathbf{W}_{f_{\mathfrak{p}}}(\mathfrak{p}, \mathfrak{q}) := \inf_{\pi \in \Pi(\mathfrak{p}, \mathfrak{q})} \mathbb{E}_{(X,Y)\sim\pi}\left[f_{\mathfrak{p}}(|X - Y|)\right]$$

induced by the concave function $f_{\mathfrak{p}}$. The above is indeed a distance since $f_{\mathfrak{p}}(0) = 0$, $f_{\mathfrak{p}}$ is strictly increasing, concave and hence also subadditive (which implies the triangular inequality). Moreover, from (6.2.2) it follows the equivalence between $\mathbf{W}_{f_{\mathfrak{p}}}$ and the usual $\mathbf{W}_1$, namely

$$\tilde{g}_{\mathfrak{p}}{}'(R)\, \mathbf{W}_1(\mathfrak{p}, \mathfrak{q}) \le \mathbf{W}_{f_{\mathfrak{p}}}(\mathfrak{p}, \mathfrak{q}) \le \tilde{g}_{\mathfrak{p}}{}'(0)\, \mathbf{W}_1(\mathfrak{p}, \mathfrak{q})\,, \tag{6.2.4}$$

For any $t \ge 0$ notice that

$$\mathbf{W}_{f_{\mathfrak{p}}}(\mathfrak{p}, \mathfrak{q}) \le \mathbf{W}_{f_{\mathfrak{p}}}(\mathfrak{p}, \mathfrak{q}\, P^{\mathfrak{p}}_t) + \mathbf{W}_{f_{\mathfrak{p}}}(\mathfrak{q}\, P^{\mathfrak{p}}_t, \mathfrak{q}) = \mathbf{W}_{f_{\mathfrak{p}}}(\mathfrak{p}\, P^{\mathfrak{p}}_t, \mathfrak{q}\, P^{\mathfrak{p}}_t) + \mathbf{W}_{f_{\mathfrak{p}}}(\mathfrak{q}\, P^{\mathfrak{p}}_t, \mathfrak{q})\,. \tag{6.2.5}$$

In order to bound the first Wasserstein distance appearing in the upper bound (6.2.5), namely $\mathbf{W}_{f_\mathfrak{p}}(\mathfrak{p}\, P_t^\mathfrak{p},\, \mathfrak{q}\, P_t^\mathfrak{p})$, fix an initial coupling $(X_0, X_0^\mathfrak{q}) \sim \pi^\star$ distributed according to the optimal coupling for $\mathbf{W}_{f_\mathfrak{p}}(\mathfrak{p}, \mathfrak{q})$ and consider the reflection coupling diffusion processes

$$\begin{cases} \mathrm{d}X_t = -\frac{1}{2}\nabla U_\mathfrak{p}(X_t)\,\mathrm{d}t + \mathrm{d}B_t \\ \mathrm{d}X_t^\mathfrak{q} = -\frac{1}{2}\nabla U_\mathfrak{p}(X_t^\mathfrak{q})\,\mathrm{d}t + \mathrm{d}\hat{B}_t \quad \forall\, t \in [0,\tau) \text{ and } X_t = X_t^\mathfrak{q} \quad \forall\, t \geq \tau \\ (X_0, X_0^\mathfrak{q}) \sim \pi^\star\,, \end{cases}$$

where $\tau := \inf\{s \geq 0 : X_s^\mathfrak{q} = X_s\}$, and $(\hat{B}_t)_{t\geq 0}$ is defined as

$$\mathrm{d}\hat{B}_t := (\mathrm{Id} - 2\, e_t\, e_t^\top\, \mathbf{1}_{\{t<\tau\}})\,\mathrm{d}B_t \qquad \text{where} \quad e_t := \begin{cases} \frac{Z_t}{|Z_t|} & \text{when } r_t > 0\,, \\ u & \text{when } r_t = 0\,. \end{cases}$$

where $Z_t := X_t - X_t^\mathfrak{q}$, $r_t := |Z_t|$ and $u \in \mathbb{R}^d$ is a fixed (arbitrary) unit-vector. By Lévy's characterisation, $(\hat{B}_t)_{t\geq 0}$ is a d-dimensional Brownian motion. As a result, $X_t \sim \mathfrak{p}$ and $X_t^\mathfrak{q} \sim \mathfrak{q}P_t^\mathfrak{p}$ for any $t \geq 0$. In addition $\mathrm{d}W_t := e_t^\top \mathrm{d}B_t$ is a one-dimensional Brownian motion. Let us notice that for any $t < \tau$ it holds

$$\begin{aligned} \mathrm{d}Z_t &= -\,2^{-1}\,(\nabla U_\mathfrak{p}(X_t) - \nabla U_\mathfrak{p}(X_t^\mathfrak{q}))\,\mathrm{d}t + 2\,e_t\,\mathrm{d}W_t\,, \\ \mathrm{d}r_t^2 &= -\,\langle Z_t,\, \nabla U_\mathfrak{p}(X_t) - \nabla U_\mathfrak{p}(X_t^\mathfrak{q})\rangle\,\mathrm{d}t + 4\,\mathrm{d}t + 4\,\langle Z_t,\, e_t\rangle\,\mathrm{d}W_t\,, \\ \mathrm{d}r_t &= -\,2^{-1}\,\langle e_t,\, \nabla U_\mathfrak{p}(X_t) - \nabla U_\mathfrak{p}(X_t^\mathfrak{q})\rangle\,\mathrm{d}t + 2\,\mathrm{d}W_t\,. \end{aligned}$$

Now, an application of Ito-Tanaka formula [RY99, Chapter VI, Theorem 1.5] to the concave function $f_\mathfrak{p} \in \mathcal{C}^1((0,+\infty),\mathbb{R}_+) \cap \mathcal{C}^2((0,R)\cup(R,+\infty),\mathbb{R}_+)$, gives for any $t < \tau$

$$f_\mathfrak{p}(r_t) = f_\mathfrak{p}(r_0) + \int_0^t f_\mathfrak{p}{}'(r_s)\,\mathrm{d}r_s + \frac{1}{2}\int_\mathbb{R} L_t^a\,\mu_{f_\mathfrak{p}}(\mathrm{d}a)\,, \tag{6.2.6}$$

where $(L_t^a)_t$ denotes the right-continuous local time of the semimartingale $(r_t)_t$, whereas $\mu_{f_\mathfrak{p}}$ is the non-positive measure representing $f_\mathfrak{p}{}''$ in the sense of distributions, *i.e.*, $\mu_{f_\mathfrak{p}}([a,b]) = f_\mathfrak{p}{}'(b) - f_\mathfrak{p}{}'(a)$ for any $a \leq b$. Let us further notice that the Meyer Wang occupation times formula [Kal21, Theorem 29.5], which for any measurable function $H\colon \mathbb{R} \to [0,+\infty)$ reads as

$$\int_0^t H(r_s)\,\mathrm{d}[r]_s = \int_\mathbb{R} H(a) L_t^a\,\mathrm{d}a\,,$$

implies that the random set $\{s \in [0,\tau]\colon r_s = R\}$ has almost surely zero Lebesgue measure. Particularly, since $\mu_{f_\mathfrak{p}}$ is non-positive, this combined with the above formula implies

$$\begin{aligned} \frac{1}{2}\int_\mathbb{R} L_t^a\,\mu_{f_\mathfrak{p}}(\mathrm{d}a) \leq \frac{1}{2}\int_\mathbb{R} \mathbf{1}_{\{a\neq R\}} L_t^a f_\mathfrak{p}{}''(a)\mathrm{d}a &= 2\int_0^t \mathbf{1}_{\{r_s\neq R\}} f_\mathfrak{p}{}''(r_s)\mathrm{d}s \\ &= 2\int_0^t f_\mathfrak{p}{}''(r_s)\mathrm{d}s. \end{aligned} \tag{6.2.7}$$

As a byproduct of (6.2.6) and (6.2.7) we may finally state that for any $t < \tau$ it almost surely holds

$$
\begin{aligned}
\mathrm{d}f_{\mathfrak{p}}(r_t) &\leq -\frac{f_{\mathfrak{p}}'(r_t)}{2}\,\langle e_t,\ \nabla U_{\mathfrak{p}}(X_t) - \nabla U_{\mathfrak{p}}(X_t^{\mathfrak{q}})\rangle\,\mathrm{d}t + 2\,f_{\mathfrak{p}}''(r_t)\,\mathrm{d}t + 2\,f_{\mathfrak{p}}'(r_t)\,\mathrm{d}W_t \\
&\leq \left(2\,f_{\mathfrak{p}}''(r_t) - \frac{r_t\,f_{\mathfrak{p}}'(r_t)}{2}\kappa_{U_{\mathfrak{p}}}(r_t)\right)\mathrm{d}t + 2\,f_{\mathfrak{p}}'(r_t)\,\mathrm{d}W_t \\
&\overset{(6.2.3)}{\leq} -\lambda_{\mathfrak{p}}\,f_{\mathfrak{p}}(r_t)\,\mathrm{d}t + 2\,f_{\mathfrak{p}}'(r_t)\,\mathrm{d}W_t\,.
\end{aligned}
$$

By recalling that $f_{\mathfrak{p}}(r_t) = f_{\mathfrak{p}}(0) = 0$ as soon as $t \geq \tau$, by taking expectation, integrating over time and by applying Gronwall Lemma we have finally proven that for any $t \geq 0$ it holds

$$
\begin{aligned}
\mathbf{W}_{f_{\mathfrak{p}}}(\mathfrak{p}\,P_t^{\mathfrak{p}},\,\mathfrak{q}\,P_t^{\mathfrak{p}}) \leq \mathbb{E}[f_{\mathfrak{p}}(|X_t - X_t^{\mathfrak{q}}|)] &\leq e^{-\lambda_{\mathfrak{p}}\,t}\,\mathbb{E}[f_{\mathfrak{p}}(|X_0 - X_0^{\mathfrak{q}}|)] \\
&= e^{-\lambda_{\mathfrak{p}}\,t}\,\mathbf{W}_{f_{\mathfrak{p}}}(\mathfrak{p},\,\mathfrak{q})\,.
\end{aligned}
\tag{6.2.8}
$$

Let us now provide a bound for the second Wasserstein distance appearing in the upper bound (6.2.5), namely $\mathbf{W}_{f_{\mathfrak{p}}}(\mathfrak{q}\,P_t^{\mathfrak{p}},\,\mathfrak{q})$, by relying on the synchronous coupling technique. Therefore fix an initial random variable $Y_0 \sim \mathfrak{q}$ and a d-dimensional Brownian motion $(B_t)_{t\geq 0}$, and consider now the diffusion processes

$$
\begin{cases}
\mathrm{d}X_t^{\mathfrak{q}} = -\frac{1}{2}\nabla U_{\mathfrak{p}}(X_t^{\mathfrak{q}})\,\mathrm{d}t + \mathrm{d}B_t \\
\mathrm{d}Y_t = -\frac{1}{2}\nabla U_{\mathfrak{q}}(Y_t)\,\mathrm{d}t + \mathrm{d}B_t \\
X_0^{\mathfrak{q}} = Y_0 = Y_0 \sim \mathfrak{q}\,.
\end{cases}
$$

Notice that for any $t \geq 0$, $Y_t \sim \mathfrak{q}$ whereas $X_t^{\mathfrak{q}} \sim \mathfrak{q}P_t^{\mathfrak{p}}$. If we set now $\bar{r}_t := |X_t^{\mathfrak{q}} - Y_t|$ we then have

$$
\begin{aligned}
\mathrm{d}(X_t^{\mathfrak{q}} - Y_t) &= -2^{-1}(\nabla U_{\mathfrak{p}}(X_t^{\mathfrak{q}}) - \nabla U_{\mathfrak{q}}(Y_t))\,\mathrm{d}t\,, \\
\mathrm{d}\bar{r}_t^2 &= -\langle X_t^{\mathfrak{q}} - Y_t,\ \nabla U_{\mathfrak{p}}(X_t^{\mathfrak{q}}) - \nabla U_{\mathfrak{q}}(Y_t)\rangle\,\mathrm{d}t\,.
\end{aligned}
$$

At this point we would like to apply the square-root function, however the latter fails to be $\mathcal{C}^2$ in the origin whereas $\bar{r}_t$ may be equal to zero (e.g., we already start with $\bar{r}_0 = 0$). For this reason we are going to perform an approximation argument. Fix $\delta > 0$ and consider the function $\rho_\delta(r) := \sqrt{r+\delta}$. Then it holds

$$
\begin{aligned}
\mathrm{d}\rho_\delta(\bar{r}_t^2) &= -(2\,\rho_\delta(\bar{r}_t^2))^{-1}\langle X_t^{\mathfrak{q}} - Y_t,\ \nabla U_{\mathfrak{p}}(X_t^{\mathfrak{q}}) - \nabla U_{\mathfrak{q}}(Y_t)\rangle\,\mathrm{d}t \\
&= -(2\,\rho_\delta(\bar{r}_t^2))^{-1}\langle X_t^{\mathfrak{q}} - Y_t,\ \nabla U_{\mathfrak{p}}(X_t^{\mathfrak{q}}) - \nabla U_{\mathfrak{p}}(Y_t)\rangle\,\mathrm{d}t \\
&\qquad\qquad - (2\,\rho_\delta(\bar{r}_t^2))^{-1}\langle X_t^{\mathfrak{q}} - Y_t,\ \nabla U_{\mathfrak{p}}(Y_t) - \nabla U_{\mathfrak{q}}(Y_t)\rangle\,\mathrm{d}t \\
&\leq -2^{-1}\,\frac{\bar{r}_t^2}{\rho_\delta(\bar{r}_t^2)}\,\kappa_{\mathfrak{p}}(\bar{r}_t)\,\mathrm{d}t + 2^{-1}\,\frac{\bar{r}_t}{\rho_\delta(\bar{r}_t^2)}\,|\nabla U_{\mathfrak{p}} - \nabla U_{\mathfrak{q}}|(Y_t)\mathrm{d}t \\
&\leq -2^{-1}\,\frac{\bar{r}_t^2}{\rho_\delta(\bar{r}_t^2)}\,(\alpha_{\mathfrak{p}} - \tilde{G}_{\mathfrak{p}})\,\mathrm{d}t + 2^{-1}\,\frac{\bar{r}_t}{\rho_\delta(\bar{r}_t^2)}\,|\nabla U_{\mathfrak{p}} - \nabla U_{\mathfrak{q}}|(Y_t)\mathrm{d}t\,,
\end{aligned}
$$

where in the last step we have relied on the sublinearity of $\tilde{g}_\mathfrak{p} \in \tilde{\mathcal{G}}$ (namely that $\tilde{g}_\mathfrak{p}(r) \leq \tilde{G}_\mathfrak{p}\, r$ for some positive constant $\tilde{G}_\mathfrak{p} > 0$). Therefore it holds

$$\mathrm{d}\rho_\delta(\bar{r}_t^2) \leq \frac{(\alpha_\mathfrak{p} - \tilde{G}_\mathfrak{p})^+}{2}\, \rho_\delta(\bar{r}_t^2)\mathrm{d}t \, + 2^{-1}|\nabla U_\mathfrak{p} - \nabla U_\mathfrak{q}|(Y_t)\mathrm{d}t\,.$$

By taking expectation and integrating over time the above bound gives

$$\begin{aligned}\mathbb{E}[\rho_\delta(\bar{r}_t^2)] &\leq \mathbb{E}[\rho_\delta(\bar{r}_0^2)] + \frac{(\alpha_\mathfrak{p} - \tilde{G}_\mathfrak{p})^+}{2}\int_0^t \mathbb{E}[\rho_\delta(\bar{r}_s^2)]\,\mathrm{d}s \\ &\qquad + 2^{-1}\int_0^t \mathbb{E}[|\nabla U_\mathfrak{p} - \nabla U_\mathfrak{q}|(Y_s)]\,\mathrm{d}s \\ &= \sqrt{\delta} + \frac{(\alpha_\mathfrak{p} - \tilde{G}_\mathfrak{p})^+}{2}\int_0^t \mathbb{E}[\rho_\delta(\bar{r}_s^2)]\,\mathrm{d}s + \frac{t}{2}\int |\nabla U_\mathfrak{p} - \nabla U_\mathfrak{q}|\,\mathrm{d}\mathfrak{q}\,,\end{aligned}$$

where in the last step we have relied on the fact that $Y_t \sim \mathfrak{q}$ for any $t \geq 0$ and that $\bar{r}_0 = 0$. Therefore Gronwall Lemma yields to

$$\mathbb{E}[\bar{r}_t] \leq \mathbb{E}[\rho_\delta(\bar{r}_t^2)] \leq \exp\left(\frac{t}{2}(\alpha_\mathfrak{p} - \tilde{G}_\mathfrak{p})^+\right)\left[\frac{t}{2}\int |\nabla U_\mathfrak{p} - \nabla U_\mathfrak{q}|\mathrm{d}\mathfrak{q} + \sqrt{\delta}\right].$$

By letting δ to zero in the above right-hand-side, we obtain the desired upper bound

$$\begin{aligned}\mathbf{W}_{f_\mathfrak{p}}(\mathfrak{q}\,P_t^\mathfrak{p}, \mathfrak{q}) &\overset{(6.2.4)}{\leq} \tilde{g}_\mathfrak{p}{}'(0)\,\mathbf{W}_1(\mathfrak{q}\,P_t^\mathfrak{p}, \mathfrak{q}) \leq \tilde{g}_\mathfrak{p}{}'(0)\,\mathbb{E}[|X_t^\mathfrak{q} - Y_t|] = \tilde{g}_\mathfrak{p}{}'(0)\,\mathbb{E}[\bar{r}_t] \\ &\leq \tilde{g}_\mathfrak{p}{}'(0)\,\frac{t}{2}\,\exp\left(\frac{t}{2}\,(\alpha_\mathfrak{p} - \tilde{G}_\mathfrak{p})^+\right)\int |\nabla U_\mathfrak{p} - \nabla U_\mathfrak{q}|\,\mathrm{d}\mathfrak{q}\,.\end{aligned}$$

By putting together the last estimate with (6.2.5) and (6.2.8) we have proven that

$$\begin{aligned}\mathbf{W}_{f_\mathfrak{p}}(\mathfrak{p}, \mathfrak{q}) &\leq e^{-\lambda_\mathfrak{p}\, t}\,\mathbf{W}_{f_\mathfrak{p}}(\mathfrak{p}, \mathfrak{q}) \\ &\qquad + \tilde{g}_\mathfrak{p}{}'(0)\,\frac{t}{2}\,\exp\left(\frac{t}{2}\,(\alpha_\mathfrak{p} - \tilde{G}_\mathfrak{p})^+\right)\int |\nabla U_\mathfrak{p} - \nabla U_\mathfrak{q}|\,\mathrm{d}\mathfrak{q}\,,\end{aligned}$$

or equivalently that

$$\mathbf{W}_{f_\mathfrak{p}}(\mathfrak{p}, \mathfrak{q}) \leq \frac{t/2}{1 - e^{-\lambda_\mathfrak{p}\, t}}\,\tilde{g}_\mathfrak{p}{}'(0)\,\exp\left(\frac{t}{2}\,(\alpha_\mathfrak{p} - \tilde{G}_\mathfrak{p})^+\right)\int |\nabla U_\mathfrak{p} - \nabla U_\mathfrak{q}|\,\mathrm{d}\mathfrak{q}\,,$$

which in the t vanishing limit reads as

$$\mathbf{W}_{f_\mathfrak{p}}(\mathfrak{p}, \mathfrak{q}) \leq \frac{\tilde{g}_\mathfrak{p}{}'(0)}{2\,\lambda_\mathfrak{p}}\int |\nabla U_\mathfrak{p} - \nabla U_\mathfrak{q}|\,\mathrm{d}\mathfrak{q}\,.$$

Combining the above bound with the equivalence (6.2.4) concludes the proof. □

Remark 6.2.2 (Explicit constants for $\tilde{g}_{\mathfrak{p}}$ as in (6.0.14))**.** *Particularly, when $\tilde{g}_{\mathfrak{p}}$ is as in* (6.0.14) *in Remark 6.0.1, i.e.*

$$\tilde{g}_{\mathfrak{p}}(r) = 2\,(L)^{1/2}\,\tanh(r\,L^{1/2}/2)$$

for some L, then in the previous proof we have $R = 2(L)^{1/2}(L^{-1} + 2/\alpha_{\mathfrak{p}})$ and therefore

$$\mathbf{W}_1(\mathfrak{p},\ \mathfrak{q}) \le \frac{\cosh^4(2(L)^{1/2}(L^{-1} + 2/\alpha_{\mathfrak{p}}))}{\alpha_{\mathfrak{p}} + L} \int |\nabla U_{\mathfrak{p}} - \nabla U_{\mathfrak{q}}| \mathrm{d}\mathfrak{q}\,.$$

Let us conclude this section by showing how the previous result and the propagation of lower-bounds for integrated convexity profiles (studied in the previous section) yield to the key estimates that will be employed in the proof of the main results.

Corollary 6.2.3. *Assume A6. Then, for any $x \in \mathbb{R}^d$, it holds*

$$\mathbf{W}_1(\pi_T^{x,\psi^n}, \pi_T^{x,\psi^\star}) \le \gamma_n^\nu \int |\nabla\psi^n - \nabla\psi^\star|\, \mathrm{d}\pi_T^{x,\psi^\star}, \tag{6.2.9}$$

and similarly

$$\mathbf{W}_1(\pi_T^{x,\varphi^{n+1}}, \pi_T^{x,\varphi^\star}) \le \gamma_n^\mu \int \left|\nabla\varphi^{n+1} - \nabla\varphi^\star\right| \mathrm{d}\pi_T^{x,\varphi^\star}, \tag{6.2.10}$$

where for any $\mathfrak{p} \in \{\mu, \nu\}$, and for any $n \in \mathbb{N}$ it holds

$$\gamma_n^{\mathfrak{p}} := \frac{\tilde{g}_{\mathfrak{p}}{}'(0)^2}{\tilde{g}_{\mathfrak{p}}{}'\left(\|\tilde{g}_{\mathfrak{p}}\|_\infty \left(\frac{1}{\tilde{g}_{\mathfrak{p}}{}'(0)} + \frac{2}{\alpha_{\mathfrak{p},n} + T^{-1}}\right)\right)^2} \frac{1}{\alpha_{\mathfrak{p},n} + T^{-1} + \tilde{g}_{\mathfrak{p}}{}'(0)} \tag{6.2.11}$$

where $(\alpha_{\mu,n})_{n\in\mathbb{N}} \subseteq (\alpha_\mu - T^{-1}, \alpha_\mu - T^{-1} + (\beta_\nu T^2)^{-1}]$ and $(\alpha_{\nu,n})_{n\in\mathbb{N}} \subseteq (\alpha_\nu - T^{-1}, \alpha_\nu - T^{-1} + (\beta_\mu T^2)^{-1}]$ are the monotone increasing sequences built in Theorem 6.1.1.

Proof. Inequality (6.2.9) follows from the previous theorem when considering $\mathfrak{p} = \pi_T^{x,\psi^n}$ and $\mathfrak{q} = \pi_T^{x,\psi^\star}$. Indeed these two probabilities are the invariant measures associated to (6.2.1) with $U_{\mathfrak{p}}(y) = (2T)^{-1}\,|y - x|^2 + \psi^n(y)$ and $U_{\mathfrak{q}}(y) = (2T)^{-1}\,|y - x|^2 + \psi^\star(y)$ respectively. Theorem 6.1.1 guarantees that there exist $\alpha_{\nu,n},\ \alpha_{\psi^\star} > -T^{-1}$ such that

$$\kappa_{U_{\mathfrak{p}}}(r) \ge T^{-1} + \alpha_{\nu,n} - r^{-1}\tilde{g}_\nu(r) \quad \text{and} \quad \kappa_{U_{\mathfrak{q}}}(r) \ge T^{-1} + \alpha_{\psi^\star} - r^{-1}\tilde{g}_\nu(r)\,.$$

Therefore $\mathfrak{p}$ and $\mathfrak{q}$ satisfy Assumption **AO**1 and Theorem 6.2.1 gives (6.2.9).

We omit the details for the proof of (6.2.10) since it can be obtained in the same way, this time considering $\mathfrak{p} = \pi_T^{x,\varphi^{n+1}}$ and $\mathfrak{q} = \pi_T^{x,\varphi^\star}$. □

6.3 Exponential convergence of the gradients and in $\mathbf{W}_1$ along Sinkhorn's algorithm

The convergence of the gradients will follow by iterating the following result.

Proposition 6.3.1. *Assume A6 holds. Then for any $n \geq 0$ it holds*

$$\int |\nabla\varphi^{n+1} - \nabla\varphi^\star| \mathrm{d}\mu \leq T^{-1}\,\gamma_n^\nu \int |\nabla\psi^n - \nabla\psi^\star| \mathrm{d}\nu\,, \tag{6.3.1}$$

and similarly

$$\int |\nabla\psi^{n+1} - \nabla\psi^\star| \mathrm{d}\nu \leq T^{-1}\,\gamma_n^\mu \int |\nabla\varphi^{n+1} - \nabla\varphi^\star| \mathrm{d}\mu\,, \tag{6.3.2}$$

where γ_n^μ and γ_n^ν are given in (6.2.11).

Proof. Let us start by showing (6.3.1). From (6.0.1) we immediatly have

$$\begin{cases} \varphi^{n+1} - \varphi^\star = \log P_T \exp(-\psi^n) - \log P_T \exp(-\psi^\star) \\ \psi^{n+1} - \psi^\star = \log P_T \exp(-\varphi^{n+1}) - \log P_T \exp(-\varphi^\star) \end{cases}$$

and since for $h = \varphi^{n+1}$, $\varphi^\star$ the gradient along the semigroup has the explicit formulation (cf. Proposition 6.A.2)

$$\nabla \log P_T \exp(-h)(x) = \frac{1}{T}\int (y-x)\,\pi_T^{x,h}(\mathrm{d}y)\,,$$

and we may deduce that

$$\begin{aligned} |\nabla\varphi^{n+1} - \nabla\varphi^\star|(x) &= T^{-1}\left|\int y\,\pi_T^{x,\psi^n}(\mathrm{d}y) - \int y\,\pi_T^{x,\psi^\star}(\mathrm{d}y)\right| \\ &\leq T^{-1}\mathbf{W}_1(\pi_T^{x,\psi^n}, \pi_T^{x,\psi^\star})\,. \end{aligned} \tag{6.3.3}$$

Combining the above observation with Corollary 6.2.3 we end up with

$$|\nabla\varphi^{n+1} - \nabla\varphi^\star|(x) \leq \; T^{-1}\,\gamma_n^\nu \int |\nabla\psi^n - \nabla\psi^\star|\,\mathrm{d}\pi_T^{x,\psi^\star}, \tag{6.3.4}$$

with γ_n^ν as introduced in (6.2.11). Combining (6.0.6) with (6.3.4) gives (6.3.1). The contractive estimate (6.3.2) can be proven in the same manner by relying on (6.2.10) by noticing that

$$|\nabla\psi^{n+1} - \nabla\psi^\star|(x) \leq T^{-1}\mathbf{W}_1(\pi_T^{x,\varphi^{n+1}}, \pi_T^{x,\varphi^\star}) \;\text{ and }\; \int \pi_T^{x,\varphi^\star}(\mathrm{d}y)\,\nu(\mathrm{d}x) = \mu(\mathrm{d}y).$$

□

As a corollary we immediately deduce our first exponential convergence result.

Theorem 6.3.2. *Assume that A6 holds. Then for any $n \geq 1$*

$$\begin{aligned}\int |\nabla\varphi^n - \nabla\varphi^\star| \mathrm{d}\mu &\leq \frac{T}{\gamma^\mu_{n-1}} \prod_{k=0}^{n-1} \frac{\gamma^\mu_k\, \gamma^\nu_k}{T^2} \int |\nabla\psi^0 - \nabla\psi^\star| \mathrm{d}\nu \;, \\ \int |\nabla\psi^n - \nabla\psi^\star| \mathrm{d}\nu &\leq \prod_{k=0}^{n-1} \frac{\gamma^\mu_k\, \gamma^\nu_k}{T^2} \int |\nabla\psi^0 - \nabla\psi^\star| \mathrm{d}\nu \;,\end{aligned} \tag{6.3.5}$$

and

$$\begin{aligned}\mathbf{W}_1(\pi^{n,n}, \boldsymbol{\pi}^T) &\leq T \prod_{k=0}^{n-1} \frac{\gamma^\mu_k\, \gamma^\nu_k}{T^2} \int |\nabla\psi^0 - \nabla\psi^\star| \mathrm{d}\nu \;, \\ \mathbf{W}_1(\pi^{n+1,n}, \boldsymbol{\pi}^T) &\leq \gamma^\nu_n \prod_{k=0}^{n-1} \frac{\gamma^\mu_k\, \gamma^\nu_k}{T^2} \int |\nabla\psi^0 - \nabla\psi^\star| \mathrm{d}\nu \;,\end{aligned} \tag{6.3.6}$$

where $(\gamma^\mu_k)_{k\in\mathbb{N}}$ and $(\gamma^\nu_k)_{k\in\mathbb{N}}$ are non-negative non-increasing sequences defined at (6.2.11)*, depending on α_μ, β_ν, $\tilde{g}_\mu$, g_ν, T and on α_ν, β_μ, $\tilde{g}_\nu$, g_μ, T respectively. Denoting $\gamma^\zeta_\infty := \lim_{k\to+\infty} \gamma^\zeta_k$ for $\zeta \in \{\mu,\nu\}$, as a corollary, as soon as $T^2 > \gamma^\mu_\infty\, \gamma^\nu_\infty$, the asymptotic rate is strictly less than one, and for any $T^{-2}\gamma^\mu_\infty\, \gamma^\nu_\infty < \lambda < 1$, there exists $C \geq 0$ such that for any $n \in \mathbb{N}^*$,*

$$\begin{aligned}\int |\nabla\varphi^n - \nabla\varphi^\star| \mathrm{d}\mu + \int |\nabla\psi^n - \nabla\psi^\star| \mathrm{d}\nu &\leq C\lambda^n \int |\nabla\psi^0 - \nabla\psi^\star| \mathrm{d}\nu \;, \\ \mathbf{W}_1(\pi^{n,n}, \boldsymbol{\pi}^T) + \mathbf{W}_1(\pi^{n+1,n}, \boldsymbol{\pi}^T) &\leq C\lambda^n \int |\nabla\psi^0 - \nabla\psi^\star| \mathrm{d}\nu \;,\end{aligned}$$

Remark 6.3.3 (Explicit rates). *Let us simply remark here that the sequences $(\gamma^\mu_n)_{n\in\mathbb{N}}$ and $(\gamma^\nu_n)_{n\in\mathbb{N}}$ given in* (6.2.11) *are non-increasing, i.e., they provide faster convergence as the index n increases. Indeed this follows from the fact that $(\alpha_{\mu,n})_{n\in\mathbb{N}} \subseteq (\alpha_\mu - T^{-1}, \alpha_\mu - T^{-1} + (\beta_\nu\, T^2)^{-1}]$ and $(\alpha_{\nu,n})_{n\in\mathbb{N}} \subseteq (\alpha_\nu - T^{-1}, \alpha_\nu - T^{-1} + (\beta_\mu\, T^2)^{-1}]$ are monotone increasing sequences, built in Theorem 6.1.1. If $\alpha^\star_\varphi$ and $\alpha^\star_\psi$ denote their respective limits, then the asymptotic rates of convergence read as*

$$\begin{aligned}\gamma^\mu_\infty &:= \frac{\tilde{g}_\mu{}'(0)^2}{\tilde{g}_\mu{}'\left(\|\tilde{g}_\mu\|_\infty \left(\frac{1}{\tilde{g}_\mu{}'(0)} + \frac{2}{\alpha_{\varphi^\star} + T^{-1}}\right)\right)^2} \frac{1}{\alpha_{\varphi^\star} + T^{-1} + \tilde{g}_\mu{}'(0)} \;, \\ \gamma^\nu_\infty &:= \frac{\tilde{g}_\nu{}'(0)^2}{\tilde{g}_\nu{}'\left(\|\tilde{g}_\nu\|_\infty \left(\frac{1}{\tilde{g}_\nu{}'(0)} + \frac{2}{\alpha_{\psi^\star} + T^{-1}}\right)\right)^2} \frac{1}{\alpha_{\psi^\star} + T^{-1} + \tilde{g}_\nu{}'(0)} \;.\end{aligned}$$

From the above expressions it can be deduced that the condition for exponential convergence $T^{-2}\gamma^\mu_\infty\gamma^\nu_\infty < 1$ is always satisfied for large enough values of T. To be more precise, a sufficient condition for the exponential convergence of Sinkhorn algorithm is *the following*

$$T^2 > \frac{\tilde{g}_\mu{}'(0)\,(\alpha_\mu + \tilde{g}_\mu{}'(0))^{-1}}{\tilde{g}_\mu{}'\left(\|\tilde{g}_\mu\|_\infty\left(\frac{1}{\tilde{g}_\mu{}'(0)} + \frac{2}{\alpha_\mu}\right)\right)^2}\;\frac{\tilde{g}_\nu{}'(0)\,(\alpha_\nu + \tilde{g}_\nu{}'(0))^{-1}}{\tilde{g}_\nu{}'\left(\|\tilde{g}_\nu\|_\infty\left(\frac{1}{\tilde{g}_\nu{}'(0)} + \frac{2}{\alpha_\nu}\right)\right)^2}\,. \tag{6.3.7}$$

Expressions simplify when considering $\tilde{g}_\mathfrak{p}$ *as in* (6.0.14) *in Remark 6.0.1, i.e.*

$$\tilde{g}_\mathfrak{p}(r) = 2\,(L_\mathfrak{p})^{1/2}\,\tanh(r\,L_\mathfrak{p}^{1/2}/2)$$

for some $L_\mathfrak{p} \geq 0$*, since the previous asymptotic rates would read as*

$$\gamma^\mu_\infty = \frac{1}{\alpha_{\varphi^\star} + T^{-1} + L_\mu}\;\cosh^4\left(2(L_\mu)^{1/2}\left(\frac{1}{L_\mu} + \frac{2}{\alpha_{\varphi^\star} + T^{-1}}\right)\right),$$
$$\gamma^\nu_\infty = \frac{1}{\alpha_{\psi^\star} + T^{-1} + L_\nu}\;\cosh^4\left(2(L_\nu)^{1/2}\left(\frac{1}{L_\nu} + \frac{2}{\alpha_{\psi^\star} + T^{-1}}\right)\right),$$

whereas (6.3.7) *would read as*

$$T^2 > \frac{\cosh^4\left(\frac{2}{L_\mu^{1/2}} + \frac{4\,L_\mu^{1/2}}{\alpha_\mu}\right)}{\alpha_\mu + L_\mu}\;\frac{\cosh^4\left(\frac{2}{L_\nu^{1/2}} + \frac{4\,L_\nu^{1/2}}{\alpha_\nu}\right)}{\alpha_\nu + L_\nu}\,.$$

Finally, explicit expressions for $(\gamma^\mu_k)_{k\in\mathbb{N}}$, $(\gamma^\nu_k)_{k\in\mathbb{N}}$ *and their limits fully simplify under **A7**, i.e., when* μ *and* ν *are strongly log-concave. We focus on this particular scenario in Section 6.6 where we give simple expression for* $(\gamma^\mu_k)_{k\in\mathbb{N}}$, $(\gamma^\nu_k)_{k\in\mathbb{N}}$, γ^μ_∞, γ^ν_∞*. In particular we show that Sinkhorn's algorithm converges as soon as*

$$T > \frac{\beta_\mu\beta_\nu - \alpha_\mu\alpha_\nu}{\sqrt{\alpha_\mu\,\beta_\mu\,\alpha_\nu\,\beta_\nu\,(\alpha_\mu + \beta_\mu)(\alpha_\nu + \beta_\nu)}}\,.$$

As $\beta_\mu = \beta_\nu = +\infty$*, this expression simply reads as* $T > (\alpha_\mu\,\alpha_\nu)^{-1/2}$*. Lastly, notice that in the Gaussian quadratic case, that is when* $\beta_\mu = \alpha_\mu$ *and* $\beta_\nu = \alpha_\nu$*, then the exponential convergence condition reads as* $T > 0$ *which means that Sinkhorn's algorithm converges exponentially fast for any fixed* $T > 0$*.*

Proof of Theorem 6.3.2. As concerns (6.3.5), it is enough concatenating the bounds proven in Proposition 6.3.1, as already sketched at (6.0.8) and (6.0.9).

The $\mathbf{W}_1$-convergence bound (6.3.6) can be deduce from (6.3.5) and Corollary 6.2.3 since $\pi^{n,n} \in \Pi(\star,\nu)$, $\pi^{n+1,n} \in \Pi(\mu,\star)$ and hence

$$\mathbf{W}_1(\pi^{n,n}, \boldsymbol{\pi}^T) \leq \int \mathbf{W}_1(\pi_T^{x,\varphi^n}, \pi_T^{x,\varphi^\star})\,\mathrm{d}\nu\,,$$
$$\mathbf{W}_1(\pi^{n+1,n}, \boldsymbol{\pi}^T) \leq \int \mathbf{W}_1(\pi_T^{x,\psi^n}, \pi_T^{x,\psi^\star})\,\mathrm{d}\mu\,.$$

□

Remark 6.3.4 (Convergence along adjusted marginals)**.** *By relying on the previous result, we are also able to prove exponential convergence of the* $\mathrm{L}^1(\mu^n)$*-norms and* $\mathrm{L}^1(\nu^n)$*-norms of the difference of the gradients, along the adjusted marginals* μ^n, ν^n*. This follows from the fact that Sinkhorn's iterates may be considered as potentials of appropriate Schrödinger problems. Indeed, the decomposition given in* (2.2.19) *implies that*

- *the couple* (φ^{n+1}, ψ^n) *corresponds to a couple of Schrödinger potentials (as defined in* (6.0.1)*) associated to the Schrödinger problem with reference measure* $\mathrm{R}_{0,T}$ *and with marginals* μ *and* $\nu^n := (\mathrm{proj}_y)_\sharp \pi^{n+1,n}$*;*

- *the couple* $(\varphi^{n+1}, \psi^{n+1})$ *corresponds to a couple of Schrödinger potentials (as defined in* (6.0.1)*) associated to the Schrödinger problem with reference measure* $\mathrm{R}_{0,T}$ *and with marginals* $\mu^{n+1} := (\mathrm{proj}_x)_\sharp \pi^{n+1,n+1}$ *and* ν*.*

This simple observation, the bound (6.3.3)*, the conditional property of* π_T^{x,ψ^n}[1]*, and arguing as in Corollary 6.2.3 (this time with* $\mathfrak{p} = \pi_T^{x,\psi^\star}$ *and* $\mathfrak{q} = \pi_T^{x,\psi^n}$*) prove that*

$$\int |\nabla\varphi^n - \nabla\varphi^\star| \mathrm{d}\mu^n \leq \frac{\gamma_\infty^\nu}{T} \prod_{k=0}^{n-2} \frac{\gamma_k^\mu\, \gamma_k^\nu}{T^2} \int |\nabla\psi^0 - \nabla\psi^\star| \mathrm{d}\nu\,,$$

$$\int |\nabla\psi^n - \nabla\psi^\star| \mathrm{d}\nu^n \leq \frac{\gamma_\infty^\mu}{\gamma_{n-1}^\mu} \prod_{k=0}^{n-1} \frac{\gamma_k^\mu\, \gamma_k^\nu}{T^2} \int |\nabla\psi^0 - \nabla\psi^\star| \mathrm{d}\nu\,.$$

Our coupling approach can also be employed in order to prove pointwise convergence result. In order to establish such a result the assumptions we impose on the regularization parameter T are more stringent than the ones we need for Theorem 6.3.2. In addition we must impose an additional assumption (6.3.8) on ψ^0. As we explain below there is a natural choice for ψ^0 that guarantees that (6.3.8) holds.

Theorem 6.3.5. *Assume A6 holds,* $\psi^0 \in \mathrm{C}^1(\mathbb{R}^d)$ *and that there exist two positive constants* A, $B > 0$ *such that for any* $x \in \mathbb{R}^d$

$$|\nabla\psi^0 - \nabla\psi^\star|(x) \leq A\,|x| + B\,. \tag{6.3.8}$$

Then for any $n \in \mathbb{N}^*$ *and* $x \in \mathbb{R}^d$

$$\begin{aligned} |\nabla\varphi^n - \nabla\varphi^\star|(x) &\leq \frac{T}{\hat\gamma_{n-1}^\mu} \prod_{k=0}^{n-1} \frac{\hat\gamma_k^\mu\, \hat\gamma_k^\nu}{T^2} (A|x| + B) \\ |\nabla\psi^n - \nabla\psi^\star|(x) &\leq \prod_{k=0}^{n-1} \frac{\hat\gamma_k^\mu\, \hat\gamma_k^\nu}{T^2}\, (A|x| + B)\,, \end{aligned} \tag{6.3.9}$$

[1] Let us recall that $\pi^{n,n}$ is the optimal coupling for the Schrödinger problem with marginals μ^n and ν, that the corresponding potentials are given by φ^n, ψ^n and therefore $\pi_T^{x,\psi^n}(\mathrm{d}y)\mu^n(\mathrm{d}x) = \pi^{n,n}(\mathrm{d}x\,\mathrm{d}y)$.

where

$$\hat{\gamma}_k^\mu := \gamma_k^\mu \max\left\{ (T\,\alpha_{\varphi^\star}+1)^{-1}\,,\left(1+\frac{A}{B}\frac{1+\|\tilde{g}_\mu\|_\infty+|\nabla\varphi^\star(0)|}{\alpha_{\varphi^\star}+T^{-1}}\right)\right\},$$

and

$$\hat{\gamma}_k^\nu := \gamma_k^\nu \max\left\{ (T\,\alpha_{\psi^\star}+1)^{-1}\,,\left(1+\frac{A}{B}\frac{1+\|\tilde{g}_\nu\|_\infty+|\nabla\psi^\star(0)|}{\alpha_{\psi^\star}+T^{-1}}\right)\right\},$$

where $(\gamma_k^\mu)_{k\in\mathbb{N}}$, $(\gamma_k^\nu)_{k\in\mathbb{N}}$ *are the non-decreasing non-negative sequences defined in Remark 6.3.3,* $\alpha_{\varphi^\star}\in(\alpha_\mu-T^{-1},\alpha_\mu-T^{-1}+(\beta_\nu\,T^2)^{-1}]$ *and* $\alpha_{\psi^\star}\in(\alpha_\nu-T^{-1},\alpha_\nu-T^{-1}+(\beta_\mu\,T^2)^{-1}]$ *are given in Theorem 6.1.1. In particular, denoting the corresponding limit rates as* $\gamma_\infty^\zeta:=\lim_{k\to+\infty}\gamma_k^\zeta$ *and* $\hat{\gamma}_\infty^\zeta:=\lim_{k\to+\infty}\hat{\gamma}_k^\zeta$ *for* $\zeta\in\{\mu,\nu\}$*, if* T *is large enough, e.g., if*

$$T>\max\left\{\alpha_\mu^{-1},\,\alpha_\nu^{-1},\,\gamma_\infty^\mu+\frac{\gamma_\infty^\mu A}{\alpha_\mu B}(1+\|\tilde{g}_\mu\|_\infty+|\nabla\varphi^\star(0)|),\right.$$
$$\left.\gamma_\infty^\nu+\frac{\gamma_\infty^\nu A}{\alpha_\nu B}(1+\|\tilde{g}_\nu\|_\infty+|\nabla\psi^\star(0)|)\right\},\tag{6.3.10}$$

then $T^2>\hat{\gamma}_\infty^\mu\,\hat{\gamma}_\infty^\nu$ *and as a result for any* $\lambda\in(T^{-2}\,\hat{\gamma}_\infty^\mu\hat{\gamma}_\infty^\nu,\,1)$*, there exists* $C\geq 0$ *such that for any* $x\in\mathbb{R}^d$ *and* $n\in\mathbb{N}^*$*,*

$$|\nabla\varphi^n-\nabla\varphi^\star|(x)+|\nabla\psi^n-\nabla\psi^\star|(x)\leq C\,\lambda^n\,(A|x|+B)\,.$$

As for Theorem 6.3.2, expressions of $(\hat{\gamma}_k^\mu)_{k\in\mathbb{N}}$, $(\hat{\gamma}_k^\nu)_{k\in\mathbb{N}}$ simplify as U_μ and U_ν are strongly convex. These expressions are given in Section 6.6.

We stress that the previous theorem holds for any smooth initialisation $\psi^0\in\mathcal{C}^1(\mathbb{R}^d)$ satisfying (6.3.8). A common choice would be starting at $\psi^0=U_\nu$, which corresponds to $\varphi^0=0$. This choice agrees with the usual normalisation imposed to Sinkhorn's iterates when studying its convergence [DMG20, CL20, DdBD24, Car22]. Let us also point out that if one starts Sinkhorn's algorithm one step before with the null initialisation $\varphi^0:=0$, then at the first iteration we immediately get $\psi^0=U_\nu$. Under this choice, from (2.2.17) we deduce that

$$\psi^0-\psi^\star=-\log P_T\exp(-\varphi^\star)$$

hence

$$|\nabla\psi^0-\nabla\psi^\star|(x)=\frac{1}{T}\left|\int(y-x)\,\pi_T^{x,\varphi^\star}(\mathrm{d}y)\right|\leq\frac{|x|}{T}+\frac{1}{T}\int|y|\,\mathrm{d}\pi_T^{x,\varphi^\star}(y)\,.\tag{6.3.11}$$

The latter combined with the bound we are going to give later in (6.3.16) shows that the initialisation $\psi^0=U_\nu$ automatically satisfies the linear growth condition of Theorem 6.3.5. At the same time, integrating (6.3.11) w.r.t. ν and using

(6.0.5) (exchanging the roles between first and second marginal) yields to

$$\int |\nabla\psi^0 - \nabla\psi^\star| \, \mathrm{d}\nu \le T^{-1} \int |x-y| \; \pi_T^{x,\varphi^\star}(\mathrm{d}y)\, \nu(\mathrm{d}x) = T^{-1} \int |x-y| \, \mathrm{d}\boldsymbol{\pi}^T \tag{6.3.12}$$

which allows us to state (6.3.5) in terms of the first moments of the marginals μ and ν.

Summarising the above discussion, if we start from $\psi^0 = U_\nu$, the previous results read as

Corollary 6.3.6. *Assume A6. If we set the initial Sinkhorn's iterate equal to $\psi^0 = U_\nu$ (or equivalently if we start with $\varphi^0 = 0$), then the linear-growth condition* (6.3.8) *is satisfied with*

$$A = T^{-1}\,\frac{T\,\alpha_{\varphi^\star}+2}{T\,\alpha_{\varphi^\star}+1} \quad \text{and} \quad B = \frac{1+\|\tilde g_\mu\|_\infty + |\nabla\varphi^\star(0)|}{T\,\alpha_{\varphi^\star}+1}\,, \tag{6.3.13}$$

where $\alpha_{\varphi^\star} \in (\alpha_\mu - T^{-1}, \alpha_\mu - T^{-1} + (\beta_\nu T^2)^{-1}]$ is given in Theorem 6.1.1.

Moreover, the integrated bounds (6.3.5) *read as*

$$\int |\nabla\varphi^n - \nabla\varphi^\star|(x)\mu(dx) \le \frac{1}{\gamma^\mu_{n-1}} \prod_{k=0}^{n-1} \frac{\gamma^\mu_k\,\gamma^\nu_k}{T^2}\,(M_1(\mu)+M_1(\nu))\,,$$

$$\int |\nabla\psi^n - \nabla\psi^\star|(y)\nu(dy) \le \prod_{k=0}^{n-1} \frac{\gamma^\mu_k\,\gamma^\nu_k}{T^2}\,(M_1(\mu)+M_1(\nu))\,.$$

As a consequence of the previous corollary, starting from $\psi^0 = U_\nu$, we have exponential pointwise convergence of the gradients as soon as T is large enough, as mentioned in Theorem 6.3.5.

The proof of Theorem 6.3.5 relies on a contractive technique which is based on the linear growth condition and reads as follows.

Lemma 6.3.7. *Assume that A6 holds true. If there are positive constants $A, B > 0$ such that for any $x \in \mathbb{R}^d$, $|\nabla\psi^n - \nabla\psi^\star|(x) \le A|x| + B$, then it holds for any $x \in \mathbb{R}^d$,*

$$|\nabla\varphi^{n+1} - \nabla\varphi^\star|(x) \le T^{-1}\hat\gamma^\nu_n\,(A|x|+B)\,,$$

with

$$\hat\gamma^\nu_n := \gamma^\nu_n \,\max\left\{(T\,\alpha_{\psi^\star}+1)^{-1}\,, \left(1+\frac{A}{B}\frac{1+\|\tilde g_\nu\|_\infty+|\nabla\psi^\star(0)|}{\alpha_{\psi^\star}+T^{-1}}\right)\right\}\,,$$

and γ^ν_n is given in (6.2.11).

Similarly if for any $x \in \mathbb{R}^d$, $|\nabla\varphi^{n+1} - \nabla\varphi^\star|(x) \le A|x| + B$, then it holds for any $x \in \mathbb{R}^d$,

$$|\nabla\psi^{n+1} - \nabla\psi^\star|(x) \le T^{-1}\hat\gamma^\mu_n\,(A|x|+B)\,,$$

with

$$\hat{\gamma}^{\mu}_n := \gamma^{\mu}_n \ \max\left\{ (T\,\alpha_{\varphi^\star}+1)^{-1}\,, \left(1+\frac{A}{B}\frac{1+\|\tilde{g}_\mu\|_\infty+|\nabla\varphi^\star(0)|}{\alpha_{\varphi^\star}+T^{-1}}\right)\right\}\,,$$

and γ^{μ}_n *is given in* (6.2.11).

Proof. We will prove only the first inequality since the proof of the second one can be achieved by following the same argument. Owing to the computations performed in Proposition 6.3.1 and Corollary 6.2.3, let us consider (6.3.4) as starting point of our proof here. Therefore we have

$$|\nabla\varphi^{n+1}-\nabla\varphi^\star|(x) \leq \ T^{-1}\,\gamma^{\nu}_n \int |\nabla\psi^n-\nabla\psi^\star|\,\mathrm{d}\pi^{x,\psi^\star}_T\,,$$

which combined with our assumption yields to

$$|\nabla\varphi^{n+1}-\nabla\varphi^\star|(x) \leq T^{-1}\,\gamma^{\nu}_n\left(A\,\mathbb{E}_{\pi^{x,\psi^\star}_T}[|Y|]+B\right)\,. \tag{6.3.14}$$

Therefore our proof follows once we provide a bound on the above right-hand-side. In order to do that, let us denote by $Y^\star$ the strong solution of

$$\begin{cases}\mathrm{d}Y^\star_t = -\left(\frac{Y^\star_t-x}{2T}+\frac{1}{2}\,\nabla\psi^\star(Y^\star_t)\right)\mathrm{d}t+\mathrm{d}B_t\\ Y^\star_0\sim\pi^{x,\psi^\star}_T\,.\end{cases}$$

Then Ito formula, Corollary 6.1.3 and the boundedness of $\tilde{g}_\nu\in\tilde{\mathcal{G}}$ imply that

$$\begin{aligned}\mathrm{d}|Y^\star_t|^2 &= -T^{-1}|Y^\star_t|^2\,\mathrm{d}t+T^{-1}\langle Y^\star_t,\,x\rangle\,\mathrm{d}t-\langle Y^\star_t,\,\nabla\psi^\star(Y^\star_t)\rangle\,\mathrm{d}t+1\,\mathrm{d}t\\ &\qquad +2\,Y^\star_t\cdot\mathrm{d}B_t\\ &\leq -(\alpha_{\psi^\star}+T^{-1})|Y^\star_t|^2\,\mathrm{d}t+(1+T^{-1}|x|+\|\tilde{g}_\nu\|_\infty+|\nabla\psi^\star(0)|)|Y^\star_t|\,\mathrm{d}t\\ &\qquad +2\,Y^\star_t\cdot\mathrm{d}B_t\,,\end{aligned}$$

and therefore for any $\varepsilon\in(0,\alpha_{\psi^\star}+T^{-1})$ we have

$$\begin{aligned}\mathrm{d}|Y^\star_t|^2 \leq &-(\alpha_{\psi^\star}+T^{-1}-\varepsilon)|Y^\star_t|^2\,\mathrm{d}t\\ &+(4\varepsilon)^{-1}(1+T^{-1}|x|+\|\tilde{g}_\nu\|_\infty+|\nabla\psi^\star(0)|)^2\mathrm{d}t+2\,Y^\star_t\cdot\mathrm{d}B_t\,.\end{aligned}$$

If we consider the stopping time $\tau_M:=\inf\{t\geq 0\,:\ |Y^\star_t|>M\}$, where we set $\inf(\emptyset):=+\infty$, by integrating over time on $[0,t\wedge\tau_M]$, taking expectation and owing to the Optional Stopping Theorem we deduce that

$$\begin{aligned}\mathbb{E}[|Y^\star_{t\wedge\tau_M}|^2] \leq \mathbb{E}[|Y^\star_0|^2]-(\alpha_{\psi^\star}+T^{-1}-\varepsilon)\int_0^t\mathbb{E}[\mathbf{1}_{s\leq\tau_M}\,|Y^\star_s|^2]\,\mathrm{d}s\\ +\frac{t}{4\varepsilon}(1+T^{-1}|x|+\|\tilde{g}_\nu\|_\infty+|\nabla\psi^\star(0)|)^2\,.\end{aligned}$$

Since $\tau_M \uparrow +\infty$ almost surely as $M \uparrow +\infty$ (as a consequence of Corollary 6.1.3), in the asymptotic regime Fatou Lemma implies for any $t \geq 0$ that

$$\begin{aligned}\mathbb{E}[|Y^\star_t|^2] + (\alpha_{\psi^\star} + T^{-1} - \varepsilon) \int_0^t \mathbb{E}[|Y^\star_s|^2]\,\mathrm{d}s \\ \leq \mathbb{E}[|Y^\star_0|^2] + \frac{t}{4\varepsilon}(1 + T^{-1}|x| + \|\tilde{g}_\nu\|_\infty + |\nabla\psi^\star(0)|)^2,\end{aligned}$$

which combined with the stationarity of the process $Y^\star_s \sim \pi_T^{x,\psi^\star}$, gives

$$\mathbb{E}_{\pi_T^{x,\psi^\star}}[|Y|] \leq \mathbb{E}_{\pi_T^{x,\psi^\star}}[|Y|^2]^{1/2} \leq \frac{1 + T^{-1}|x| + \|\tilde{g}_\nu\|_\infty + |\nabla\psi^\star(0)|}{\sqrt{4\varepsilon(\alpha_{\psi^\star} + T^{-1} - \varepsilon)}}\,.$$

By minimising over $\varepsilon \in (0, \alpha_{\psi^\star} + T^{-1})$ we finally get the desired upper-bound

$$\begin{aligned}\mathbb{E}_{\pi_T^{x,\psi^\star}}[|Y|] &\leq \frac{1 + T^{-1}|x| + \|\tilde{g}_\nu\|_\infty + |\nabla\psi^\star(0)|}{\alpha_{\psi^\star} + T^{-1}} \\ &= \frac{|x|}{T\,\alpha_{\psi^\star} + 1} + \frac{1 + \|\tilde{g}_\nu\|_\infty + |\nabla\psi^\star(0)|}{\alpha_{\psi^\star} + T^{-1}}\,.\end{aligned} \tag{6.3.15}$$

By combining the above bound with (6.3.14) we finally conclude that

$$\begin{aligned}|\nabla\varphi^{n+1} - \nabla\varphi^\star|(x) \leq& T^{-1}\,\gamma_n^\nu\left(\frac{A}{T\,\alpha_{\psi^\star} + 1}\,|x| + A\frac{1 + \|\tilde{g}_\nu\|_\infty + |\nabla\psi^\star(0)|}{\alpha_{\psi^\star} + T^{-1}} + B\right) \\ \leq& T^{-1}\hat{\gamma}_n^\nu(A\,|x| + B)\end{aligned}$$

where the last step holds true because of the choice of $\hat{\gamma}_n^\nu$. This concludes the proof of the first part of the statement. The proof of the second one is similar and for this reason we omit it. Let us just mention here that the same reasoning yields to the moment bound

$$\mathbb{E}_{\pi_T^{x,\varphi^\star}}[|Y|] < \frac{|x|}{T\,\alpha_{\varphi^\star} + 1} + \frac{1 + \|\tilde{g}_\mu\|_\infty + |\nabla\varphi^\star(0)|}{\alpha_{\varphi^\star} + T^{-1}}\,. \tag{6.3.16}$$

□

Proof of Theorem 6.3.5. It is enough concatenating the bounds we have deduced in Lemma 6.3.7 and observing that at each step ψ^k and φ^k satisfy (6.3.8) with constants A_k, B_k such that the ratio $A^k/B^k \equiv A/B$. □

6.4 Exponential pointwise and entropic convergence of Sinkhorn's algorithm

It is possible to infer the convergence of Sinkhorn's iterates $(\varphi^n)_{n\in\mathbb{N}}$, $(\psi^n)_{n\in\mathbb{N}}$ from the convergence of their gradients. Since Schrödinger potentials are unique up to a trivial additive shift, let us recall that we impose the symmetric normalisation (2.2.13), that is we suppose it holds

$$\int \varphi^\star \,\mathrm{d}\mu + \mathrm{Ent}(\mu) = \int \psi^\star \,\mathrm{d}\nu + \mathrm{Ent}(\nu) = \frac{1}{2}\Big(\mathrm{Ent}(\mu) + \mathrm{Ent}(\nu) - \mathscr{H}(\boldsymbol{\pi}^T|\mathrm{R}_{0,T})\Big).$$

This normalisation has already been used while showing convergence of the (rescaled) Schrödinger potentials and their gradients to Kantorovich potentials and the Brenier map respectively, see Chapter 3. In what concerns Sinkhorn's iterates, we work with the normalisation already considered in (5.1.8), *i.e.*, we consider the shifted iterates

$$\varphi^{\diamond n} = \varphi^n - \left(\int \varphi^n \mathrm{d}\mu - \int \varphi^\star \mathrm{d}\mu\right), \qquad \psi^{\diamond n} = \psi^n - \left(\int \psi^n \mathrm{d}\nu - \int \psi^\star \mathrm{d}\nu\right), \tag{6.4.1}$$

This choice guarantees that

$$\int \varphi^{\diamond n}\mathrm{d}\mu + \mathrm{Ent}(\mu) = \int \varphi^\star \mathrm{d}\mu + \mathrm{Ent}(\mu) = \int \psi^\star \mathrm{d}\nu + \mathrm{Ent}(\nu) = \int \psi^{\diamond n}\mathrm{d}\nu + \mathrm{Ent}(\nu).$$

Theorem 6.4.1. *Assume that **A**6 and* (6.3.8) *hold. Then for any $n \geq 1$ and $x \in \mathbb{R}^d$ it holds*

$$|\varphi^{\diamond n} - \varphi^\star|(x) \leq \frac{T}{\hat{\gamma}^\mu_{n-1}} \prod_{k=0}^{n-1} \frac{\hat{\gamma}^\mu_k\, \hat{\gamma}^\nu_k}{T^2} \Big[A\,|x|^2 + (A\, M_1(\mu) + B)\,|x| + B\, M_1(\mu) + 2\,A\, M_2(\mu)\Big],$$

$$|\psi^{\diamond n} - \psi^\star|(x) \leq \prod_{k=0}^{n-1} \frac{\hat{\gamma}^\mu_k\, \hat{\gamma}^\nu_k}{T^2} \Big[A\,|x|^2 + (A\, M_1(\nu) + B)\,|x| + B\, M_1(\nu) + 2\,A\, M_2(\nu)\Big].$$

Hence, for T large enough (e.g., (6.3.10)*), the pointwise exponential convergence of Sinkhorn's iterates holds. Finally, if the initial iteration is set equal to $\psi^0 = U_\nu$ (i.e., $\varphi^0 = 0$), then the above bounds hold true with A and B given at* (6.3.13).

Let us mention here that a straightforward adaptation of the proof of Theorem 6.4.1 implies that this result also holds true under a pointwise normalisation (e.g., $\psi^\star(0) = \psi^{\diamond n}(0) = U_\nu(0)$) or for the symmetric zero-mean option considered in [DMG20, CL20, DdBD24, Car22].

Proof of Theorem 6.4.1. Owing to the normalisations (2.2.13), (6.4.1) and to Theorem 6.3.5, we immediately deduce that

$$
\begin{aligned}
|\varphi^{\diamond n} - \varphi^\star|(x) =& \left|\varphi^{\diamond n}(x) - \int \varphi^{\diamond n} \mathrm{d}\mu - \varphi^\star(x) + \int \varphi^\star \mathrm{d}\mu \right| \\
=& \left| \int (\varphi^n - \varphi^\star)(x) - (\varphi^n - \varphi^\star)(y) \, \mathrm{d}\mu(y) \right| \\
\leq& \int \left| (\varphi^n - \varphi^\star)(x) - (\varphi^n - \varphi^\star)(y) \right| \mathrm{d}\mu(y) \\
\leq& \int \int_0^1 |\nabla(\varphi^n - \varphi^\star)|(y + t(x-y)) \, |x-y| \, \mathrm{d}t \, \mathrm{d}\mu(y) \\
\overset{(6.3.9)}{\leq}& \frac{T}{\hat{\gamma}^\mu_{n-1}} \prod_{k=0}^{n-1} \frac{\hat{\gamma}^\mu_k \, \hat{\gamma}^\nu_k}{T^2} \int \int_0^1 (A|y + t(x-y)| + B)|x-y| \, \mathrm{d}t \, \mathrm{d}\mu(y) \\
\leq& \frac{T}{\hat{\gamma}^\mu_{n-1}} \prod_{k=0}^{n-1} \frac{\hat{\gamma}^\mu_k \, \hat{\gamma}^\nu_k}{T^2} \Big[A \, |x|^2 + (A \, M_1(\mu) + B) \, |x| + B \, M_1(\mu) \\
& \qquad + 2 \, A \, M_2(\mu) \Big] \, .
\end{aligned}
$$

The second pointwise bound can be proven in the same manner. □

As a consequence of Theorem 6.4.1 we may also deduce the convergence of the L^1-norms along the adjusted marginals and along the real marginals.

Corollary 6.4.2. *Assume* ***A6*** *and* (6.3.8) *for some positive constants* $A, B > 0$*. Then for any* $n \geq 1$ *it holds*

$$
\begin{aligned}
\|\varphi^{\diamond n} - \varphi^\star\|_{\mathrm{L}^1(\mu)} \leq& \frac{T}{\hat{\gamma}^\mu_{n-1}} \prod_{k=0}^{n-1} \frac{\hat{\gamma}^\mu_k \, \hat{\gamma}^\nu_k}{T^2} \Big[3 \, A \, M_2(\mu) + (A \, M_1(\mu) + B) \, M_1(\mu) \\
& \qquad + B \, M_1(\mu) \Big] \, , \\
\|\psi^{\diamond n} - \psi^\star\|_{\mathrm{L}^1(\nu)} \leq& \prod_{k=0}^{n-1} \frac{\hat{\gamma}^\mu_k \, \hat{\gamma}^\nu_k}{T^2} \Big[3 \, A \, M_2(\nu) + (A \, M_1(\nu) + B) \, M_1(\nu) \\
& \qquad + B \, M_1(\nu) \Big] ,
\end{aligned}
\tag{6.4.2}
$$

and

$$
\begin{aligned}
\|\varphi^{\diamond n} - \varphi^\star\|_{\mathrm{L}^1(\mu^n)} \leq& \; C(A, B, \mu) \frac{T}{\hat{\gamma}^\mu_{n-1}} \prod_{k=0}^{n-1} \frac{\hat{\gamma}^\mu_k \, \hat{\gamma}^\nu_k}{T^2} \, , \\
\|\psi^{\diamond n} - \psi^\star\|_{\mathrm{L}^1(\nu^n)} \leq& \; C(A, B, \nu) \prod_{k=0}^{n-1} \frac{\hat{\gamma}^\mu_k \, \hat{\gamma}^\nu_k}{T^2} ,
\end{aligned}
\tag{6.4.3}
$$

where

$$C(A,B,\mu) := \Big[3\,A\,M_2(\mu) + (A\,M_1(\mu)+B)\,M_1(\mu) + B\,M_1(\mu)\Big]$$
$$+A\,C_2(\mu)\left(\sqrt{\mathcal{H}(\mu^1|\mu)} + \frac{\mathcal{H}(\mu^1|\mu)}{2}\right) + (A\,M_1(\mu)+B)\,C_1(\mu)\,\sqrt{\mathcal{H}(\mu^1|\mu)} \tag{6.4.4}$$

and

$$C(A,B,\nu) := \Big[3\,A\,M_2(\nu) + (A\,M_1(\nu)+B)\,M_1(\nu) + B\,M_1(\nu)\Big]$$
$$+A\,C_2(\nu)\left(\sqrt{\mathcal{H}(\nu^0|\nu)} + \frac{\mathcal{H}(\nu^0|\nu)}{2}\right) + (A\,M_1(\nu)+B)\,C_1(\nu)\,\sqrt{\mathcal{H}(\nu^0|\nu)}\,,$$

with $C_1(\zeta)$ and $C_2(\zeta)$ are positive constants defined below in Lemma 6.4.3 for $\zeta \in \{\mu,\nu\}$.

Proof. The proof of the integrated bounds along the marginals (6.4.2) is a straightforward consequence of the pointwise convergence, whereas the bounds along the adjusted marginals follow from the weighted Csiszár-Kullback-Pinsker inequalities [BV05, Theorem 2.1] which imply for any $\zeta \in \{\mu,\,\nu\}$ that

$$M_1(\zeta^n) \le M_1(\zeta) + C_1(\zeta)\,\sqrt{\mathcal{H}(\zeta^n|\zeta)}\,,$$
$$M_2(\zeta^n) \le M_2(\zeta) + C_2(\zeta)\left(\sqrt{\mathcal{H}(\zeta^n|\zeta)} + \frac{\mathcal{H}(\zeta^n|\zeta)}{2}\right)$$

where $C_1(\zeta)$, $C_2(\zeta)$ are positive constants (independent from $n \in \mathbb{N}$). For sake of clarity we postponed the proof of the above moment bounds to Lemma 6.4.3, below. The proof of (6.4.3) then follows from the fact that the two sequences $(\mathscr{H}(\mu^n|\mu))_{n\in\mathbb{N}}$ and $(\mathscr{H}(\nu^n|\nu))_{n\in\mathbb{N}}$ are monotone decreasing along Sinkhorn's algorithm [Nut21, Proposition 6.10]. □

In the next lemma we show how the weighted Csiszár-Kullback-Pinsker inequalities imply the above moments inequalities along the adjusted marginals.

Lemma 6.4.3. *Assume A6 and let $\zeta \in \{\mu,\nu\}$. Then for any probability measure $\mathfrak{p} \in \mathcal{P}(\mathbb{R}^d)$ such that $\mathcal{H}(\mathfrak{p}|\zeta) < +\infty$ it holds*

$$M_1(\mathfrak{p}) \le M_1(\zeta) + C_1(\zeta)\,\sqrt{\mathcal{H}(\mathfrak{p}|\zeta)}\,,$$
$$M_2(\mathfrak{p}) \le M_2(\zeta) + C_2(\zeta)\left(\sqrt{\mathcal{H}(\mathfrak{p}|\zeta)} + \frac{\mathcal{H}(\mathfrak{p}|\zeta)}{2}\right),$$

where $C_1(\zeta)$, $C_2(\zeta)$ are positive constants (independent of $\mathfrak{p}$) defined as

$$C_1(\zeta) := \inf_{\sigma_\zeta \in (0,\alpha_\zeta/2)} \left(\frac{2}{\sigma_\zeta} + \frac{2}{\sigma_\zeta} \log \int e^{\sigma_\zeta |x|^2} \, \mathrm{d}\zeta \right)^{1/2},$$

$$C_2(\zeta) := \inf_{\sigma_\zeta \in (0,\alpha_\zeta/2)} \left(\frac{3}{\sigma_\zeta} + \frac{2}{\sigma_\zeta} \int e^{\sigma_\zeta |x|^2} \mathrm{d}\zeta \right).$$

Proof. For any $\sigma_\nu \in (0, \alpha_\nu/2)$, from the weighted Csiszár-Kullback-Pinsker inequalities [BV05, Theorem 2.1] applied to the measurable functions $F_1(x) = \sigma_\nu^{1/2}|x|$ and $F_2(x) = \sigma_\nu |x|^2/2$ we immediately deduce

$$\begin{aligned}
M_1(\mathfrak{p}) &= M_1(\nu) + \sigma_\nu^{-1/2} \int \sigma_\nu^{1/2} \, |x| \, \mathrm{d}(\mathfrak{p} - \nu) \\
&\leq M_1(\nu) + \sqrt{\mathcal{H}(\mathfrak{p}|\nu)} \left(\frac{2}{\sigma_\nu} + \frac{2}{\sigma_\nu} \log \int e^{\sigma_\nu |x|^2} \, \mathrm{d}\nu \right)^{1/2} \\
M_2(\mathfrak{p}) &= M_2(\nu) + \frac{2}{\sigma_\nu} \int \frac{\sigma_\nu}{2} \, |x|^2 \, \mathrm{d}(\mathfrak{p} - \nu) \\
&\leq M_2(\nu) + \left(\sqrt{\mathcal{H}(\mathfrak{p}|\nu)} + \frac{\mathcal{H}(\mathfrak{p}|\nu)}{2} \right) \left(\frac{3}{\sigma_\nu} + \frac{2}{\sigma_\nu} \int e^{\sigma_\nu |x|^2} \mathrm{d}\nu \right)
\end{aligned}$$

which are finite by Lemma 6.A.1 and **A**6. Minimising over $\sigma_\nu \in (0, \alpha_\nu/2)$ concludes the proof of the first claim.

Lastly, the moment bounds corresponding to the choice of reference μ, can be proven in the same way. □

6.4.1 Exponential entropic convergence of Sinkhorn's plans

Finally, let us conclude the section with the exponential entropic convergence of Sinkhorn's algorithm on the primal side, *i.e.*, for Sinkhorn's plans $(\pi^{n,n})_{n\in\mathbb{N}}$ and $(\pi^{n+1,n})_{n\in\mathbb{N}}$ defined in (2.2.19) and for the adjusted marginals $(\mu^n)_{n\in\mathbb{N}}$ and $(\nu^n)_{n\in\mathbb{N}}$, using the symmetrised version of the relative entropy $\mathscr{H}^{\mathrm{sym}}$ (as defined at (4.0.1)). As observed in [CCGT23, GN22], measuring distances between plans with this divergence leads to tractable expressions.

Theorem 6.4.4 (Exponential convergence of Sinkhorn on the primal side)**.** *Assume that **A**6 and* (6.3.8) *hold. Then, for any $n \geq 1$ it holds*

$$\mathscr{H}^{\mathrm{sym}}(\pi^{n,n}, \boldsymbol{\pi}^T) \leq D(A, B, \mu) \frac{T}{\hat{\gamma}^\mu_{n-1}} \prod_{k=0}^{n-1} \frac{\hat{\gamma}^\mu_k \hat{\gamma}^\nu_k}{T^2},$$

$$\mathscr{H}^{\mathrm{sym}}(\pi^{n+1,n}, \boldsymbol{\pi}^T) \leq D(A, B, \nu) \prod_{k=0}^{n-1} \frac{\hat{\gamma}^\mu_k \hat{\gamma}^\nu_k}{T^2},$$

and

$$\mathscr{H}^{\mathrm{sym}}(\pi^{n+1,n},\pi^{n,n}) = \mathscr{H}^{\mathrm{sym}}(\mu^n,\mu) \leq D(A,B,\mu)\frac{T}{\hat{\gamma}^{\mu}_{n-1}}\prod_{k=0}^{n-1}\frac{\hat{\gamma}^{\mu}_k\hat{\gamma}^{\nu}_k}{T^2}\,,$$

$$\mathscr{H}^{\mathrm{sym}}(\pi^{n+1,n+1},\pi^{n+1,n}) = \mathscr{H}^{\mathrm{sym}}(\nu^n,\mu) \leq D(A,B,\nu)\prod_{k=0}^{n-1}\frac{\hat{\gamma}^{\mu}_k\hat{\gamma}^{\nu}_k}{T^2}\,,$$

where the multiplicative constants expressions $D(A,B,\mu)$, $D(A,B,\nu)$ are explicitly given in (6.4.5) *and* (6.4.6).

As a consequence, for T large enough (e.g. (6.3.10)*), entropic exponential convergence of Sinkhorn's plans and adjusted marginals holds. Finally, if the initial iteration is set equal to $\psi^0 = U_\nu$ (i.e., $\varphi^0 = 0$), then the above bounds hold true with A and B given at* (6.3.13).

In Remark 6.4.5 we show how the multiplicative constant $D(A,B,\cdot)$ can be further improved. The benefit of considering symmetric relative entropies in Theorem 6.4.4 is twofold: not only it allows us to bound these relative entropies in terms of $\varphi^n-\varphi^\star$ and $\psi^n-\psi^\star$, but also allows us to translate it in terms of $\varphi^{\diamond n}-\varphi^\star$ and $\psi^{\diamond n}-\psi^\star$ and therefore apply the results of Theorem 6.4.1.

Proof of Theorem 6.4.4. Let us preliminary point out that as a first consequence of Corollary 6.4.2, for any $n\geq 1$ it holds

$$\varphi^{\diamond n}-\varphi^\star \in \mathrm{L}^1(\mu)\cap\mathrm{L}^1(\mu^n) \quad\text{and}\quad \psi^{\diamond n}-\psi^\star \in \mathrm{L}^1(\nu)\cap\mathrm{L}^1(\nu^n)\,,$$

which will guarantee that the following integrals (and corresponding summations) are all well-defined.

Now, (6.0.1) and (2.2.19) imply that

$$\log\frac{\mathrm{d}\boldsymbol{\pi}^T}{\mathrm{d}\pi^{n,n}}(x,y) = \varphi^n(x)-\varphi^\star(x)+\psi^n(y)-\psi^\star(y)\,,$$

and hence the symmetric relative entropies can be rewritten as

$$\begin{aligned}
&\mathscr{H}^{\mathrm{sym}}(\pi^{n,n},\boldsymbol{\pi}^T) = \mathscr{H}(\pi^{n,n}|\boldsymbol{\pi}^T)+\mathscr{H}(\boldsymbol{\pi}^T|\pi^{n,n})\\
&=\int(\varphi^n-\varphi^\star)\oplus(\psi^n-\psi^\star)\,\mathrm{d}\boldsymbol{\pi}^T-\int(\varphi^n-\varphi^\star)\oplus(\psi^n-\psi^\star)\,\mathrm{d}\pi^{n,n}\\
&=\int(\varphi^{\diamond n}-\varphi^\star)\oplus(\psi^{\diamond n}-\psi^\star)\,\mathrm{d}\boldsymbol{\pi}^T-\int(\varphi^{\diamond n}-\varphi^\star)\oplus(\psi^{\diamond n}-\psi^\star)\,\mathrm{d}\pi^{n,n}\\
&=\int(\varphi^{\diamond n}-\varphi^\star)\,\mathrm{d}\mu^n-\int(\varphi^{\diamond n}-\varphi^\star)\,\mathrm{d}\mu\,.
\end{aligned}$$

By combining the above with Corollary 6.4.2, we then deduce that

$$\mathscr{H}(\pi^{n,n},\boldsymbol{\pi}^T) \leq \frac{D(A,B,\mu)}{T^{-1}\hat{\gamma}^{\mu}_{n-1}}\prod_{k=0}^{n-1}\frac{\hat{\gamma}^{\mu}_k\,\hat{\gamma}^{\nu}_k}{T^2}\,,$$

with

$$\begin{aligned}
D(A,B,\mu) &:= C(A,B,\mu) + 3\,A\ M_2(\mu) + (A\ M_1(\mu)+B)\,M_1(\mu) + B\ M_1(\mu)\\
&= 2\Big[3\,A\ M_2(\mu) + (A\ M_1(\mu)+B)\,M_1(\mu) + B\ M_1(\mu)\Big]\\
&\quad + A\,C_2(\mu)\left(\sqrt{\mathcal{H}(\mu^1|\mu)} + \frac{\mathcal{H}(\mu^1|\mu)}{2}\right)\\
&\quad + (A\,M_1(\mu)+B)\,C_1(\mu)\,\sqrt{\mathcal{H}(\mu^1|\mu)},
\end{aligned} \tag{6.4.5}$$

with $C_1(\mu)$, $C_2(\mu)$ being the constants introduced in Lemma 6.4.3.

Similarly (6.0.1) and (2.2.19) imply

$$\log \frac{\mathrm{d}\boldsymbol{\pi}^T}{\mathrm{d}\pi^{n+1,n}}(x,y) = \varphi^{n+1}(x) - \varphi^\star(x) + \psi^n(y) - \psi^\star(y)\,,$$

hence

$$\mathscr{H}(\pi^{n+1,n}|\boldsymbol{\pi}^T) + \mathscr{H}(\boldsymbol{\pi}^T|\pi^{n+1,n}) = \int (\psi^{\diamond n} - \psi^\star)\,\mathrm{d}\nu^n - \int (\psi^{\diamond n} - \psi^\star)\,\mathrm{d}\nu\,.$$

The latter combined with Corollary 6.4.2 yields to

$$\begin{aligned}
\mathscr{H}^{\mathrm{sym}}(\pi^{n+1,n},\boldsymbol{\pi}^T) &= \mathscr{H}(\pi^{n+1,n}|\boldsymbol{\pi}^T) + \mathscr{H}(\boldsymbol{\pi}^T|\pi^{n+1,n})\\
&\le D(A,B,\nu)\,\prod_{k=0}^{n-1}\frac{\hat\gamma_k^\mu\ \hat\gamma_k^\nu}{T^2},
\end{aligned}$$

with

$$\begin{aligned}
D(A,B,\nu) &:= C(A,B,\nu) + 3\,A\ M_2(\nu) + (A\ M_1(\nu)+B)\,M_1(\nu) + B\ M_1(\nu)\\
&= 2\Big[3\,A\ M_2(\nu) + (A\ M_1(\nu)+B)\,M_1(\nu) + B\ M_1(\nu)\Big]\\
&\quad + A\,C_2(\nu)\left(\sqrt{\mathcal{H}(\nu^0|\nu)} + \frac{\mathcal{H}(\nu^0|\nu)}{2}\right)\\
&\quad + (A\,M_1(\nu)+B)\,C_1(\nu)\,\sqrt{\mathcal{H}(\nu^0|\nu)}.
\end{aligned} \tag{6.4.6}$$

Finally, the proof of the last claims runs exactly as in the last part of the proof of Theorem 5.2.7 and for this reason we omit it here. □

Remark 6.4.5. *Let us remark here that from Theorem 6.4.4 we can consider a sharper multiplicative constant $C_S(A,B,\mu)$ in Corollary 6.4.2. Indeed, instead of relying on the monotonicity of relative entropies along Sinkhorn's algorithm, in* (6.4.4) *we could define $C_S(A,B,\mu)$ as*

$$C_S(A,B,\mu) := \Big[3\,A\ M_2(\mu) + (A\ M_1(\mu)+B)\,M_1(\mu) + B\ M_1(\mu)\Big] + \varepsilon_\mu(n)\,,$$

where

$$\varepsilon_\mu(n) := A\, C_2(\mu)\left(\sqrt{\mathcal{H}(\mu^n|\mu)} + \frac{\mathcal{H}(\mu^n|\mu)}{2}\right) + (A\, M_1(\mu) + B)\, C_1(\mu)\, \sqrt{\mathcal{H}(\mu^n|\mu)}$$

is a positive constant exponentially small as $n \uparrow +\infty$ thanks to Theorem 6.4.4. Then, in Theorem 6.4.4 instead of (6.4.5)*, we can consider the sharper multiplicative constant*

$$D_S(A,B,\mu) := 2\left[3\,A\;M_2(\mu) + (A\;M_1(\mu) + B)\,M_1(\mu) + B\;M_1(\mu)\right] + \varepsilon_\mu(n)\;.$$

The same reasoning applies for $C_S(A,B,\nu)$ and $D_S(A,B,\nu)$.

6.5 Exponential convergence of the Hessians along Sinkhorn's algorithm

In this section we show that the pointwise convergence of the iterates' gradients $(\nabla\varphi^n)_{n\in\mathbb{N}}$ and $(\nabla\psi^n)_{n\in\mathbb{N}}$ implies the pointwise convergence of the corresponding Hessian matrices $(\nabla^2\varphi^n)_{n\in\mathbb{N}}$ and $(\nabla^2\psi^n)_{n\in\mathbb{N}}$, with the exact same exponential convergence rate. We will measure this convergence through the Frobenius norm, that is defined for any matrix $\mathbf{A} \in \mathbb{R}^{d\times d}$

$$\|\mathbf{A}\|_{\mathrm{F}} := \sqrt{\mathrm{Tr}(\mathbf{A}\mathbf{A}^\mathsf{T})}\;.$$

The proof of the convergence for the Hessians follows a scheme similar to the one for the gradients, replacing the representation formula for the derivative of Sinkhorn's potentials with its second order counterpart, that is

$$\nabla^2 \log P_T \exp(-h)(x) = -T^{-1}\,\mathrm{Id} + T^{-2}\,\mathrm{Cov}(\pi_T^{x,h})\;, \tag{6.5.1}$$

which is proven as a part of Proposition 6.A.2.

For notations sake for any couple of vectors $v,\ w \in \mathbb{R}^d$ we will denote by $v \otimes w = v\,w^\mathsf{T}$ the matrix given by their tensor product. Then, for any $v,\ w \in \mathbb{R}^d$ the Frobenius norm of their product reads as

$$\|v \otimes w\|_{\mathrm{F}} = |v|\,|w|\;. \tag{6.5.2}$$

Through this section we will always assume the validity of the hypothesis of Theorem 6.3.5, *i.e.*, that both **A**6, and (6.3.8) are met.

Lemma 6.5.1. *There exist two positive constants C_1, $C_2 > 0$ independent of x and T, such that for any coupling $\pi \in \Pi(\pi_T^{x,\psi^\star},\ \pi_T^{x,\psi^n})$ it holds*

$$\|\nabla^2\varphi^{n+1} - \nabla^2\varphi^\star\|_{\mathrm{F}}(x) \le T^{-2}\,\mathbb{E}_{(Y,Z)\sim\pi}\left[|Y-Z|\,(|Y| + |Z| + C_1\,T^{-1}\,|x| + C_2)\right]. \tag{6.5.3}$$

Proof. As a byproduct of (2.2.18) and (6.5.1) we immediately have for any $x \in \mathbb{R}^d$ and for any coupling $\pi \in \Pi(\pi_T^{x,\psi^\star}, \pi_T^{x,\psi^n})$

$$\begin{aligned} T^2\,(\nabla^2\varphi^{n+1} - \nabla^2\varphi^\star)(x) =\ & \mathrm{Cov}(\pi_T^{x,\psi^\star}) - \mathrm{Cov}(\pi_T^{x,\psi^n}) \\ =\ & \mathbb{E}_{(Y,Z)\sim\pi}\left[Y^{\otimes 2} - Z^{\otimes 2}\right] + \mathbb{E}_{Z\sim\pi_T^{x,\psi^n}}[Z]^{\otimes 2} \\ & - \mathbb{E}_{Y\sim\pi_T^{x,\psi^\star}}[Y]^{\otimes 2} \\ =\ & \mathbb{E}_{(Y,Z)\sim\pi}\left[(Y-Z)\otimes Y\right] + \mathbb{E}_{(Y,Z)\sim\pi}\left[Z\otimes(Y-Z)\right] \\ & - \mathbb{E}_{(Y,Z)\sim\pi}\left[Y-Z\right]\mathbb{E}_{Z\sim\pi_T^{x,\psi^n}}[Z] \\ & - \mathbb{E}_{Y\sim\pi_T^{x,\psi^\star}}[Y]\,\mathbb{E}_{(Y,Z)\sim\pi}\left[Y-Z\right]. \end{aligned}$$

By applying the Frobenius norm, and recalling (6.5.2), we then deduce that

$$\begin{aligned} & T^2\,\|\nabla^2\varphi^{n+1} - \nabla^2\varphi^\star\|_{\mathrm{F}}(x) \\ & \quad \le \mathbb{E}_{(Y,Z)\sim\pi}\left[|Y-Z|\,\left(|Y| + |Z| + \mathbb{E}_{Z\sim\pi_T^{x,\psi^n}}[|Z|] + \mathbb{E}_{Y\sim\pi_T^{x,\psi^\star}}[|Y|]\right)\right] \end{aligned}$$

In order to bound the last two expected values in the right hand side we will proceed as in the proof of Lemma 6.3.7. Particularly, (6.3.15) already proves that

$$\mathbb{E}_{Y\sim\pi_T^{x,\psi^\star}}[|Y|] < \frac{|x|}{T\,\alpha_{\psi^\star}+1} + \frac{1+\|\tilde g_\nu\|_\infty + |\nabla\psi^\star(0)|}{\alpha_{\psi^\star}+T^{-1}}\,.$$

By reasoning in the same way we can prove that

$$\mathbb{E}_{Z\sim\pi_T^{x,\psi^n}}[|Z|] < \frac{|x|}{T\,\alpha_{\nu,n}+1} + \frac{1+\|\tilde g_\nu\|_\infty + |\nabla\psi^n(0)|}{\alpha_{\psi^n}+T^{-1}}\,,$$

Particularly, the pointwise convergence of the gradients of Theorem 6.3.5 and the convergences $\alpha_{\mu,n}\uparrow\alpha_{\varphi^\star}$, $\alpha_{\nu,n}\uparrow\alpha_{\psi^\star}$ stated in Theorem 6.1.1 yield to the uniform bound

$$\mathbb{E}_{Y\sim\pi_T^{x,\psi^\star}}[|Y|]\ \vee\ \sup_{n\in\mathbb{N}}\mathbb{E}_{Z\sim\pi_T^{x,\psi^n}}[|Z|]\ \le C_1\,T^{-1}\,|x| + C_2\,,$$

for some positive constants C_1, $C_2 > 0$ independent of x and T. This concludes our proof. □

Let us recall here (cf. Corollary 6.1.3) that $V(y) = 1+|y|^2$ is a Lyapunov function for (6.0.4) satisfying a geometric drift condition, *i.e.*, there are constants A_μ, $A_\nu > 0$ and B_μ, B_ν, independent of n (but depending on x and T), such that

$$\begin{aligned} \mathcal{L}_{\psi^\star}V(y)\ \vee\ \mathcal{L}_{\psi^n}V(y)\ &\le B_\nu - A_\nu\,V(y)\,, \\ \mathcal{L}_{\varphi^\star}V(y)\ \vee\ \mathcal{L}_{\varphi^n}V(y)\ &\le B_\mu - A_\mu\,V(y)\,, \end{aligned} \tag{6.5.4}$$

where $\mathcal{L}_h := \Delta/2 - \frac{1}{2}\langle T^{-1}(y-x) + \nabla h(y), \nabla\rangle$ is the generator associated to (6.0.4). The possibility of choosing parameters A_μ, A_ν, B_μ, B_ν independently from $n \in \mathbb{N}$ follows from the pointwise convergence of the gradients of Theorem 6.3.5 and the convergences $\alpha_{\mu,n} \uparrow \alpha_{\varphi^\star}$, $\alpha_{\nu,n} \uparrow \alpha_{\psi^\star}$ stated in Theorem 6.1.1.

We are now ready to state and prove the convergence of the Hessians. Our proof relies on the construction of a concave function (similar to the one considered in Section 6.2). For exposition's purposes we have partially postponed these computations to Section 6.5.1.

Theorem 6.5.2. *Assume that **A**6 and* (6.3.8) *hold. Then, for any $n \geq 1$ it holds*

$$\|\nabla^2\varphi^{n+1} - \nabla^2\varphi^\star\|_{\mathrm{F}}(x) \leq C(x,T,\nu,A,B) \prod_{k=0}^{n-1} \frac{\hat{\gamma}_k^\mu \hat{\gamma}_k^\nu}{T^2},$$
$$\|\nabla^2\psi^{n+1} - \nabla^2\psi^\star\|_{\mathrm{F}}(x) \leq C(x,T,\mu,A,B) \frac{T}{\hat{\gamma}_n^\mu} \prod_{k=0}^{n} \frac{\hat{\gamma}_k^\mu \hat{\gamma}_k^\nu}{T^2},$$

where $C(x,T,\nu,A,B)$ and $C(x,T,\mu,A,B)$ are given in (6.5.15). *As a consequence, for T large enough (e.g.,* (6.3.10)*), pointwise exponential convergence of Sinkhorn's Hessians holds in Frobenius norm. Finally, if the initial potential is set equal to $\psi^0 = U_\nu$ (i.e., $\varphi^0 = 0$), then the above bounds hold true with A and B given at* (6.3.13).

Proof. For sake of notations let us introduce the constant

$$C_x := \max\{1,\, C_1\, T^{-1}\,|x| + C_2\}\,,$$

with C_1, $C_2 > 0$ as in Lemma 6.5.1. We will proceed as in Corollary 6.2.3, this time considering a distorted Wasserstein semi-distance

$$\mathbf{W}_{f_x^\nu}(\cdot,\cdot) := \inf_{\pi\in\Pi(\cdot,\cdot)} \mathbb{E}_{(Y,Z)\sim\pi}\Big[f_x^\nu(|Y-Z|)\,(1+\varepsilon V(Y)+\varepsilon V(Z))\Big]\,. \tag{6.5.5}$$

In the above definition $V(y) := 1 + |y|^2$ and we consider the bounded concave function f_x^ν and the parameter $\varepsilon \in (0,1)$ built in Section 6.5.1 below, by following [EGZ19b, Theorem 2.2]. Let us just state here that there is no dependence from the index $n \in \mathbb{N}$, $\varepsilon \in (0,1)$ is taken small enough such that (6.5.18) will hold, $\mathbf{W}_{f_x^\nu}$ contracts along the semigroup $(P_t^n)_{t\geq 0}$ associated to the SDE

$$\mathrm{d}Y_t^n = -\left(\frac{Y_t^n - x}{2T} + \frac{1}{2}\nabla\psi^n(Y_t^n)\right)\mathrm{d}t + \mathrm{d}B_t\,,$$

and there exists a rate $\lambda > 0$ (independent from $n \in \mathbb{N}$, explicitly given at (6.5.27)) such that for any μ_1, $\mu_2 \in \mathcal{P}(\mathbb{R}^d)$

$$\mathbf{W}_{f_x^\nu}(\mu_1\, P_t^n,\, \mu_2\, P_t^n) \leq e^{-\lambda\, t}\,\mathbf{W}_{f_x^\nu}(\mu_1,\, \mu_2)\,. \tag{6.5.6}$$

Moreover, $\mathbf{W}_{f_x^\nu}$ satisfies a weak triangle inequality, *i.e.*, for any μ_1, μ_2, $\mu_3 \in \mathcal{P}(\mathbb{R}^d)$

$$\mathbf{W}_{f_x^\nu}(\mu_1,\mu_2) \leq C_\Delta(\mathbf{W}_{f_x^\nu}(\mu_1,\mu_3) + \mathbf{W}_{f_x^\nu}(\mu_3,\mu_2))\,, \tag{6.5.7}$$

with $C_\Delta = \max\{3,\, 2+2\,R_2^2\}$, with the radius $R_2>0$ given in (6.5.17) such that

$$\begin{aligned} &\frac{C_I}{2}\,r \le f_x^\nu(r) \le r \quad \text{and} \quad \frac{1}{2} \le (f_x^\nu)'(r) \le 1 \quad \forall\, r < R_2\,, \\ &f_x^\nu(r) = f_x^\nu(R_2) \quad \text{and hence} \quad (f_x^\nu)'(r) = 0 \quad \forall\, r > R_2\,, \end{aligned} \tag{6.5.8}$$

for a positive constant $C_I \in (0,1)$ given in (6.5.19). This particularly implies that for any coupling $(Y,Z)\sim \pi \in \Pi(\pi_T^{x,\psi^\star},\, \pi_T^{x,\psi^n})$ we have

$$\begin{aligned} T^{-2}|Y-Z|\,(|Y|+|Z|+C_x) &\le \frac{C_x}{T^2\,\varepsilon^{1/2}}|Y-Z|\,(1+\varepsilon^{1/2}\,|Y|+\varepsilon^{1/2}\,|Z|) \\ &\le \begin{cases} \frac{2}{C_I}\,\frac{C_x}{T^2\,\varepsilon^{1/2}}\,f_x^\nu(|Y-Z|)\,(1+\varepsilon^{1/2}\,|Y|+\varepsilon^{1/2}\,|Z|) & \text{if } |Y-Z|\le R_2\,, \\ \frac{C_x}{T^2\,\varepsilon}\,f_x^\nu(R_2)^{-1}\,f_x^\nu(R_2)\,(1+\varepsilon^{1/2}\,|Y|+\varepsilon^{1/2}\,|Z|)^2 & \text{if } |Y-Z|>R_2 \end{cases} \\ &\le \frac{10}{3}\,\frac{C_x}{T^2\,\varepsilon^{1/2}}\left(\frac{2}{C_I}\vee\frac{f_x^\nu(R_2)^{-1}}{\varepsilon^{1/2}}\right)f_x^\nu(|Y-Z|)\,(1+\varepsilon\,V(Y)+\varepsilon\,V(Z))\,. \end{aligned}$$

Using (6.5.3) and by minimising the above over $\pi\in\Pi(\pi_T^{x,\psi^\star},\,\pi_T^{x,\psi^n})$, we deduce then that

$$\|\nabla^2\varphi^{n+1}-\nabla^2\varphi^\star\|_{\mathrm{F}}(x) \le \frac{10}{3}\,\frac{C_x}{T^2\,\varepsilon^{1/2}}\left(\frac{2}{C_I}\vee\frac{f_x^\nu(R_2)^{-1}}{\varepsilon^{1/2}}\right)\mathbf{W}_{f_x^\nu}(\pi_T^{x,\psi^\star},\,\pi_T^{x,\psi^n})\,. \tag{6.5.9}$$

Now, recall that $\pi_T^{x,\psi^\star}$ corresponds to the invariant probability of the SDE

$$\mathrm{d}Y_t^\star = -\left(\frac{Y_t^\star - x}{2T}+\frac{1}{2}\nabla\psi^\star(Y_t^\star)\right)\mathrm{d}t+\mathrm{d}B_t$$

and let $(P_t^\star)_{t\ge0}$ be its corresponding semigroup. Particularly we have $\pi_T^{x,\psi^\star}\,P_t^\star = \pi_T^{x,\psi^\star}$ for any $t\ge0$. Similarly, for any $t\ge0$ it holds $\pi_T^{x,\psi^n}\,P_t^n=\pi_T^{x,\psi^n}$, with $(P_t^n)_{t\ge0}$ being the semigroup introduced above.

Given the above premises, from the weak triangle inequality (6.5.7) we deduce that

$$\begin{aligned} \mathbf{W}_{f_x^\nu}(\pi_T^{x,\psi^\star},\,\pi_T^{x,\psi^n}) &= \mathbf{W}_{f_x^\nu}(\pi_T^{x,\psi^\star}\,P_t^\star,\,\pi_T^{x,\psi^n}\,P_t^n) \\ &\le C_\Delta\left(\mathbf{W}_{f_x^\nu}(\pi_T^{x,\psi^\star}\,P_t^\star,\,\pi_T^{x,\psi^\star}\,P_t^n)+\mathbf{W}_{f_x^\nu}(\pi_T^{x,\psi^\star}\,P_t^n,\,\pi_T^{x,\psi^n}\,P_t^n)\right) \\ &\overset{(6.5.6)}{\le} C_\Delta\,\mathbf{W}_{f_x^\nu}(\pi_T^{x,\psi^\star}\,P_t^\star,\,\pi_T^{x,\psi^\star}\,P_t^n)+C_\Delta\,e^{-\lambda\,t}\,\mathbf{W}_{f_x^\nu}(\pi_T^{x,\psi^\star},\,\pi_T^{x,\psi^n})\,. \end{aligned}$$

Therefore for any $t>\lambda^{-1}\,\log C_\Delta$ it holds

$$\mathbf{W}_{f_x^\nu}(\pi_T^{x,\psi^\star},\,\pi_T^{x,\psi^n}) \le \frac{C_\Delta}{1-C_\Delta\,e^{-\lambda\,t}}\,\mathbf{W}_{f_x^\nu}(\pi_T^{x,\psi^\star}\,P_t^\star,\,\pi_T^{x,\psi^\star}\,P_t^n)\,. \tag{6.5.10}$$

We will show that $\mathbf{W}_{f_x^\nu}(\pi_T^{x,\psi^\star} P_t^\star,\, \pi_T^{x,\psi^\star} P_t^n)$ is exponentially small as $n \uparrow +\infty$. Hence consider the diffusion processes

$$\begin{cases} \mathrm{d}Y_t^\star = -\Big(\frac{Y_t^\star - x}{2T} + \frac{1}{2}\nabla\psi^\star(Y_t^\star) \Big)\mathrm{d}t + \mathrm{d}B_t \,, \\ \mathrm{d}Z_t = -\Big(\frac{Z_t - x}{2T} + \frac{1}{2}\nabla\psi^n(Z_t) \Big)\mathrm{d}t + \mathrm{d}B_t \,, \\ Y_0^\star = Z_0 \sim \pi_T^{x,\psi^\star} \,, \end{cases} \tag{6.5.11}$$

where $(B_t)_{t\geq 0}$ is the same d-dimensional Brownian motion. Notice $Y_t^\star \sim \pi_T^{x,\psi^\star} P_t^\star$ whereas $Z_t \sim \pi_T^{x,\psi^\star} P_t^n$, which yields to

$$\mathbf{W}_{f_x^\nu}(\pi_T^{x,\psi^\star} P_t^\star,\, \pi_T^{x,\psi^\star} P_t^n) \leq \mathbb{E}_{Y_t^\star, Z_t}\Big[f_x^\nu(|Y_t^\star - Z_t|)\,(1 + \varepsilon V(Y_t^\star) + \varepsilon V(Z_t^\star)) \Big] \,. \tag{6.5.12}$$

By construction, if we introduce $r_t := |Y_t^\star - Z_t|$, we immediately get

$$\mathrm{d}r_t^2 = -\frac{r_t^2}{T}\,\mathrm{d}t - \langle \nabla\psi^\star(Y_t^\star) - \nabla\psi^n(Z_t),\, Y_t^\star - Z_t \rangle\ \mathrm{d}t \,.$$

By reasoning as in the proof of Theorem 6.2.1, fix $\delta > 0$ and consider the function $\rho_\delta(r) := \sqrt{r+\delta}$. Then it holds

$$\mathrm{d}\rho_\delta(\bar r_t^2) = -(2T)^{-1}\frac{r_t^2}{\rho_\delta(r_t^2)}\mathrm{d}t - (2\,\rho_\delta(\bar r_t^2))^{-1}\,\langle \nabla\psi^\star(Y_t^\star) - \nabla\psi^n(Z_t),\, Y_t^\star - Z_t\rangle\,\mathrm{d}t,$$

and hence we deduce that it holds

$$\begin{aligned} \mathrm{d}f_x^\nu(\rho_\delta(r_t^2)) &= -\frac{(f_x^\nu)'(\rho_\delta(r_t^2))}{2}\Big(T^{-1}\frac{r_t^2}{\rho_\delta(r_t^2)} \\ &\quad + (\rho_\delta(r_t^2)^{-1}\,\langle \nabla\psi^\star(Y_t^\star) - \nabla\psi^n(Z_t),\, Y_t^\star - Z_t\rangle \Big)\,\mathrm{d}t \\ &\leq -(T^{-1} + \kappa_{\psi^n}(r_t))\,\frac{(f_x^\nu)'(\rho_\delta(r_t^2))}{2}\,\frac{r_t^2}{\rho_\delta(r_t^2)}\,\mathrm{d}t \\ &\quad + \frac{(f_x^\nu)'(\rho_\delta(r_t^2))}{2}\,\frac{r_t}{\rho_\delta(r_t^2)}\,|\nabla\psi^\star - \nabla\psi^n|(Y_t^\star)\mathrm{d}t \\ &\overset{(6.1.3)}{\leq} -(T^{-1} + \alpha_{\nu,n} - r_t^{-1}\tilde g_\nu(r_t))\,\frac{(f_x^\nu)'(\rho_\delta(r_t^2))}{2}\,\frac{r_t^2}{\rho_\delta(r_t^2)}\,\mathrm{d}t \\ &\quad + \frac{(f_x^\nu)'(\rho_\delta(r_t^2))}{2}\,\frac{r_t}{\rho_\delta(r_t^2)}\,|\nabla\psi^\star - \nabla\psi^n|(Y_t^\star)\mathrm{d}t \,. \end{aligned}$$

Since $\alpha_{\nu,n} > \alpha_\nu - T^{-1}$ (cf. Theorem 6.1.1), the sublinearity $\tilde{g}_\nu(r) \leq \tilde{G}_\nu\, r$ and the upper bound (6.5.8) for $(f_x^\nu)'$, imply

$$\begin{aligned}
\mathrm{d} f_x^\nu(\rho_\delta(r_t^2)) &\leq (\tilde{G}_\nu - \alpha_\nu)^+ \,\frac{(f_x^\nu)'(\rho_\delta(r_t^2))}{2}\, \rho_\delta(r_t^2)\,\mathrm{d}t + \frac{1}{2}\,|\nabla\psi^\star - \nabla\psi^n|(Y_t^\star)\mathrm{d}t \\
&\overset{(6.5.8)}{\leq} C_I^{-1}\,(\tilde{G}_\nu - \alpha_\nu)^+\, f_x^\nu(\rho_\delta(r_t^2))\,\mathrm{d}t + \frac{1}{2}\,|\nabla\psi^\star - \nabla\psi^n|(Y_t^\star)\mathrm{d}t\,.
\end{aligned}$$

Gronwall Lemma and the monotonicity of f_x^ν (*i.e.*, $(f_x^\nu)' \geq 0$, cf. (6.5.8)) finally yield for any $t \geq 0$ to

$$\begin{aligned}
f_x^\nu(|Y_t^\star - Z_t|) \leq f_x^\nu(\rho_\delta(r_t^2)) &\leq \exp(C_I^{-1}\,(\tilde{G}_\nu - \alpha_\nu)^+\, t)\,\frac{1}{2}\int_0^t |\nabla\psi^\star - \nabla\psi^n|(Y_s^\star)\,\mathrm{d}s \\
&\overset{(6.3.9)}{\leq} \exp(C_I^{-1}\,(\tilde{G}_\nu - \alpha_\nu)^+\, t)\,\frac{1}{2}\int_0^t (A|Y_s^\star| + B)\mathrm{d}\,s \prod_{k=0}^{n-1} \frac{\hat{\gamma}_k^\mu\,\hat{\gamma}_k^\nu}{T^2}\,.
\end{aligned}$$

where the last step follows from Theorem 6.3.5. By recalling (6.5.12) and taking expectation, so far we have proven that

$$\begin{aligned}
&\mathbf{W}_{f_x^\nu}(\pi_T^{x,\psi^\star}\, P_t^\star, \pi_T^{x,\psi^\star}\, P_t^n) \\
&\quad \leq \frac{1}{2}\, e^{C_I^{-1}(\tilde{G}_\nu - \alpha_\nu)^+ t} \int_0^t \mathbb{E}\Big[(A|Y_s^\star| + B)(1 + \varepsilon V(Y_t^\star) + \varepsilon V(Z_t))\Big]\,\mathrm{d}s \prod_{k=0}^{n-1} \frac{\hat{\gamma}_k^\mu\,\hat{\gamma}_k^\nu}{T^2}.
\end{aligned} \tag{6.5.13}$$

Next, we claim that the above integral is bounded by a constant independent from $n \in \mathbb{N}$. From Young's inequality and the stationarity of $Y_s^\star \sim \pi_T^{x,\psi^\star}$ we have

$$\begin{aligned}
&\int_0^t \mathbb{E}\Big[(A|Y_s^\star| + B)(1 + \varepsilon V(Y_t^\star) + \varepsilon V(Z_t))\Big]\,\mathrm{d}s \\
&\qquad \leq Bt(1 + \varepsilon\mathbb{E}[V(Y_0^\star)] + \varepsilon\mathbb{E}[V(Z_t)]) \\
&\qquad\quad + A\,t\,\mathbb{E}[|Y_0^\star|] + \frac{A\varepsilon t}{2}(2\,\mathbb{E}[|Y_0^\star|^2] + \mathbb{E}[V(Y_0^\star)^2] + \mathbb{E}[V(Z_t)^2])\,.
\end{aligned} \tag{6.5.14}$$

At this point it is enough noticing that the geometric drift condition (6.1.10) obtained in the proof Corollary 6.1.3 guarantees the finiteness of the fourth moments of the random variables appearing in the last display and hence the finiteness of the above expected values. For exposition's clarity we provide a proof of this last statement in Corollary 6.A.4 in the Appendix, where we show in (6.A.4) that $U(t, \nu, x, A, B, T)$, the upper-bounding constant, can be chosen independently from $n \in \mathbb{N}$.

As a byproduct of (6.5.9), (6.5.10) and (6.5.13), and by minimising over $t > \lambda^{-1}\,\log C_\Delta$ we have proven that

$$\|\nabla^2\varphi^{n+1} - \nabla^2\varphi^\star\|_{\mathrm{F}}(x) \leq C(x, T, \nu, A, B) \prod_{k=0}^{n-1} \frac{\hat{\gamma}_k^\mu\,\hat{\gamma}_k^\nu}{T^2}\,,$$

where the above constant $C(x,T,\nu,A,B)$ is equal to

$$C := \inf_{t>\lambda^{-1}\log C_\Delta} \frac{10}{3}\frac{C_x}{T^2\varepsilon^{1/2}}\left(\frac{2}{C_I} \vee \frac{f_x^\nu(R_2)^{-1}}{\varepsilon^{1/2}}\right)\frac{C_\Delta}{1-C_\Delta e^{-\lambda t}}\frac{1}{2}e^{C_I^{-1}(\tilde{G}_\nu-\alpha_\nu)^+ t}U\,, \tag{6.5.15}$$

where $U = U(t,\nu,x,A,B,T)$ is the constant defined at (6.A.4).

By following the same line of reasoning, it is possible proving that

$$\|\nabla^2\psi^{n+1} - \nabla^2\psi^\star\|_{\mathrm{F}(x)} \le C(x,T,\mu,A,B)\,\frac{T}{\hat{\gamma}_n^\mu}\prod_{k=0}^{n}\frac{\hat{\gamma}_k^\mu\,\hat{\gamma}_k^\nu}{T^2}\,,$$

and the constant $C(x,T,\mu,A,B)$ can be built in analogy to (6.5.15). □

6.5.1 Explicit construction and contractive properties of the $\mathrm{W}_{f_x^\nu}$-distance

Here we carry out the explicit construction of f_x^ν and the proof of (6.5.6) and (6.5.7). The following result follows form [EGZ19b, Theorem 2.2].

Firstly, notice the geometric drift condition (6.5.4) implies

$$\begin{aligned}
&\mathcal{L}_{\psi^n}V(z) + \mathcal{L}_{\psi^n}V(y) < 0 \quad \forall (z,y) \notin \mathrm{B}_{2\mathrm{d}}(2B_\nu/A_\nu)\,,\\
&\varepsilon\mathcal{L}_{\psi^n}V(z) + \varepsilon\mathcal{L}_{\psi^n}V(y) < -\frac{A_\nu}{2}(1\wedge 4B_\nu\varepsilon)(1+\varepsilon V(z)+\varepsilon V(y))\\
&\qquad\qquad\qquad\qquad\qquad\forall (z,y) \notin \mathrm{B}_{2\mathrm{d}}(4B_\nu(1+A_\nu^{-1}))\,,
\end{aligned} \tag{6.5.16}$$

where $\mathrm{B}_{2\mathrm{d}}(r^2) \subseteq \mathbb{R}^{2d}$ denotes the centred Euclidean ball of radius r. For later convenience let us also define the radii

$$\begin{aligned}
R_1 &:= \sup\{|x-y| \,:\, (x,y) \in \mathrm{B}_{2\mathrm{d}}(2B_\nu/A_\nu)\}\,,\\
R_2 &:= \sup\{|x-y| \,:\, (x,y) \in \mathrm{B}_{2\mathrm{d}}(4B_\nu(1+A_\nu^{-1}))\}\,.
\end{aligned} \tag{6.5.17}$$

Now, take $\varepsilon \in (0,1)$ satisfying the condition

$$(4B_\nu\varepsilon)^{-1} \ge \int_0^{R_1}\int_0^s \exp\left(\frac{(\tilde{G}_\nu-\alpha_\nu)^+}{4}(s^2-r^2) + 2\,\varepsilon^{1/2}(s-r)\right)\mathrm{d}r\,\mathrm{d}s\,, \tag{6.5.18}$$

which is always possible since the left hand side diverges as ε vanishes, whereas the right hand side is bounded.

Finally, define

$$f_x^\nu(r) := \int_0^{r\wedge R_2}\phi(s)g(s)\,\mathrm{d}s, \quad \text{with } \phi(r) := \exp\left(-\frac{(\tilde{G}_\nu-\alpha_\nu)^+}{8}r^2 - 2\,\varepsilon^{1/2}\,r\right),$$

$$\Phi(r) := \int_0^r \phi(s)\,\mathrm{d}s \quad \text{and} \quad g(r) := 1 - \frac{\int_0^{r\wedge R_1}\frac{\Phi(s)}{\phi(s)}\mathrm{d}s}{4\int_0^{R_1}\frac{\Phi(s)}{\phi(s)}\mathrm{d}s} - \frac{\int_0^{r\wedge R_2}\frac{\Phi(s)}{\phi(s)}\mathrm{d}s}{4\int_0^{R_2}\frac{\Phi(s)}{\phi(s)}\mathrm{d}s}.$$

Let us also consider the positive quantities

$$\xi^{-1} := \int_0^{R_1} \frac{\Phi(s)}{\phi(s)} \mathrm{d}s \quad \text{and} \quad \beta^{-1} := \int_0^{R_2} \frac{\Phi(s)}{\phi(s)} \mathrm{d}s \,,$$

and notice that (6.5.18) equivalently reads as $4B_\nu\varepsilon \leq \xi$. The function f_x^ν is clearly bounded, increasing and concave. Moreover, from the above definitions we immediately deduce the validity of the properties stated at (6.5.8) with

$$C_I := \phi(R_2) \,, \tag{6.5.19}$$

and the inequality regarding the first derivative can actually be strengthen since for any $r < R_2$ it holds

$$(f_x^\nu)'(r)\; \phi(r)^{-1} = g(r) \in [1/2, 1] \,.$$

Finally, a straightforward differentiation shows that for any $r \in (0, R_1) \cup (R_1, R_2)$ it holds

$$\begin{aligned}(f_x^\nu)''(r) &= -\left(\frac{(\tilde{G}_\nu - \alpha_\nu)^+}{4}\, r + 2\,\varepsilon^{1/2}\right) \phi(r) g(r) - \frac{\mathbf{1}_{\{r<R_1\}}\xi + \beta}{4}\, \Phi(r) \\ &\leq -\left(\frac{(\tilde{G}_\nu - \alpha_\nu)^+}{4}\, r + 2\,\varepsilon^{1/2}\right) (f_x^\nu)'(r) - \frac{\mathbf{1}_{\{r<R_1\}}\xi + \beta}{4}\, f_x^\nu(r) \,.\end{aligned} \tag{6.5.20}$$

We conclude with the proof of the triangle inequality and the contractive property for the distorted Wasserstein semi-distance introduced at (6.5.5)

$$\mathbf{W}_{f_x^\nu}(\cdot,\cdot) := \inf_{\pi \in \Pi(\cdot,\cdot)} \mathbb{E}_{(Y,Z)\sim\pi}\left[f_x^\nu(|Y-Z|)\,(1 + \varepsilon V(Y) + \varepsilon V(Z)) \right] .$$

Proof of the weak triangle inequality (6.5.7). The following proof is an adaptation of [HMS11, Lemma 4.14]. It is enough showing that there exists $C_\Delta > 0$ such that for any $y,\, z,\, p \in \mathbb{R}^d$ it holds

$$\begin{aligned}&f_x^\nu(|y-z|)(1+\varepsilon V(y)+\varepsilon V(z)) \\ &\quad \leq C_\Delta \left[f_x^\nu(|y-p|)(1+\varepsilon V(y)+\varepsilon V(p)) + f_x^\nu(|p-z|)(1+\varepsilon V(p)+\varepsilon V(z)) \right] .\end{aligned} \tag{6.5.21}$$

Firstly, notice that for any $y, z \in \mathbb{R}^d$ such that $|y-z| \leq R_2$ it holds

$$V(y) \leq \max\{2,\, 1 + 2\,R_2^2\}\, V(z) \,. \tag{6.5.22}$$

Without loss of generalities assume that $|z| \leq |y|$ (and hence $V(z) \leq V(y)$). Since $f_x^\nu(r) \leq f_x^\nu(R_2)$, if $|y-p| \geq R_2$ then

$$\begin{aligned}f_x^\nu(|y-z|)(1+\varepsilon V(y)+\varepsilon V(z)) &\leq f_x^\nu(R_2)(1+2\,\varepsilon V(y)) \\ &= f_x^\nu(|y-p|)(1+2\varepsilon V(y)) \\ &\leq 2\, f_x^\nu(|y-p|)(1+\varepsilon V(y)+\varepsilon V(p)).\end{aligned} \tag{6.5.23}$$

On the other hand, if $|y-p| \le R_2$ then (6.5.22) and the subadditivity of f_x^ν (which is guaranteed by its concavity) imply

$$
\begin{aligned}
&f_x^\nu(|y-z|)(1+\varepsilon V(y)+\varepsilon V(z)) \\
&\quad \le (f_x^\nu(|y-p|)+f_x^\nu(|p-z|))(1+\varepsilon V(y)+\varepsilon V(z)) \\
&\quad \le f_x^\nu(|y-p|)(1+2\varepsilon V(y))+f_x^\nu(|p-z|)(1+\max\{2,\,1+2\,R_2^2\}\,\varepsilon V(z)+\varepsilon V(z)).
\end{aligned}
\tag{6.5.24}
$$

As a byproduct of (6.5.23) and (6.5.24) we get the validity of (6.5.21) (and hence of (6.5.7)) with

$$C_\Delta := \max\{3,\, 2+2\,R_2^2\}\ .$$

□

Proof of the contraction (6.5.6). Fix two probability measure μ_1, $\mu_2 \in \mathcal{P}(\mathbb{R}^d)$ and let $\pi \in \Pi(\mu_1,\mu_2)$ be a coupling between them. Our proof starts by considering the coupling by reflection, *i.e.*, the diffusion processes

$$
\begin{cases}
\mathrm{d}Z_t = -\left(\frac{Z_t-x}{2T}+\frac{1}{2}\nabla\psi^n(Z_t)\right)\mathrm{d}t+\mathrm{d}B_t\ , \\
\mathrm{d}Y_t = -\left(\frac{Y_t-x}{2T}+\frac{1}{2}\nabla\psi^n(Y_t)\right)\mathrm{d}t+\mathrm{d}\hat{B}_t \quad \forall\, t\in[0,\tau) \text{ and } Y_t=Z_t \quad \forall\, t\ge\tau\ , \\
(Z_0,Y_0)\sim\pi\ ,
\end{cases}
$$

where $\tau := \inf\{s\ge 0 : Z_s=Y_s\}$, and $(\hat{B}_t)_{t\ge 0}$ is defined as

$$
\mathrm{d}\hat{B}_t := (\mathrm{Id}-2\,e_t\,e_t^{\mathsf{T}}\,\mathbf{1}_{\{t<\tau\}})\,\mathrm{d}B_t \quad \text{where } e_t := \begin{cases} \frac{Z_t-Y_t}{|Z_t-Y_t|} & \text{when } |Z_t-Y_t|>0\,, \\ u & \text{when } |Z_t-Y_t|=0\,. \end{cases}
$$

with $u\in\mathbb{R}^d$ being a fixed (arbitrary) unit-vector. By Lévy's characterisation, $(\hat{B}_t)_{t\ge0}$ is a d-dimensional Brownian motion, hence $Z_t\sim\mu_1 P_t^n$ and $Y_t\sim\mu_2 P_t^n$, and finally $\mathrm{d}W_t := e_t^{\mathsf{T}}\mathrm{d}B_t$ is a one-dimensional Brownian motion. By setting $r_t = |Z_t-Y_t|$ and by applying Ito-Tanaka formula as in the proof of Theorem 6.2.1, the trivial bound $\alpha_{\nu,n} > \alpha_\nu - T^{-1}$ (cf. Theorem 6.1.1) and the sublinearity $\tilde{g}_\nu(r)\le\tilde{G}_\nu\, r$ imply that

$$
\begin{aligned}
\mathrm{d}f_x^\nu(r_t) &\le \left(2\,(f_x^\nu)''(r_t)-\frac{r_t\,(f_x^\nu)'(r_t)}{2}(\alpha_{\nu,n}+T^{-1}-r_t^{-1}\tilde{g}_\nu(r_t))\right)\mathrm{d}t \\
&\qquad\qquad +2\,(f_x^\nu)'(r_t)\,\mathrm{d}W_t \\
&\le \left(2\,(f_x^\nu)''(r_t)+\frac{r_t\,(f_x^\nu)'(r_t)}{2}(\tilde{G}_\nu-\alpha_\nu)^+\right)\mathrm{d}t+2\,(f_x^\nu)'(r_t)\,\mathrm{d}W_t \qquad (6.5.25)\\
&\overset{(6.5.20)}{\le} -4\,\varepsilon^{1/2}\,(f_x^\nu)'(r_t)\mathrm{d}t-\mathbf{1}_{\{r_t<R_1\}}\frac{\xi}{2}f_x^\nu(r_t)\,\mathrm{d}t-\mathbf{1}_{\{r_t<R_2\}}\frac{\beta}{2}f_x^\nu(r_t)\,\mathrm{d}t \\
&\qquad\qquad +2\,(f_x^\nu)'(r_t)\,\mathrm{d}W_t.
\end{aligned}
$$

The validity of the above Ito-Tanaka formula for f_x^ν can be proven as we have already done in the proof of Theorem 6.2.1 and for this reason we omit it; we refer the reader to [EGZ19b, Formula (5.15)] for a detailed discussion on that.

Next, we notice that from the geometric drift condition (6.5.4) and the definition of R_1, R_2 it follows that

$$\begin{aligned}
\mathrm{d}(1+\varepsilon V(Z_t)+\varepsilon V(Y_t)) &= \varepsilon(\mathcal{L}_{\psi^n}V(Z_t)+\mathcal{L}_{\psi^n}V(Y_t))\,\mathrm{d}t+2\varepsilon\,\langle Z_t+Y_t,\,\mathrm{d}B_t\rangle \\
&\quad -4\varepsilon\,\langle Y_t,\,e_t\rangle\,\mathrm{d}W_t \\
&\leq 2\varepsilon\,B_\nu\,\mathrm{d}t-\varepsilon\,A_\nu(V(Z_t)+V(Y_t))\,\mathrm{d}t+2\varepsilon\,\langle Z_t+Y_t,\,\mathrm{d}B_t\rangle-4\varepsilon\,\langle Y_t,\,e_t\rangle\,\mathrm{d}W_t \\
&\overset{(6.5.18)}{\leq} \mathbf{1}_{\{r_t<R_1\}}\Big(\xi/2-\varepsilon\,A_\nu(V(Z_t)+V(Y_t))\Big)\mathrm{d}t \\
&\quad +\varepsilon\,\mathbf{1}_{\{r_t\in[R_1,R_2)\}}(\mathcal{L}_{\psi^n}V(Z_t)+\mathcal{L}_{\psi^n}V(Y_t))\,\mathrm{d}t \\
&\quad +\varepsilon\,\mathbf{1}_{\{r_t\geq R_2\}}(\mathcal{L}_{\psi^n}V(Z_t)+\mathcal{L}_{\psi^n}V(Y_t))\,\mathrm{d}t+2\varepsilon\,\langle Z_t+Y_t,\,\mathrm{d}B_t\rangle \\
&\quad -4\varepsilon\,\langle Y_t,\,e_t\rangle\,\mathrm{d}W_t \\
&\overset{(6.5.16)}{\leq} \Big(\mathbf{1}_{\{r_t<R_1\}}\,\xi/2-\mathbf{1}_{\{r_t\geq R_2\}}\,\lambda\,(1+\varepsilon V(Z_t)+\varepsilon V(Y_t))\Big)\,\mathrm{d}t \\
&\quad +2\varepsilon\,\langle Z_t+Y_t,\,\mathrm{d}B_t\rangle-4\varepsilon\,\langle Y_t,\,e_t\rangle\,\mathrm{d}W_t
\end{aligned} \tag{6.5.26}$$

where in the last step we have taken

$$\lambda := \min\{\beta,\,A_\nu,\,4A_\nu B_\nu\varepsilon\}/2\,. \tag{6.5.27}$$

Finally, notice that the choice of coupling by reflection gives to the covariation between $f_x^\nu(r_t)$ and $1+\varepsilon V(Z_t)+\varepsilon V(Y_t)$ the expression

$$\begin{aligned}
\mathrm{d}[f_x^\nu(r_\cdot),\,1+\varepsilon V(Z_\cdot)+\varepsilon V(Y_\cdot)]_t &= 4\varepsilon\,r_t\,(f_x^\nu)'(r_t)\,\mathrm{d}t \\
&< 4\varepsilon^{1/2}\,(1+\varepsilon V(y)+\varepsilon V(z))\,(f_x^\nu)'(r_t)\,\mathrm{d}t\,.
\end{aligned} \tag{6.5.28}$$

where the last step follows from the trivial series of inequalities

$$\begin{aligned}
4\varepsilon\,|y-z| &\leq 4\varepsilon(1+\varepsilon V(y)+\varepsilon V(z))\left(\frac{|y|}{1+\varepsilon V(y)}+\frac{|z|}{1+\varepsilon V(z)}\right) \\
&\leq 4\varepsilon(1+\varepsilon V(y)+\varepsilon V(z))\,\sup_{y\in\mathbb{R}^d}\frac{2|y|}{1+\varepsilon V(y)} = 4(1+\varepsilon V(y)+\varepsilon V(z))\sqrt{\frac{\varepsilon}{1+\varepsilon}}\,.
\end{aligned}$$

If for sake of notation we set $F_V(z,y) := f_x^\nu(|y-z|)(1+\varepsilon V(z)+\varepsilon V(y))$ the inequalities (6.5.25), (6.5.26) and (6.5.28) gives

$$\begin{aligned}
\mathrm{d}F_V(Z_t,\, Y_t) &= (1+\varepsilon V(Z_t)+\varepsilon V(Y_t))\,\mathrm{d}f^\nu_x(r_t) + f^\nu_x(r_t)\,\mathrm{d}(1+\varepsilon V(Z_t)+\varepsilon V(Y_t)) \\
&\qquad + \mathrm{d}[f^\nu_x(r_\cdot),\, 1+\varepsilon V(Z_\cdot)+\varepsilon V(Y_\cdot)]_t \\
&\le -\mathbf{1}_{\{r_t<R_1\}}\frac{\xi}{2}\,(F_V(Z_t,Y_t)-f^\nu_x(r_t))\,\mathrm{d}t - \left(\mathbf{1}_{\{r_t<R_2\}}\frac{\beta}{2}+\mathbf{1}_{\{r_t\ge R_2\}}\lambda\right)F_V(Z_t,Y_t)\,\mathrm{d}t \\
&\qquad + \mathrm{d}M_t \\
&\le -\lambda\, F_V(Z_t,Y_t)\,\mathrm{d}t + \mathrm{d}M_t\,,
\end{aligned}$$

where

$$\begin{aligned}
\mathrm{d}M_t := 2\,(f^\nu_x)'(r_t)(1+\varepsilon V(Z_t)+\varepsilon V(Y_t))\,\mathrm{d}W_t + 2\varepsilon\, f^\nu_x(r_t)\,\langle Z_t+Y_t,\,\mathrm{d}B_t\rangle \\
-4\varepsilon\, f^\nu_x(r_t)\,\langle Y_t,\, e_t\rangle\,\mathrm{d}W_t
\end{aligned}$$

is a local martingale. Hence $e^{\lambda t}F_V(Z_t,Y_t)$ is a local-supermartingale.

Now, for any $N\in\mathbb{N}$ consider now the stopping time

$$\tau_N := \inf\{t\ge 0\colon |Z_t-Y_t|\le N^{-1} \;\text{ or }\; |Y_t|\vee|Z_t|\ge N\}\,,$$

and notice that $\tau_N\uparrow+\infty$ as N grows (cf. Corollary 6.1.3). Then the previous discussion, Gronwall Lemma and Fatou Lemma give

$$e^{\lambda t}\,\mathbb{E}[F_V(Z_t,Y_t)] \le \liminf_{M\to+\infty}\mathbb{E}[\mathbf{1}_{\{t<\tau_M\}}\,e^{\lambda t}F_V(Z_t,Y_t)] = \mathbb{E}[F_V(Z_0,Y_0)]\,.$$

In conclusion, since $Z_t\sim\mu_1P^n_t$ and $Y_t\sim\mu_2P^n_t$ we have

$$\mathbf{W}_{f^\nu_x}(\mu_1\,P^n_t,\mu_2\,P^n_t) \le \mathbb{E}[F_V(Z_t,Y_t)] \le e^{-\lambda\,t}\,\mathbb{E}_\pi[F_V(Z_0,Y_0)]\,.$$

By minimising the above bound over $\pi\in\Pi(\mu_1,\mu_2)$ concludes the proof of (6.5.6). □

6.6 Convergence rates for marginals with strictly log-concave densities

In this section we further develop the discussion started in Remark 6.3.3. Therefore assume the validity of **A7**, or equivalently **A6** with $\tilde g_\mu=\tilde g_\nu=g_\mu=g_\nu\equiv 0$. Let us firstly observe that Theorem 6.1.1 in this particular setting simply reads as

Theorem 6.6.1. *Assume **A6** with $\tilde g_\mu=\tilde g_\nu=g_\mu=g_\nu\equiv 0$. Then there exist two monotone increasing sequences $(\alpha_{\mu,n})_{n\in\mathbb{N}}\subseteq(\alpha_\mu-T^{-1},\alpha_\mu-T^{-1}+(\beta_\nu\,T^2)^{-1}]$ and $(\alpha_{\nu,n})_{n\in\mathbb{N}}\subseteq(\alpha_\nu-T^{-1},\alpha_\nu-T^{-1}+(\beta_\mu\,T^2)^{-1}]$ such that for any $n\ge 1$ for any $n\in\mathbb{N}$ it holds $r>0$ it holds*

$$\kappa_{\varphi^n}(r)\ge\alpha_{\mu,n}\quad\text{ and }\quad\kappa_{\psi^n}(r)\ge\alpha_{\nu,n}\,.$$

These two sequences are defined as

$$\begin{cases} \alpha_{\mu,0} := \alpha_\mu - T^{-1}\,, \\ \alpha_{\mu,n+1} := \alpha_\mu - T^{-1} + \Big(T^2\,\beta_\nu + (\alpha_{\mu,n} + T^{-1})^{-1}\Big)^{-1}\,, \quad n \in \mathbb{N}\,, \end{cases}$$

and

$$\begin{cases} \alpha_{\nu,0} := \alpha_\nu - T^{-1}\,, \\ \alpha_{\nu,n+1} := \alpha_\nu - T^{-1} + \Big(T^2\,\beta_\mu + (\alpha_{\nu,n} + T^{-1})^{-1}\Big)^{-1}\,, \quad n \in \mathbb{N}\,. \end{cases}$$

Moreover, both sequences converge respectively to

$$\begin{aligned} \alpha_{\varphi^\star} &:= \frac{1}{2}\Big(\alpha_\mu + \sqrt{\alpha_\mu^2 + 4\alpha_\mu/(T^2\beta_\nu)}\Big) - T^{-1}\,, \\ \alpha_{\psi^\star} &:= \frac{1}{2}\Big(\alpha_\nu + \sqrt{\alpha_\nu^2 + 4\alpha_\nu/(T^2\beta_\mu)}\Big) - T^{-1}\,, \end{aligned} \tag{6.6.1}$$

and for any $r > 0$ it holds

$$\kappa_{\varphi^\star}(r) \geq \alpha_{\varphi^\star} \quad \text{and} \quad \kappa_{\psi^\star}(r) > \alpha_{\psi^\star}\,,$$

where $\varphi^\star$ and $\psi^\star$ are the Schrödinger potentials introduced in (6.0.1).

Proof. This is a particular instance of Theorem 6.1.1 when $\tilde{g}_\mu = \tilde{g}_\nu = g_\mu = g_\nu \equiv 0$. The only statement that does not follow from that theorem is the identification of the limit values $\alpha_{\varphi^\star}$ and $\alpha_{\psi^\star}$ in (6.6.1). We will only prove the first one since the second identity can be proven in the same way. From Theorem 6.1.1 we already know that $\alpha_{\mu,n} \uparrow \alpha_{\varphi^\star} \in (\alpha_\mu - T^{-1}, \alpha_\mu - T^{-1} + (\beta_\nu\,T^2)^{-1}]$. Consider the shifted sequence $\theta_n^\mu := \alpha_{\mu,n} + T^{-1}$. Clearly $\theta_n^\mu > 0$, $\theta_n^\mu \uparrow \theta_\infty^\mu := \alpha_{\varphi^\star} + T^{-1}$ and the latter limit value can be seen as a fixed point for the iteration

$$\theta_{n+1}^\mu = \alpha_\nu + (T^2\,\beta_\nu + (\theta_n^\mu)^{-1})^{-1}\,.$$

A straightforward computation shows that the there are just two possible fixed point solutions, namely

$$\frac{1}{2}\Big(\alpha_\mu - \sqrt{\alpha_\mu^2 + 4\alpha_\mu/(T^2\beta_\nu)}\Big) \text{ and } \frac{1}{2}\Big(\alpha_\mu + \sqrt{\alpha_\mu^2 + 4\alpha_\mu/(T^2\beta_\nu)}\Big)\,.$$

Since one solution is negative, whereas $(\theta_n^\mu)_{n\in\mathbb{N}}$ is a positive increasing sequence, we immediately deduce that θ_∞^μ equals the largest (and positive) fixed point. This proves (6.6.1). □

From the previous result we immediately deduce the explicit expressions for the rates of convergence appearing in Remark 6.3.3. Indeed Theorem 6.6.1 implies the strict convexity of the Schrödinger potentials, which allows to perform all our contraction estimates without relying on reflection coupling. Namely, by relying on synchronous coupling we are able to improve on Theorem 6.3.2 and consider $\mathbf{W}_2$-distances, obtaining

Theorem 6.6.2. *Assume A7. Then for any $n \geq 1$*

$$\begin{aligned} \|\nabla\varphi^n - \nabla\varphi^\star\|_{L^2(\mu)} &\leq \frac{T}{\gamma^\mu_{n-1}} \prod_{k=0}^{n-1} \frac{\gamma^\mu_k\,\gamma^\nu_k}{T^2}\,\|\nabla\psi^0 - \nabla\psi^\star\|_{L^2(\nu)}\,, \\ \|\nabla\psi^n - \nabla\psi^\star\|_{L^2(\nu)} &\leq \prod_{k=0}^{n-1} \frac{\gamma^\mu_k\,\gamma^\nu_k}{T^2}\,\|\nabla\psi^0 - \nabla\psi^\star\|_{L^2(\nu)}\,, \end{aligned} \tag{6.6.2}$$

and

$$\begin{aligned} \mathbf{W}_2(\pi^{n,n},\boldsymbol{\pi}^T) &\leq T \prod_{k=0}^{n-1} \frac{\gamma^\mu_k\,\gamma^\nu_k}{T^2}\,\|\nabla\psi^0 - \nabla\psi^\star\|_{L^2(\nu)}\,, \\ \mathbf{W}_2(\pi^{n+1,n},\boldsymbol{\pi}^T) &\leq \gamma^\nu_n \prod_{k=0}^{n-1} \frac{\gamma^\mu_k\,\gamma^\nu_k}{T^2}\,\|\nabla\psi^0 - \nabla\psi^\star\|_{L^2(\nu)}\,, \end{aligned} \tag{6.6.3}$$

where $(\gamma^\mu_k)_{k\in\mathbb{N}}$ and $(\gamma^\nu_k)_{k\in\mathbb{N}}$ are non-negative non-increasing sequences satisfying

$$\begin{cases} \gamma^\mu_0 := \alpha_\mu^{-1} \\ \gamma^\mu_{k+1} := (\alpha_\mu + (T^2\,\beta_\nu + \gamma^\mu_k)^{-1})^{-1} \end{cases} \quad \text{and} \quad \begin{cases} \gamma^\nu_0 := \alpha_\nu^{-1} \\ \gamma^\nu_{k+1} := (\alpha_\nu + (T^2\,\beta_\mu + \gamma^\nu_k)^{-1})^{-1} \end{cases} \tag{6.6.4}$$

and which converge respectively to the limit rates

$$\begin{aligned} \gamma^\mu_\infty &:= 2\left(\alpha_\mu + \sqrt{\alpha_\mu^2 + 4\alpha_\mu/(T^2\beta_\nu)}\right)^{-1} \\ \gamma^\nu_\infty &:= 2\left(\alpha_\nu + \sqrt{\alpha_\nu^2 + 4\alpha_\nu/(T^2\beta_\mu)}\right)^{-1}. \end{aligned} \tag{6.6.5}$$

Henceforth, as soon as

$$T > \frac{\beta_\mu\beta_\nu - \alpha_\mu\alpha_\nu}{\sqrt{\alpha_\mu\,\beta_\mu\,\alpha_\nu\,\beta_\nu\,(\alpha_\mu + \beta_\mu)(\alpha_\nu + \beta_\nu)}}\,, \tag{6.6.6}$$

the asymptotic rate is strictly less than one, and for any $T^{-2}\gamma^\mu_\infty\,\gamma^\nu_\infty < \lambda < 1$, there exists $C \geq 0$ such that for any $n \in \mathbb{N}^*$,

$$\begin{aligned} \|\nabla\varphi^n - \nabla\varphi^\star\|_{L^2(\mu)} + \|\nabla\psi^n - \nabla\psi^\star\|_{L^2(\nu)} &\leq C\lambda^n \|\nabla\psi^0 - \nabla\psi^\star\|_{L^2(\nu)}\,, \\ \mathbf{W}_2(\pi^{n,n},\boldsymbol{\pi}^T) + \mathbf{W}_2(\pi^{n+1,n},\boldsymbol{\pi}^T) &\leq C\lambda^n \|\nabla\psi^0 - \nabla\psi^\star\|_{L^2(\nu)}\,. \end{aligned}$$

Proof. The key $\mathbf{W}_2$-contractive estimates

$$\mathbf{W}_2^2(\pi_T^{x,\psi^n},\pi_T^{x,\psi^\star}) \leq (\gamma_n^\nu)^2 \int |\nabla\psi^n - \nabla\psi^\star|^2 \,\mathrm{d}\pi_T^{x,\psi^\star}$$
$$\text{with } \gamma_n^\nu = (\theta_n^\nu)^{-1} = (\alpha_{\nu,n} + T^{-1})^{-1},$$

and

$$\mathbf{W}_2^2(\pi_T^{x,\varphi^{n+1}},\pi_T^{x,\varphi^\star}) \leq (\gamma_n^\mu)^2 \int \left|\nabla\varphi^{n+1} - \nabla\varphi^\star\right|^2 \mathrm{d}\pi_T^{x,\varphi^\star}$$
$$\text{with } \gamma_n^\mu = (\theta_n^\mu)^{-1} = (\alpha_{\mu,n} + T^{-1})^{-1}.$$

can be obtained as in Corollary 6.2.3 and Proposition 6.3.1, this time directly considering a synchronous coupling (which allows to get estimates in Wasserstein $\mathbf{W}_2$-distance). Because of Theorem 6.6.1, these two rates sequences are non-negative, non-increasing and satisfy (6.6.4) and converge to γ_∞^μ, γ_∞^ν as in (6.6.5).

The proof of (6.6.2) and (6.6.3) follows as in Theorem 6.3.2.

Finally, from the explicit expressions (6.6.5), the exponential convergence condition, that is $T^{-2}\gamma_\infty^\mu\gamma_\infty^\nu < 1$, can be obtained as follows. By solving $T > \theta\,\gamma_\infty^\mu$ and $T > \theta^{-1}\,\gamma_\infty^\nu$, we deduce that for any $\theta \in (0,\infty)$ it holds

$$\begin{aligned} T > \theta\,\alpha_\mu^{-1} - \theta^{-1}\beta_\nu^{-1} \quad &\Leftrightarrow \quad T > \theta\,\gamma_\infty^\mu\,, \\ T > \theta^{-1}\alpha_\nu^{-1} - \theta\,\beta_\mu^{-1} \quad &\Leftrightarrow \quad T > \theta^{-1}\,\gamma_\infty^\nu\,, \end{aligned} \tag{6.6.7}$$

and therefore if

$$\begin{aligned} T &> \inf_{\theta\in(0,\infty)} \max\{\theta\,\alpha_\mu^{-1} - \theta^{-1}\beta_\nu^{-1},\, \theta^{-1}\alpha_\nu^{-1} - \theta\,\beta_\mu^{-1}\} \\ &= \frac{\alpha_\mu^{-1}\alpha_\nu^{-1} - \beta_\mu^{-1}\beta_\nu^{-1}}{\sqrt{(\alpha_\mu^{-1}+\beta_\mu^{-1})(\alpha_\nu^{-1}+\beta_\nu^{-1})}} = \frac{\beta_\mu\beta_\nu - \alpha_\mu\alpha_\nu}{\sqrt{\alpha_\mu\,\beta_\mu\,\alpha_\nu\,\beta_\nu\,(\alpha_\mu+\beta_\mu)(\alpha_\nu+\beta_\nu)}}, \end{aligned}$$

then we are guaranteed that $T^2 > \gamma_\infty^\mu\,\gamma_\infty^\nu$, and hence the exponential convergence of Sinkhorn's algorithm. □

Notice that if we start Sinkhorn's algorithm at $\psi^0 = U_\nu$, then the constant term appearing in the above right hand sides can always be bounded as in (6.3.12), which this time yields to

$$\|\nabla\psi^0 - \nabla\psi^\star\|^2_{\mathrm{L}^2(\nu)} = \|\nabla\log P_T\exp(-\varphi^\star)\|^2_{\mathrm{L}^2(\nu)} \leq T^{-2}\int |x-y|^2\,\mathrm{d}\boldsymbol{\pi}^T$$
$$\leq 2T^{-2}(M_2(\mu)+M_2(\nu))\,.$$

Clearly, the above result implies the L^1 and $\mathbf{W}_1$ results of Theorem 6.3.2 in the log-concave setting, with the rates defined as in Theorem 6.6.2.

Remark 6.6.3. *If the marginals are Gaussian distributions, i.e., if* $U_\mu = \alpha_\mu |x|^2/2$ *and* $U_\nu = \alpha_\nu |x|^2/2$ *then Theorem 6.6.2 proves the exponential convergence of Sinkhorn's algorithm for any choice of* $T > 0$ *since* $\beta_\mathfrak{p} = \alpha_\mathfrak{p}$ *for* $\mathfrak{p} \in \{\mu, \nu\}$ *and therefore the right hand side in* (6.6.6) *is null.*

Similarly, we may also give explicit expressions for the exponential rates $(\hat{\gamma}_k^\mu)_{k\in\mathbb{N}}$, $(\hat{\gamma}_k^\nu)_{k\in\mathbb{N}}$ appearing in Theorem 6.3.5, Theorem 6.4.1, Theorem 6.4.4 and Theorem 6.5.2 which in the log-concave setting read as

$$\begin{aligned}
\hat{\gamma}_k^\mu &:= \gamma_k^\mu \max\left\{T^{-1}\gamma_\infty^\mu\,, \left(1+\gamma_\infty^\mu\,\frac{A(1+|\nabla\varphi^\star(0)|)}{B}\right)\right\},\\
\hat{\gamma}_k^\nu &:= \gamma_k^\nu \max\left\{T^{-1}\gamma_\infty^\nu\,, \left(1+\gamma_\infty^\nu\,\frac{A(1+|\nabla\psi^\star(0)|)}{B}\right)\right\}.
\end{aligned}$$

Finally, when $\psi^0 = U_\nu$ (*i.e.*, $\varphi^0 = 0$) the above expressions read as

$$\begin{aligned}
\hat{\gamma}_k^\mu &:= \gamma_k^\mu \max\left\{T^{-1}\gamma_\infty^\mu\,, 2+T^{-1}\gamma_\infty^\mu\right\} = \gamma_k^\mu\,(2+T^{-1}\gamma_\infty^\mu)\,,\\
\hat{\gamma}_k^\nu &:= \gamma_k^\nu \max\left\{T^{-1}\gamma_\infty^\nu\,, \left(1+\left(T^{-1}\gamma_\infty^\nu+\frac{\gamma_\infty^\nu}{\gamma_\infty^\mu}\right)\frac{1+|\nabla\psi^\star(0)|}{1+|\nabla\varphi^\star(0)|}\right)\right\}.
\end{aligned}$$

Notice that, if for instance we assume the validity of (6.6.7) for $\theta = 1$, then the asymptotic rates read as

$$\begin{aligned}
&\hat{\gamma}_\infty^\mu = \gamma_\infty^\mu\,(2+T^{-1}\gamma_\infty^\mu) < 3\,\gamma_\infty^\mu\,,\\
&\hat{\gamma}_\infty^\nu = \gamma_\infty^\nu\left(1+\left(T^{-1}\gamma_\infty^\nu+\frac{\gamma_\infty^\nu}{\gamma_\infty^\mu}\right)\frac{1+|\nabla\psi^\star(0)|}{1+|\nabla\varphi^\star(0)|}\right) < \gamma_\infty^\nu\left(1+M\,\frac{1+|\nabla\psi^\star(0)|}{1+|\nabla\varphi^\star(0)|}\right),\\
&\qquad\text{with } M = 1+\sup_{s\geq 0}\frac{\alpha_\mu\,s+\sqrt{\alpha_\mu^2\,s^2+4\alpha_\mu/\beta_\nu}}{\alpha_\nu\,s+\sqrt{\alpha_\nu^2\,s^2+4\alpha_\nu/\beta_\mu}} < +\infty\,.
\end{aligned}$$

Therefore the exponential convergence of Sinkhorn's algorithm in Theorem 6.3.5, Theorem 6.4.1, Theorem 6.4.4 and Theorem 6.5.2 holds for any

$$\begin{aligned}
T > \max\Bigg\{&3\,\alpha_\mu^{-1} - 3^{-1}\beta_\nu^{-1},\\
&\left(1+M\,\frac{1+|\nabla\psi^\star(0)|}{1+|\nabla\varphi^\star(0)|}\right)\alpha_\nu^{-1} - \left(1+M\,\frac{1+|\nabla\psi^\star(0)|}{1+|\nabla\varphi^\star(0)|}\right)^{-1}\beta_\mu^{-1}\Bigg\}.
\end{aligned}$$

Bibliographical Remarks

When working on discrete spaces, convergence of Sinkhorn's algorithm is well-known and we refer to the book [PC19] for an extensive overview. However, let us mention that in the discrete setting, convergence of Sinkhorn's algorithm dates back at least to [Sin64] and [SK67]. More recently, [FL89, BLN94] show that Sinkhorn's iterates are equivalent to a sequence of iterates associated to an appropriate contraction in the Hilbert projective metrics, therefore proving the convergence of the algorithm boils down to studying a fixed-point problem, whose (exponential) convergence can be deduced from Birckoff's Theorem. In the case of continuous state spaces, the use of Birckoff's Theorem for the Hilbert metrics has also been employed by [CGP16a, DdBD24, Ber20], in order to establish the exponential convergence of Sinkhorn's algorithm under the condition that the state space is compact or that the cost function is bounded.

When considering non compact spaces with possibly unbounded costs and marginals, results are scarcer and the techniques developed to handle the discrete setting cannot apply as such and new ideas have emerged. When it comes to results that allow for unbounded costs, [Rus95] shows qualitative convergences of iterates in relative entropy and total variation for Sinkhorn's plans. More recently, [NW23] establishes qualitative convergence both on the primal and dual sides under mild assumptions. Concerning convergence rates, [Lég21] gives an interpretation of Sinkhorn's algorithm as a block coordinate descent on the dual problem and obtains convergence of marginal distributions in relative entropy at a linear rate n^{-1} under minimal assumptions. [EN22b] derives polynomial rates of convergence in Wasserstein distance assuming, among other things, that marginals admit exponential moments. Lastly, [GN22] improves existent polynomial convergence rates for optimal plans with respect to a symmetric relative entropy.

Let us further report on results available for multimarginal entropic optimal transport problem and the natural extension of Sinkhorn's algorithm in this setting. For bounded costs and marginals, (or equivalently compact spaces) [CL20] shows well-posedness of the Schrödinger system and smooth dependence of Schrödinger potentials on the marginal inputs. In [DMG20] the authors manage to show qualitative convergence of Sinkhorn's iterates towards the Schrödinger potentials in L^p-norms using tools from calculus of variations: their results require bounded costs but apply to multimarginal problems. These results have been subsequently improved by Carlier [Car22] who establishes exponential convergence.

The results proven in this chapter have been presented in [CDG23]. In contrast to the above existing works, the main contribution of [CDG23] is to establish exponential convergence bounds for the gradients and Hessians of Sinkhorn's iterates as well as for the optimal plans. To the best of our knowledge these findings are new both in their dual and primal formulation in that they represent the first exponential convergence results that hold for unbounded costs and marginals. They also are among the very few results that yield convergence of derivatives of potentials. On this subject, let us also mention the recent work [GNCD23] (from which Chapter 5 is based) where the state space is the d-dimensional torus and therefore deals with bounded cost functions. As [GNCD23], our proofs are mainly probabilistic, and differ from other proposed methodologies in that they rely on one-sided integrated semiconvexity estimates for potentials along Sinkhorn's iterates. These estimates are by themselves a new result, that has potentially several further implications. Though both the approaches we proposed in [CDG23] and in [GNCD23] are inspired by coupling methods and stochastic control, there is a fundamental difference. In [GNCD23] exponential convergence is achieved through Lipschitz estimates on potentials. In the current setup, we make assumptions on the integrated log-concavity profile of the marginals; these assumption are of geometric nature and not perturbative.

In addition, to the best of our knowledge and understanding, ours are the very first exponential convergence results of Sinkhorn's iterates and plans for unbounded costs and marginals. Moreover, when solely considering Gaussian marginals (that is when $\alpha_\mu = \beta_\mu$ and $\alpha_\nu = \beta_\nu$, see Remark 6.6.3) our results holds for any positive regularising parameter $T > 0$, which is crucial when considering application where we take $\nabla\varphi^\star$ as proxy for the Brenier map.

We should also mention that, after [GNCD23] and [CDG23] appeared, very recently [Eck23] has managed to extend the approach based on Hilbert's projective metric for unbounded settings in the general EOT problem, solely for marginals satisfying a light-tail condition. Though their result applies to general EOT problems for any regularising parameter $\varepsilon > 0$, still they can't cover the landmark example of quadratic cost with Gaussian marginals (as we do here as well for any regularising parameter $T > 0$).

Appendix 6

6.A Technical results

Lemma 6.A.1 (Exponential integrability of the marginals). *Let $\zeta \in \mathcal{P}_{\mathrm{alc}}(\mathbb{R}^d)$ associated with $U : \mathbb{R}^d \to \mathbb{R}$ satisfying* (6.0.10). *Then for any $\sigma \in (0, \alpha_U/2)$ it holds $\int \exp(\sigma|x|^2)\mathrm{d}\zeta(x) < +\infty$.*

Proof. It is enough noticing that for any $x \in \mathbb{R}^d$ it holds

$$\begin{aligned}\langle \nabla U(x),\, x\rangle = \langle \nabla U(x) - \nabla U(0),\, x\rangle + \langle \nabla U(0),\, x\rangle \geq \kappa_U(|x|)\,|x|^2 - |\nabla U(0)|\,|x| \\ \geq \alpha_U\,|x|^2 - (\tilde{g}_U(|x|) + |\nabla U(0)|)|x| \geq \alpha_U\,|x|^2 - \bar{G}_U\,|x|\end{aligned}$$

where above we have set $\bar{G}_U = \|\tilde{g}_U\|_\infty + |\nabla U(0)|$. Therefore for any $x \in \mathbb{R}^d$ it holds

$$\begin{aligned}U(x) =& U(0) + \int_0^1 \langle \nabla U(tx), x\rangle\,\mathrm{d}t \\ \geq& U(0) + \int_0^1 (\alpha_U\, t\,|x|^2 - \bar{G}_U\,|x|)\,\mathrm{d}t = U(0) + \frac{\alpha_U}{2}\,|x|^2 - \bar{G}_U\,|x|\ .\end{aligned}$$

Finally we may deduce for any $\sigma \in (0, \alpha_U/2)$

$$\begin{aligned}\|\exp(\sigma|x|^2)\|_{\mathrm{L}^1(U)} = \int_{\{|x|\leq R\}} \exp(\sigma|x|^2)\mathrm{d}\zeta(x) + \int_{\{|x|>R\}} \exp(\sigma|x|^2)\mathrm{d}\zeta(x) \\ \leq \mathrm{e}^{\sigma|R|^2}\, U(\{|x| \leq R\}) + \mathrm{e}^{-U(0)} \int_{\{|x|>R\}} \exp(-(\alpha_U/2 - \sigma)|x|^2 + \bar{G}_U|x|)\,\mathrm{d}x \\ \leq \mathrm{e}^{\sigma|R|^2}\, U(\{|x| \leq R\}) + \mathrm{e}^{-U(0)} \int_{\{|x|>R\}} \exp(-(\alpha_U/4 - \sigma/2)|x|^2)\,\mathrm{d}x < +\infty\ ,\end{aligned}$$

where above we have set $R \coloneqq 2\bar{G}_U\,(\alpha_U/2 - \sigma)^{-1}$. □

The following results proves the identities stated in (6.0.3).

Proposition 6.A.2. *Assume A6 holds. Then for any $n \in \mathbb{N}^*$, $h \in \{\psi^\star, \psi^n, \varphi^\star, \varphi^n\}$ it holds*

$$\begin{aligned}\nabla \log P_T \mathrm{e}^{-h}(x) &= T^{-1} \int (y - x)\, \pi_T^{x,h}(\mathrm{d}y)\ , \\ \nabla^2 \log P_T \mathrm{e}^{-h}(x) &= -\,T^{-1}\,\mathrm{Id} + T^{-2}\,\mathrm{Cov}(\pi_T^{x,h})\ .\end{aligned} \tag{6.A.1}$$

Proof. We will only prove the case $h = \psi^n$ since the other cases can be proven with the same argument. The proof will run as in [Con24, Proposition 5.2], once we have noticed that **A**6 and Lemma 6.A.1 guarantees the validity of $\exp(\sigma_\nu |x|^2) \in \mathrm{L}^1(\nu)$ for some positive $\sigma_\nu > 0$. Therefore from (2.2.18) we know that

$$\int_{\mathbb{R}^d\times\mathbb{R}^d} \exp\left(\sigma_\nu |y|^2 - \varphi^n(x) - \psi^n(y) - \frac{|x-y|^2}{2T}\right) \mathrm{d}x\,\mathrm{d}y = \int_{\mathbb{R}^d} \sigma_\nu |y|^2\,\mathrm{d}\nu(y) < +\infty,$$

and hence there exists at least one point $\bar{x} \in \mathbb{R}^d$ such that

$$\int_{\mathbb{R}^d} \exp\left(\sigma_\nu |y|^2 - \psi^n(y) - \frac{|\bar{x}-y|^2}{2T}\right) \mathrm{d}y < +\infty\,.$$

Since for any $x \in \mathbb{R}^d$ we can always write

$$|x-y|^2 = |\bar{x}-y|^2 - 2\langle x-\bar{x}, y\rangle + |x|^2 - |\bar{x}|^2 \geq |\bar{x}-y|^2 - 2|x-\bar{x}|\,|y| + |x|^2 - |\bar{x}|^2\,,$$

for any $\bar{\sigma} < \sigma_\nu$ we have

$$\int_{\mathbb{R}^d} \exp\left(\bar{\sigma}|y|^2 - \psi^n(y) - \frac{|\bar{x}-y|^2}{2T}\right) \mathrm{d}y < +\infty \quad \forall\, x \in \mathbb{R}^d\,. \tag{6.A.2}$$

This allows to differentiate under the integral sign in

$$\log P_T \exp(-\psi^n)(x) = -\frac{d}{2}\log(2\pi T) + \log \int \exp\left(-\psi^n(y) - \frac{|x-y|^2}{2T}\right) \mathrm{d}y$$

and get the validity of (6.A.1) for $h = \psi^n$, *i.e.*

$$\begin{aligned}\nabla \log P_T \exp(-\psi^n)(x) &= -\frac{x}{T} + \frac{1}{T}\,\frac{\int y \exp(-\psi^n(y) - \frac{|x-y|^2}{2T})\mathrm{d}y}{\int \exp(-\psi^n(y) - \frac{|x-y|^2}{2T})\mathrm{d}y}\\ &= T^{-1}\int (y-x)\,\pi_T^{x,\psi^n}(\mathrm{d}y)\,.\end{aligned}$$

The bound (6.A.2) guarantees to differentiate again the above integral and finally deduce our thesis. □

Finally, let us conclude by giving explicit upper-bounds for the fourth moments appearing in the proof of Theorem 6.5.2, under the geometric-drift condition (6.1.10) obtained in the proof of Corollary 6.1.3. More precisely if for any even $p \geq 2$ we set $V_p(y) = 1 + |y|^p$, similarly to what happened in (6.5.4), Corollary 6.1.3 and (6.1.10), the pointwise convergence of the gradients of Theorem 6.3.5 and the convergences $\alpha_{\mu,n} \uparrow \alpha_{\varphi^\star}$, $\alpha_{\nu,n} \uparrow \alpha_{\psi^\star}$ (stated in Theorem 6.1.1) imply the existence of constants $A_{\mu,p}, A_{\nu,p} > 0$ and $B_{\mu,p}, B_{\nu,p}$, independent of n (but depending on x and T), such that

$$\begin{aligned}\mathcal{L}_{\psi^\star} V_p(y) \;\vee\; \mathcal{L}_{\psi^n} V_p(y) \;&\leq B_{\nu,p} - A_{\nu,p}\, V_p(y)\,,\\ \mathcal{L}_{\varphi^\star} V_p(y) \;\vee\; \mathcal{L}_{\varphi^n} V_p(y) \;&\leq B_{\mu,p} - A_{\mu,p}\, V_p(y)\,,\end{aligned}$$

where $\mathcal{L}_h := \Delta/2 - \frac{1}{2}\langle T^{-1}(y-x) + \nabla h(y), \nabla\rangle$ is the generator associated to (6.0.4). We will bound the moments appearing in the proof of Theorem 6.5.2 in terms of the above constants $A_{\mu,p}, A_{\nu,p} > 0$ and $B_{\mu,p}, B_{\nu,p}$.

Lemma 6.A.3. *Take $p \geq 2$ and set $V_p(y) = 1 + |y|^p$. Let $(Y_t^\star)_{t\geq 0}$ and $(Z_t)_{t\geq 0}$ be defined as in* (6.5.11) *in the proof of Theorem 6.5.2. Recall that $Z_0 = Y_0^\star \sim \pi_T^{x,\psi^\star}$, $Y_t^\star \sim \pi_T^{x,\psi^\star} P_t^\star = \pi_T^{x,\psi^\star}$ whereas $Z_t \sim \pi_T^{x,\psi^\star} P_t^n$. Then for any $t \geq 0$ it holds*

$$\begin{aligned}\mathbb{E}[V_p(Y_t^\star)] =& \mathbb{E}[V_p(Y_0^\star)] \leq B_{\nu,p}/A_{\nu,p}\\ \mathbb{E}[V_p(Z_t)] \leq& (B_{\nu,p} + B_{\nu,p}/A_{\nu,p}) \exp(t\, A_{\nu,p})\ .\end{aligned}$$

Proof. By choosing $h = \psi^\star$ we immediately deduce that

$$\begin{aligned}\mathrm{d}V_p(Y_t^\star) =& \mathcal{L}_{\psi^\star} V_p(Y_t^\star)\,\mathrm{d}t + 4|Y_t^\star|^2\langle Y_t^\star, \mathrm{d}B_t\rangle\\ \leq& -A_{\nu,p}V_p(Y_t^\star)\,\mathrm{d}t + B_{\nu,p}\,\mathrm{d}t + p\,|Y_t^\star|^{p-2}\langle Y_t^\star, \mathrm{d}B_t\rangle\ .\end{aligned}$$

Therefore, up to considering a stopping time as already detailed in the proof of the contraction (6.5.6), by taking expectation and integrating over time, from the stationarity of the process $Y_t^\star \sim \pi_T^{x,\psi^\star}$ we deduce

$$\mathbb{E}[V_p(Y_t^\star)] \leq B_{\nu,p}/A_{\nu,p} \quad \forall\, t \geq 0\ . \tag{6.A.3}$$

Similarly, when considering $h = \psi^n$ we get

$$\begin{aligned}\mathrm{d}V_p(Z_t) =& \mathcal{L}_{\psi^\star} V_p(Z_t)\,\mathrm{d}t + 4|Z_t|^2\langle Z_t, \mathrm{d}B_t\rangle\\ \leq& -A_{\nu,p}V_p(Z_t)\,\mathrm{d}t + B_{\nu,p}\,\mathrm{d}t + p\,|Z_t|^{p-2}\langle Z_t, \mathrm{d}B_t\rangle\ .\end{aligned}$$

Up to considering again a stopping time, taking expectation and integrating over time yield to

$$\mathbb{E}[V_p(Z_t)] \leq \mathbb{E}[V_p(Z_0)] + B_{\nu,p}\,t - A_{\nu,p}\int_0^t \mathbb{E}[V_p(Z_s)]\,\mathrm{d}s\ .$$

By recalling that $Z_0 = Y_0^\star$, the previous bound (6.A.3), from Gronwall lemma we finally deduce

$$\mathbb{E}[V_p(Z_t)] \leq (B_{\nu,p}\ t + B_{\nu,p}/A_{\nu,p}) \exp(t\, A_{\nu,p}) \quad \forall\, t \geq 0\ .$$

□

Then, from (6.5.14), (6.3.15) and Lemma 6.A.3 (for $p = 2, 4$) we finally conclude that

Corollary 6.A.4. *Let $(Y_t^\star)_{t\geq 0}$ and $(Z_t)_{t\geq 0}$ be defined as in* (6.5.11) *in the proof of Theorem 6.5.2. Then* (6.5.14) *can be bounded as*

$$\int_0^t \mathbb{E}\left[(A|Y_s^\star| + B)(1 + \varepsilon V_2(Y_t^\star) + \varepsilon V_2(Z_t))\right]\mathrm{d}s \leq U(t, \nu, x, A, B, T)\ ,$$

with

$$
\begin{aligned}
U(t,\nu,x,A,B,T) \coloneqq{}& Bt\left(1+\varepsilon\frac{B_{\nu,2}}{A_{\nu,2}}(1+e^{t\,A_{\nu,2}})+\varepsilon B_{\nu,2}\,t\,e^{t\,A_{\nu,2}}\right)\\
&+\frac{A\,t}{\alpha_{\psi^\star}+T^{-1}}\left(T^{-1}|x|+1+\|\tilde{g}_\nu\|_\infty+|\nabla\psi^\star(0)|\right)\\
&+A\varepsilon\,t\left(\frac{B_{\nu,2}}{A_{\nu,2}}+\frac{B_{\nu,4}}{A_{\nu,4}}(1+e^{t\,A_{\nu,4}})+B_{\nu,4}\,t\,e^{t\,A_{\nu,4}}\right).
\end{aligned}
\tag{6.A.4}
$$

Chapter 7

The kinetic Schrödinger problem

In this chapter we investigate a different Schrödinger problem, known as the *kinetic Schrödinger problem*, hereafter KSP. Contrary to what done in the rest of the thesis where we have often focused our attention on the small-time asymptotics of SP, here we discuss this model with particular emphasis on the long-time and ergodic behaviour of the corresponding Schrödinger bridges.

A heuristic formulation of KSP can be given in terms of the thought experiment originally considered by Schrödinger, in the same fashion as we already portrayed in Chapter 1 with Brownian motions. Henceforth, consider a system of $N \gg 1$ independent stationary particles $(X^1_t, \dots, X^N_t)_{t\in[0,T]}$ evolving according to the (underdamped) Langevin dynamics

$$\begin{cases} \mathrm{d}X^i_t = V^i_t \mathrm{d}t, \\ \mathrm{d}V^i_t = -\nabla U(X^i_t)\mathrm{d}t - \gamma V^i_t \mathrm{d}t + \sqrt{2\gamma}\,\mathrm{d}B^i_t, \quad i = 1, \dots, N, \end{cases}$$

where X^i_t and V^i_t denote respectively the position and the velocity of the i^{th} particle at time t. Notice that a first difference with the classic setting stems from the fact that two parameters now encode the dynamics of the particle system, namely *position* X and *velocity* V. As a consequence in order to describe particles' trajectories we need both and therefore the space of continuous trajectories considered here is equal to $\Omega_{2d} := \mathcal{C}([0,T];\mathbb{R}^{2d})$.

The Schrödinger problem is that of finding the most likely evolution of the particle system conditionally on two snapshots of the particle system at the initial time $t = 0$ and at the terminal time $t = T$. In order to turn this heuristic description into a sound mathematical problem, we introduce again the empirical path measure as in (1.1.1), which this time is

$$\mathbf{P}^N := \frac{1}{N}\sum \delta_{(X^i_\cdot, V^i_\cdot)} \in \mathcal{P}(\Omega_{2d}) \,,$$

and consider two probability measures μ, ν on $\mathbb{R}^d$, representing the observed configurations at initial and final time, that is to say

$$\frac{1}{N}\sum_{i=1}^{N}\delta_{X_0^i} \approx \mu \quad \text{and} \quad \frac{1}{N}\sum_{i=1}^{N}\delta_{X_T^i} \approx \nu.$$

Notice that the above information is only a partial condition, if compared to (1.1.2) where the snapshots' condition was $\mathbf{P}_0^N \approx \mu$, $\mathbf{P}_T^N \approx \nu$. Still, we may again leverage Sanov's Theorem [DZ10, Theorem 6.2.10], whose message is that the likelihood of a given evolution P is measured through the relative entropy

$$\mathrm{Prob}\Big[\mathbf{P}^N \approx \mathrm{P}\Big] \approx \exp(-N\mathscr{H}(\mathrm{P}|\mathrm{R})),$$

which yields this time to the variational dynamic problem

$$\mathcal{C}_T(\mu,\nu) := \inf\Big\{\mathscr{H}(\mathrm{P}|\mathrm{R}) : \mathrm{P} \in \mathcal{P}(\mathcal{C}([0,T];\mathbb{R}^{2d})),\ (X_0)_\#\mathrm{P} = \mu, (X_T)_\#\mathrm{P} = \nu\Big\}. \tag{7.0.1}$$

In the above, R is the reference probability measure, that is the law on $\mathcal{P}(\Omega_{2d})$ of

$$\begin{cases} \mathrm{d}X_t = V_t\mathrm{d}t \\ \mathrm{d}V_t = -\nabla U(X_t)\mathrm{d}t - \gamma V_t\mathrm{d}t + \sqrt{2\gamma}\,\mathrm{d}B_t \\ (X_0, V_0) \sim \mathfrak{m}, \end{cases} \tag{7.0.2}$$

where the invariant (probability) measure $\mathfrak{m}$ is given by

$$\mathfrak{m}(\mathrm{d}x, \mathrm{d}v) = \frac{1}{Z}e^{-U(x)-\frac{|v|^2}{2}}\mathrm{d}x\,\mathrm{d}v,$$

with Z being a normalising constant. The term *kinetic* in KSP comes indeed from the above reference dynamics (7.0.2), since its probability density with respect to $\mathfrak{m}$ satisfies namely the kinetic Fokker-Planck equation

$$\partial_t f_t(x,v) = \gamma\,\Delta_v f_t(x,v) - \gamma\, v\cdot\nabla_v f_t(x,v) + \nabla U\cdot\nabla_v f_t(x,v) - v\cdot\nabla_x f_t(x,v). \tag{7.0.3}$$

Because of this, analysing KSP and its ergodic behaviour requires more attention: the hypocoercive [Vil09] nature of the kinetic Fokker-Planck equation makes more challenging quantifying its trend to equilibrium which reflects, as we will see later, in the long-time behaviour of Schrödinger bridges for KSP. In order to deal with that, we rely on the important progresses made in the study of the long-time behaviour of (7.0.3) over the last fifteen years using either an analytical approach see e.g. [Bau17, DMS15, HN04, Vil09] and references therein, or a probabilistic approach, see e.g. [EGZ19a, GLWZ21], as well as on the new developments around the long-time behaviour of Schrödinger bridges, in order to gain some understanding on *controlled* versions of the kinetic Fokker-Planck equation. We refer the reader to the bibliographical remarks section at the end of this chapter for a more accurate comparison between our results and the existing literature.

Besides the change of the reference measure, another difference between KSP and the classical instances of the Schrödinger problem (e.g., (1.1.4)) lies in the fact that it is not the full marginal that is constrained at initial and final time, but only its spatial component. Even though KSP seems to be a more faithful representation of Schrödinger's thought experiment, also the problem with fully constrained marginals

$$\mathcal{C}_T^F(\bar{\mu},\bar{\nu}) := \inf\left\{\mathscr{H}(\mathrm{P}|\mathrm{R}) : \mathrm{P}\in\mathcal{P}(\Omega),\ (X_0,V_0)_{\#}\mathrm{P}=\bar{\mu},\ (X_T,V_T)_{\#}\mathrm{P}=\bar{\nu}\right\}. \tag{7.0.4}$$

where $\bar{\mu},\bar{\nu}\in\mathcal{P}(\mathbb{R}^{2d})$ is worth studying and we shall work on both problems in the sequel. We will refer to the above problem as to the *Kinetic Full Schrödinger Problem*, hereafter KFSP. Through a classical argument, namely the same considered in Section 1.2, it is possible to reduce the dynamic formulations (cf. (7.0.1) and (7.0.4)) to static ones. For example, KSP is equivalent to solving

$$\inf\left\{\mathscr{H}(\pi|\mathrm{R}_{0,T}) : \pi\in\Pi_X(\mu,\nu)\right\}, \tag{7.0.5}$$

where $\mathrm{R}_{0,T} := ((X_0,V_0),(X_T,V_T))_{\#}\,\mathrm{R}$ is the joint law of R at initial and terminal time and the set $\Pi_X(\mu,\nu)$ is defined as

$$\Pi_X(\mu,\nu) := \left\{\pi\in\mathcal{P}\big(\mathbb{R}^{2d}\times\mathbb{R}^{2d}\big)\mid(\mathrm{proj}_{x_1})_{\#}\pi=\mu,\ (\mathrm{proj}_{x_2})_{\#}\pi=\nu\right\},$$

with $\mathrm{proj}_{x_i}\big((x_1,v_1),(x_2,v_2)\big) := x_i$ for any $i=1,2$. In a similar fashion, the static formulation of KFSP is

$$\inf\left\{\mathscr{H}(\pi|\mathrm{R}_{0,T}) : \pi\in\Pi(\bar{\mu},\bar{\nu})\right\}, \tag{7.0.6}$$

where $\Pi(\bar{\mu},\bar{\nu})$ is the (usual) set of couplings of $\bar{\mu}$ and $\bar{\nu}$.

It is worth noticing that, since the stationary Langevin dynamics is not a reversible measure, $\mathcal{C}_T^F(\cdot,\cdot)$ is not symmetric in its arguments. Nevertheless, due to the "physical reversibility" of the dynamics [CGP15], that is, reversibility up to a sign flip in the velocities, it is not hard to show that $\mathcal{C}_T(\cdot,\cdot)$ is symmetric in its arguments.

Lastly, let us add a bit of notation: if $\mathbf{P}^T$ is the unique solution of (7.0.1), we call *entropic interpolation* $(\mu_t^T)_{t\in[0,T]}$ the marginal flow of $\mathbf{P}^T$ and denote with ρ_t^T its density against $\mathfrak{m}$, *i.e.*,

$$\forall t\in[0,T],\quad \mu_t^T=(X_t,V_t)_{\#}\mathbf{P}^T,\qquad \rho_t^T := \frac{\mathrm{d}\mu_t^T}{\mathrm{d}\mathfrak{m}}.$$

With the obvious small modifications, we also define the entropic interpolation $(\bar{\mu}_t^T)_{t\in[0,T]}$ and their densities $(\bar{\rho}_t^T)_{t\in[0,T]}$ in the framework of KFSP.

7.1 Stochastic control formulation and turnpike property for kinetic Schrödinger bridges

KSP can be rephrased into a stochastic optimal control problem, as we already did for the classic problem in Section 1.3. Namely, Girsanov's Theorem [Léo12b, Theorems 2.1 and 2.3] this time implies the equivalence between (7.0.1) and

$$\inf\left\{\mathscr{H}((X_0,V_0)_{\#}\mathrm{P}|\mathfrak{m})+\frac{1}{4\gamma}\mathbb{E}_{\mathrm{P}}\Big[\int_0^T|\alpha_t^{\mathrm{P}}|^2\mathrm{d}t\Big]:\mathrm{P}\in\mathcal{P}(\Omega_{2d}),\ \mathrm{P}\,\text{admissible}\right\},\tag{7.1.1}$$

where a path probability measure P is admissible if and only if under P, there exist a Brownian motion $(B_t)_{t\in[0,T]}$ adapted to the canonical filtration and an adapted process $(\alpha_t^P)_{t\in[0,T]}$ such that $\mathbb{E}_{\mathrm{P}}[\int_0^T|\alpha^{\mathrm{P}}|^2\mathrm{d}t]<+\infty$ and the canonical process satisfies

$$\begin{cases}\mathrm{d}X_t=V_t\mathrm{d}t,\\ \mathrm{d}V_t=-\nabla U(X_t)\mathrm{d}t-\gamma V_t\mathrm{d}t+\alpha_t^P\mathrm{d}t+\sqrt{2\gamma}\mathrm{d}B_t,\\ X_0\sim\mu,X_T\sim\nu.\end{cases}\tag{7.1.2}$$

The same control formulation clearly holds for KFSP, with the corresponding full-marginals constraints at time 0 and T.

The above control formulation inspired the discussion of the present chapter, where we study the long-time behaviour of the controlled SDE (7.1.2) and establish the *turnpike property*. The latter is a general principle in optimal control theory stipulating that solutions of dynamic control problems are made of three pieces: first a rapid transition from the initial state to the steady state, the *turnpike*, then a long stationary phase localised around the turnpike, and finally another rapid transition to reach the final prescribed state. For the control problem (7.1.1), the turnpike is the invariant measure $\mathfrak{m}$. Indeed, the natural tendency of the particle system is that of reaching configuration $\mathfrak{m}$ and since Schrödinger bridges aim at approximating as much as possible the unconditional dynamics while matching the observed configurations, they should also favour configurations close to $\mathfrak{m}$. Obtaining a quantitative rigorous version of this statement is one of the main objectives of this chapter and, in view of (7.1.2), it is equivalent to show that Schrödinger bridges satisfy the turnpike property. In the field of deterministic control, the turnpike phenomenon is rather well understood both in a finite and infinite dimensional setting, see either [TZ15, TZZ18] and references therein, or the monographs [Zas05, Zas19]. The understanding of this phenomenon in stochastic control seems to be much more limited: see [CLLP12, CLLP13, CP19] for results on mean field games and [CCG22, BCGL20] for results on the classical and mean field Schrödinger problems. The reason why the turnpike property for Schrödinger bridges in the present context cannot be deduced from existing results lies in the hypocoercivity of the kinetic Fokker-Planck equation (7.0.3).

Proof strategy A general idea to obtain exponential speed of convergence to equilibrium for hypocoercive equations, systematically exploited in [Vil09], is that of modifying the *natural* Lyapunov function of the system by adding some extra terms in such a way that proving exponential dissipation becomes an easier task. For the Langevin dynamics, a suitable modification of the natural Lyapunov functional, that is the relative entropy $\mathscr{H}(\cdot|\mathfrak{m})$, is obtained considering

$$\mu \mapsto a\mathscr{H}(\mu|\mathfrak{m}) + \mathcal{I}(\mu)$$

for a carefully chosen constant $a > 0$, where the Fisher information (w.r.t. $\mathfrak{m}$) we recall to be defined for any $q \ll \mathfrak{m} \in \mathcal{P}(\mathbb{R}^{2d})$ as

$$\mathcal{I}(q) := \begin{cases} \int_{\mathbb{R}^{2d}} \left|\nabla \log \frac{\mathrm{d}q}{\mathrm{d}\mathfrak{m}}\right|^2 \mathrm{d}q & \text{if } \nabla \log \frac{\mathrm{d}q}{\mathrm{d}\mathfrak{m}} \in \mathrm{L}^2(q), \\ +\infty, & \text{otherwise.} \end{cases}$$

Emulating Bakry-Émery Γ-calculus [Bau17] it is possible to show that the modified Lyapunov functional decays exponentially fast along solutions of the kinetic Fokker-Planck equation. Our proof of the turnpike property consists in implementing this abstract idea on the fg-decomposition of the entropic interpolation, as we now briefly explain. Indeed, in order to bound $\mathcal{I}(\mu_t^T)$ one is naturally led to consider the quantities

$$\int_{\mathbb{R}^{2d}} \left|\nabla \log P_s^* f^T\right|^2 \mathrm{d}\mu_s^T, \tag{7.1.3a}$$

$$\int_{\mathbb{R}^{2d}} \left|\nabla \log P_{T-s} g^T\right|^2 \mathrm{d}\mu_s^T, \tag{7.1.3b}$$

where (f^T, g^T) is the fg-decomposition of KSP (cf. Proposition 7.2.3), $(P_s)_{s\in[0,T]}$ the semigroup associated to (7.0.2) and $(P_s^*)_{s\in[0,T]}$ its $\mathrm{L}^2(\mathfrak{m})$-adjoint. However, it is not clear how to obtain a differential inequality ensuring exponential (forward) dissipation of (7.1.3a) and exponential (backward) dissipation of (7.1.3b). But, as we show at Lemma 7.4.1, it is possible to find two norms $|\cdot|_{M^{-1}}$ and $|\cdot|_{N^{-1}}$, that are equivalent to the Euclidean norm and such that if we define

$$\varphi^T(s) := \int_{\mathbb{R}^{2d}} \left|\nabla \log P_s^* f^T\right|_{N^{-1}}^2 \mathrm{d}\mu_s^T,$$
$$\psi^T(s) := \int_{\mathbb{R}^{2d}} \left|\nabla \log P_{T-s} g^T\right|_{M^{-1}}^2 \mathrm{d}\mu_s^T,$$

then $\varphi^T(s)$ and $\psi^T(s)$ satisfy the desired exponential estimates. To complete the proof, one needs to take care of the boundary conditions. This part is non trivial as it demands to prove certain regularity properties of the fg-decomposition and it is accomplished in two steps: we first show in Proposition 7.4.3 a regularising property of entropic interpolations, namely that if $\mathscr{H}(\mu|\mathfrak{m}_X), \mathscr{H}(\nu|\mathfrak{m}_X)$ are finite, then the Fisher information $\mathcal{I}(\mu_t^T)$ is finite for any $t \in (0, T)$. The proof of this property is based on a gradient bound obtained in [GW12] and is of independent interest. The second step (Proposition 7.4.2) consists in showing that for a fixed small δ, $\varphi^T(\delta)$ and $\psi^T(T-\delta)$ can be controlled with by the sum of $\mathcal{I}(\mu^T_\delta)$ and $\mathcal{I}(\mu^T_{T-\delta})$. We prove this estimate adapting an argument used in [TZ15] in the analysis of deterministic finite dimensional control problems.

7.2 Assumptions and preliminaries

In this section we collect our assumptions and useful results about the Markov semigroup associated to the kinetic Fokker-Planck equation. We conclude with structural results for KSP (and KFSP).

In what follows we write $\lesssim$ to indicate that an inequality holds up to a multiplicative positive constant depending possibly on the dimension d, the bounds on the spectrum of U, α and β, or the friction parameter γ.

7.2.1 On the assumptions

We state here the assumption on the potential U and on the constraints μ, ν, $\bar\mu$ and $\bar\nu$ that we use in the sequel. We define $\mathfrak{m}_X, \mathfrak{m}_V \in \mathcal{P}(\mathbb{R}^d)$ to be the respectively the space and velocity marginals of $\mathfrak{m}$, in particular $\mathfrak{m} = \mathfrak{m}_X \otimes \mathfrak{m}_V$.

(H1) U is a C^∞ strongly convex potential with bounded derivatives of order $k \geq 2$.

(H2) There exist $0 < \alpha < \beta$ such that

$$\sqrt{\beta} - \sqrt{\alpha} \leq \gamma\,, \qquad \text{and} \qquad \alpha \mathrm{Id}_d \leq \nabla^2 U(x) \leq \beta \mathrm{Id}_d\,, \quad \text{for all } x \in \mathbb{R}^d,$$

where $\gamma > 0$ is the friction parameter in (7.0.2).

(H3) The probability measures μ and ν on $\mathbb{R}^d$ satisfy

$$\mathscr{H}(\mu|\mathfrak{m}_X) < +\infty \quad \text{and} \quad \mathscr{H}(\nu|\mathfrak{m}_X) < +\infty\,.$$

(H4) $\mu, \nu \ll \mathfrak{m}_X$, $\frac{\mathrm{d}\mu}{\mathrm{d}\mathfrak{m}_X}, \frac{\mathrm{d}\nu}{\mathrm{d}\mathfrak{m}_X} \in \mathrm{L}^\infty(\mathfrak{m}_X)$ and are compactly supported on $\mathbb{R}^d$.

(FH3) The probability measures $\bar\mu$ and $\bar\nu$ on $\mathbb{R}^{2d}$ satisfy

$$\mathscr{H}(\bar\mu|\mathfrak{m}) < +\infty \quad \text{and} \quad \mathscr{H}(\bar\nu|\mathfrak{m}) < +\infty\,.$$

(FH4) $\bar\mu, \bar\nu \ll \mathfrak{m}$, $\frac{\mathrm{d}\bar\mu}{\mathrm{d}\mathfrak{m}}, \frac{\mathrm{d}\bar\nu}{\mathrm{d}\mathfrak{m}} \in \mathrm{L}^\infty(\mathfrak{m})$ and are compactly supported on $\mathbb{R}^{2d}$.

In what follows we report some straightforward consequences of the various assumptions listed above that we shall repeatedly use from now on. We begin by observing that assumption (H1) guarantees that $\mathfrak{m} \in \mathcal{P}_2(\mathbb{R}^{2d})$ and that

$\mathfrak{m}_X$ satisfies Talagrand's inequality because of [BGL13, Corollary 9.3.2], *i.e.*, for any $q \in \mathcal{P}(\mathbb{R}^d)$

$$\mathbf{W}_2(q,\, \mathfrak{m}_X)^2 \lesssim \mathscr{H}(q|\mathfrak{m}_X)\,. \tag{7.2.1}$$

Since the Talagrand inequality holds also for the Gaussian measure $\mathfrak{m}_V$, from [BGL13, Proposition 9.2.4] it follows that for any $q \in \mathcal{P}(\mathbb{R}^{2d})$

$$\mathbf{W}_2(q,\, \mathfrak{m})^2 \lesssim \mathscr{H}(q|\mathfrak{m})\,. \tag{7.2.2}$$

Let us also point out that (H4) implies (H3) and that under (H1) and (H3) it easily follows that $\mu,\, \nu \in \mathcal{P}_2(\mathbb{R}^d)$. Indeed,

$$\int_{\mathbb{R}^d} |x|^2\, \mathrm{d}\mu \lesssim \int_{\mathbb{R}^d} |x|^2\, \mathrm{d}\mathfrak{m}_X + \mathbf{W}_2(\mu, \mathfrak{m}_X)^2 \overset{(7.2.1)}{\lesssim} \int_{\mathbb{R}^d} |x|^2\, \mathrm{d}\mathfrak{m}_X + \mathscr{H}(\mu|\mathfrak{m}_X) < +\infty, \tag{7.2.3}$$

and similarly for the measure ν. We also remark that (FH4) implies (FH3). Moreover, from (H1) and (FH3), by means of (7.2.2), it follows that $\bar{\mu},\, \bar{\nu} \in \mathcal{P}_2(\mathbb{R}^{2d})$.

Finally, let us also notice that (H1) and (H2) guarantee the validity of a log-Sobolev inequality for $\mathfrak{m}_X$ because of [BGL13, Corollary 5.7.2], and by means of [BGL13, Proposition 5.2.7 and Proposition 5.5.1] it follows that $\mathfrak{m}$ satisfies a log-Sobolev inequality. Therefore for any $q \ll \mathfrak{m}$ it holds

$$\mathscr{H}(q|\mathfrak{m}) \lesssim \mathcal{I}(q)\,. \tag{7.2.4}$$

7.2.2 Markov semigroups and heat kernel

The generator L associated to the SDE (7.0.2) is given by

$$L = \gamma\Delta_v - \gamma v \cdot \nabla_v - \nabla U \cdot \nabla_v + v \cdot \nabla_x$$

while its adjoint in $\mathrm{L}^2(\mathfrak{m})$ reads as

$$L^* = \gamma\Delta_v - \gamma v \cdot \nabla_v + \nabla U \cdot \nabla_v - v \cdot \nabla_x\,.$$

Under (H1), it is well known that Hörmander's Theorem for parabolic hypoellipticity applies [Hör67, Theorem 1.1] to the operator L , and thus the associated semigroup $(P_t)_{t\geq 0}$ admits a probability kernel $\mathsf{p}_t((x,y),(y,w))$, which is $\mathcal{C}^\infty$ in all of the parameters, with respect to the invariant probability measure

$$\mathrm{d}\mathfrak{m}(x,v) = \frac{1}{Z} e^{-U(x) - \frac{|v|^2}{2}}\, \mathrm{d}x\mathrm{d}v,$$

where Z is a normalising constant. Sometimes, with a slight abuse of notation we will write $\mathfrak{m}(x,v)$ to denote the density of $\mathfrak{m}$ with respect to the Lebesgue measure. Similarly, we will denote by $(P_t^*)_{t\geq 0}$ the semigroup associated to L^*. Note that the function p_t also represents the density of $\mathrm{R}_{0,t}$ (the joint law at time 0 and t of the solution to (7.0.2)) with respect to $\mathfrak{m} \otimes \mathfrak{m}$. Moreover, according to [DM10, Theorem 1.1], $\mathsf{p}_t(\cdot,\cdot)$ satisfies two-sided Gaussian estimates. Importantly, p_t is locally bounded away from zero and infinity, but with constants that might depend non-trivially on the time horizon T.

For some of our proofs, we need lower bounds that are uniform in T. To this aim, we have the following consequence of the results of [DM10].

Lemma 7.2.1. *Let $T_0 > 0$ be fixed. Under assumption (H1), there exists a constant $c_{T_0} > 0$ such that for all $T \geq T_0$ and all (x, v), $(y, w) \in \mathbb{R}^{2d}$*

$$\log \mathsf{p}_T\left((x,v),\,(y,w)\right) \geq -c_{T_0}\Big(1+|x|^2+|v|^2+|y|^2+|w|^2\Big)\,. \tag{7.2.5}$$

Proof. Let $T_0 > 0$ be fixed. From Jensen's inequality we know that

$$\begin{aligned}
&\log \mathsf{p}_T((x,v),\,(y,w)) \\
&\quad= \log \int_{\mathbb{R}^{2d}} \mathsf{p}_{T-T_0/2}\left((x,v),\,(z,u)\right)\, \mathsf{p}_{T_0/2}\left((z,u),\,(y,w)\right) \mathrm{d}\mathfrak{m}(z,u) \\
&\quad\geq \int_{\mathbb{R}^{2d}} \log \mathsf{p}_{T-T_0/2}\left((x,v),\,(z,u)\right) \mathrm{d}\mathfrak{m}(z,u) \\
&\qquad+ \int_{\mathbb{R}^{2d}} \log \mathsf{p}_{T_0/2}\left((z,u),\,(y,w)\right) \mathrm{d}\mathfrak{m}(z,u)\,.
\end{aligned} \tag{7.2.6}$$

By [DM10, Theorem 1.1], there exists $C \geq 1$ depending on T_0 such that

$$\mathsf{p}_{T_0/2}\left((z,u),\,(y,w)\right) \gtrsim \mathsf{p}_{T_0/2}((z,u),(y,w))\mathfrak{m}(y,w) \geq C^{-1}\, e^{-C|\theta_{T_0/2}(z,u)-(y,w)^T|^2}$$

where $\theta_t(x_0, v_0) = (\theta_t^x,\, \theta_t^v)^T$ denotes the solution of the denoised Langevin ODE system

$$\begin{cases} \frac{\mathrm{d}}{\mathrm{d}t}\theta_t^x = \theta_t^v \\ \frac{\mathrm{d}}{\mathrm{d}t}\theta_t^v = -\theta_t^v - \nabla U(\theta_t^x) \end{cases} \qquad \text{with} \quad \theta_0 = (x_0,\, v_0)^T\,.$$

Since under (H1) there exists a large enough positive $r \in \mathbb{R}$ such that (7.2.11) holds for $(\mathrm{Id}, -r)$, from [Mon23, Theorem 1] (with $\Sigma = 0$) it follows

$$|\theta_t(y,w) - \theta_t(0,0)| \leq e^{r\,t}\left|(y,w)^T\right|\,, \quad \forall\,(y,w)^T \in \mathbb{R}^{2d},\ \forall t \geq 0\,. \tag{7.2.7}$$

Therefore, up to changing the constants C from line to line, we have

$$\begin{aligned}
&\int_{\mathbb{R}^{2d}} \log \mathsf{p}_{T_0/2}\left((z,u),\,(y,w)\right) \mathrm{d}\mathfrak{m}(z,u) \\
&\quad\geq \log C^{-1} - C\int_{\mathbb{R}^{2d}} \left|\theta_{T_0/2}(z,u) - (y,w)^T\right|^2 \mathrm{d}\mathfrak{m}(z,u) \\
&\quad\geq -C\Big(1+|y|^2+|w|^2+\int_{\mathbb{R}^{2d}} \left|\theta_{T_0/2}(z,u)\right|^2 \mathrm{d}\mathfrak{m}(z,u)\Big) \\
&\quad\geq -C\Big(1+|y|^2+|w|^2\Big)\,,
\end{aligned} \tag{7.2.8}$$

where the last step holds since $\mathfrak{m} \in \mathcal{P}_2(\mathbb{R}^{2d})$, and therefore

$$\begin{aligned}
&\int_{\mathbb{R}^{2d}} |\theta_{T_0/2}(z,u)|^2 \mathrm{d}\mathfrak{m}(z,u) \\
&\qquad \leq 2\left|\theta_{T_0/2}(0,0)\right|^2 + 2\int_{\mathbb{R}^{2d}} \left|\theta_{T_0/2}(z,u) - \theta_{T_0/2}(0,0)\right|^2 \mathrm{d}\mathfrak{m}(z,u) \\
&\qquad \overset{(7.2.7)}{\leq} 2\left|\theta_{T_0/2}(0,0)\right|^2 + 2e^{2r}\int_{\mathbb{R}^{2d}} \left(|z|^2 + |u|^2\right) \mathrm{d}\mathfrak{m}(z,u) \leq C\,.
\end{aligned}$$

Now, notice that we can rewrite the first integral of the RHS in (7.2.6) as

$$\begin{aligned}
&\int_{\mathbb{R}^{2d}} \log \mathsf{p}_{T-T_0/2}\left((x,v),\,(z,u)\right) \mathrm{d}\mathfrak{m}(z,u) \\
&= \int_{\mathbb{R}^{2d}} \log\left[\int_{\mathbb{R}^{2d}} \mathsf{p}_{T_0/2}\left((x,v),\,(q,r)\right)\, \mathsf{p}_{T-T_0}\left((q,r),\,(z,u)\right) \mathrm{d}\mathfrak{m}(q,r)\right] \mathrm{d}\mathfrak{m}(z,u).
\end{aligned}$$

Because of (7.2.9), we know that $\mathsf{p}_{T-T_0/2}\left((q,r),\,(z,u)\right)\mathrm{d}\mathfrak{m}(q,r)$ is a probability measure over $\mathbb{R}^{2d}$ and therefore by Jensen's inequality and Fubini the above displacement can be lower bounded by

$$\begin{aligned}
&\int_{\mathbb{R}^{2d}}\int_{\mathbb{R}^{2d}} \log\left[\mathsf{p}_{T_0/2}\left((x,v),\,(q,r)\right)\right]\, \mathsf{p}_{T-T_0}\left((q,r),\,(z,u)\right) \mathrm{d}\mathfrak{m}(q,r)\mathrm{d}\mathfrak{m}(z,u) \\
&\quad = \int_{\mathbb{R}^{2d}} \log \mathsf{p}_{T_0/2}\left((x,v),\,(q,r)\right) \mathrm{d}\mathfrak{m}(q,r) \\
&\quad \overset{(7.2.9)}{=} \int_{\mathbb{R}^{2d}} \log \mathsf{p}_{T_0/2}\left((q,-r),\,(x,-v)\right) \mathrm{d}\mathfrak{m}(q,r) \\
&\quad = \int_{\mathbb{R}^{2d}} \log \mathsf{p}_{T_0/2}\left((q,r),\,(x,-v)\right) \mathrm{d}\mathfrak{m}(q,r) \overset{(7.2.8)}{\geq} -C\left(1+|x|^2+|v|^2\right).
\end{aligned}$$

Putting the above lower bound and (7.2.8) into inequality (7.2.6), we get

$$\log \mathsf{p}_T\left((x,v),\,(y,w)\right) \geq -c_{T_0}\left(1+|x|^2+|v|^2+|y|^2+|w|^2\right).$$

□

Finally, let us stress that the Langevin dynamics (7.0.2) is not reversible, and in particular the probability kernel p_t is not symmetric. However, it is symmetric up to a sign-flip in the velocities,

$$\mathsf{p}_t\big((x,v),(y,w)\big) = \mathsf{p}_t\big((y,-w),(x,-v)\big) \quad \forall t \geq 0, \quad \forall (x,v),\,(y,w) \in \mathbb{R}^{2d}\,. \tag{7.2.9}$$

As we said above, this useful property is sometimes called physical reversibility.

7.2.3 Contraction of the semigroup

In our setup, due to the lack of a curvature condition CD, the standard Bakry-Emery machinery does not apply to obtain a commutation estimate for the semigroup of the type

$$|\nabla P_t h(z)| \le e^{-c\,t}\, P_t\big(\,|\nabla h|\,\big)(z)\,, \tag{7.2.10}$$

for some $c > 0$. It is still possible to obtain a commutation estimate similar to (7.2.10) by replacing the Euclidean norm $|\cdot|$ by a certain twisted norm $|\xi|_M := \sqrt{\xi \cdot M\xi}$ on $\mathbb{R}^{2d}$ for some well chosen positive definite symmetric matrix $M \in \mathbb{R}^{2d\times 2d}$. This is a common idea in the kinetic setting and it is exploited for example in [Bau17, GLWZ21, Mon23]. Particularly, Assumption (H2) implies *local* gradient contraction bounds for the semigroup of the Langevin dynamics with a certain rate $\kappa > 0$ (see Proposition 7.2.2 or [Bau17]). The exponential rate κ of Theorems 7.5.1 and 7.5.3 below is precisely the one, computed e.g. in [Mon23, BGM10], at which the synchronous coupling is contractive for the (uncontrolled) Langevin dynamics.

For instance, in [Mon23, Theorem 1] the author studies the contractive properties of the semigroup P_t associated to the SDE on $\mathbb{R}^m$

$$\mathrm{d}Z_t = b(Z_t)\mathrm{d}t + \Sigma \mathrm{d}B_t\,,$$

with the drift $b : \mathbb{R}^m \to \mathbb{R}^m$ being globally Lipschitz and Σ a constant positive-semidefinite symmetric matrix. The author shows that the condition on the Jacobian matrix J_b of the drift

$$\xi \cdot (MJ_b(z))\xi \le -\kappa\, \xi \cdot M\xi = -\kappa\, |\xi|_M^2 \qquad \forall \xi \in \mathbb{R}^m,\, \forall z \in \mathbb{R}^m\,, \tag{7.2.11}$$

where $\kappa \in \mathbb{R}$ and M is a positive definite symmetric matrix, is equivalent to the commutation estimate

$$|\nabla P_t h(z)|_{M^{-1}} \le e^{-\kappa\, t}\, P_t\big(\,|\nabla h|_{M^{-1}}\,\big)(z)\,.$$

Our setup, which is also discussed in [Mon23, Section 3.3], corresponds to the choice $m = 2d$, and

$$b(x,v) = \begin{pmatrix} v \\ -\nabla U(x) - \gamma v \end{pmatrix} \qquad \Sigma = \begin{pmatrix} 0 & 0 \\ 0 & \sqrt{2\gamma}\mathrm{Id} \end{pmatrix}$$

and therefore the Jacobian reads as

$$J_b(x,v) = \begin{pmatrix} 0 & \mathrm{Id} \\ -\nabla^2 U(x) & -\gamma \mathrm{Id} \end{pmatrix}\,.$$

In [Mon23, Proposition 5], the author shows that (7.2.11) holds with $\kappa > 0$ as long as α and β from assumption (H2) are close enough. The slightly sharper condition $\sqrt{\beta} - \sqrt{\alpha} \leq \gamma$ (which corresponds to (H2)) can be obtained by mimicking the computations in [Bau17, Theorem 2.12], where the case $\gamma = 1$ is discussed. By exploiting the symmetry of the heat kernel up to a sign flip, we obtain a similar commutation estimate also for the reversed dynamics.

In view of the above discussion we have the following.

Proposition 7.2.2. *Assume that (H1) and (H2) hold. Then, there exist a constant $\kappa > 0$ and positive definite symmetric matrices $M, N \in \mathbb{R}^{2d \times 2d}$ such that* (7.2.11) *holds and*

(i) For all $h \in C_c^1(\mathbb{R}^{2d})$, $t \geq 0$ and $z \in \mathbb{R}^{2d}$

$$|\nabla P_t h(z)|_{M^{-1}} \leq e^{-\kappa t} P_t\big(|\nabla h|_{M^{-1}} \big)(z) . \tag{7.2.12}$$

(ii) For all $h \in C_c^1(\mathbb{R}^{2d})$, $t \geq 0$ and $z \in \mathbb{R}^{2d}$

$$|\nabla P_t^* h(z)|_{N^{-1}} \leq e^{-\kappa t} P_t^*\big(|\nabla h|_{N^{-1}} \big)(z) . \tag{7.2.13}$$

Proof. A proof that (7.2.11) holds with $\kappa > 0$ under (H1) and (H2) is included in Section 7.B for the reader's convenience. Given (7.2.11), (i) follows from Theorem 1 in [Mon23].

We now derive (ii) from (i) with the help of (7.2.9). For any function f on $\mathbb{R}^{2d}$ define the transformation $\mathcal{S}f(x,v) = f(x,-v)$ and set

$$N = \begin{pmatrix} \mathrm{Id} & 0 \\ 0 & -\mathrm{Id} \end{pmatrix} M \begin{pmatrix} \mathrm{Id} & 0 \\ 0 & -\mathrm{Id} \end{pmatrix} .$$

Note that $\mathcal{S}^2 = \mathrm{Id}$, moreover in view of (7.2.9), for all $h \in C_c^1(\mathbb{R}^{2d})$, $P_t^*(\mathcal{S}h) = \mathcal{S}(P_t h)$ and $\mathcal{S}|\nabla h|_{M^{-1}} = |\nabla(\mathcal{S}h)|_{N^{-1}}$. It is then immediate to derive

$$\begin{aligned} |\nabla P_t^* h|_{N^{-1}} = \mathcal{S}\, |\nabla P_t(\mathcal{S}h)|_{M^{-1}} &\leq e^{-\kappa t}\, \mathcal{S}\Big(P_t\big(|\nabla(\mathcal{S}h)|_{M^{-1}} \big)\Big) \\ &= e^{-\kappa t} P_t^*\big(|\nabla h|_{N^{-1}} \big) \end{aligned}$$

which is the desired conclusion. □

As a result of Proposition 7.2.2 and [Mon23, Theorem 1] we have the equivalent statements, with M, N and $\kappa > 0$ as above, and all $q_1, q_2 \in \mathcal{P}(\mathbb{R}^{2d})$,

$$\mathbf{W}_{M,2}(q_1 P_t, q_2 P_t) \leq e^{-\kappa t}\, \mathbf{W}_{M,2}(q_1, q_2) , \tag{7.2.14}$$

$$\mathbf{W}_{N,2}(q_1 P_t^*, q_2 P_t^*) \leq e^{-\kappa t} \mathbf{W}_{N,2}(q_1, q_2) ,$$

where $\mathbf{W}_{M,2}(q_1, q_2)$ is the $\mathbf{W}_2$-Wasserstein distance on $\mathcal{P}(\mathbb{R}^{2d})$ with the Euclidean metric replaced by $d_M(x,y) = |x-y|_M$ and similarly for $\mathbf{W}_{N,2}(q_1, q_2)$.

7.2.4 The fg-decomposition for KSP

Optimal couplings in the Schrödinger problem are characterised by the fact that their density against the reference measure takes a product form, often called fg-decomposition, as we have established in Theorem 2.2.1. In KSP f and g have the additional property of depending only on the first and second space variables respectively.

Proposition 7.2.3. *Grant (H1), (H3). Then, for all $T > 0$,* (7.0.5) *and* (7.0.1) *admit unique solutions $\boldsymbol{\pi}^T, \mathbf{P}^T$ with $\boldsymbol{\pi}^T = ((X_0, V_0), (X_T, V_T))_{\#}\mathbf{P}^T$ and there exist two non-negative measurable functions f^T, g^T on $\mathbb{R}^d$ such that*

$$\rho^T(x,v,y,w) := \frac{\mathrm{d}\boldsymbol{\pi}^T}{\mathrm{d}\mathrm{R}_{0,T}}(x,v,y,w) = f^T(x)g^T(y), \qquad \mathrm{R}_{0,T}\text{-a.s.} \tag{7.2.15}$$

Moreover, f^T, g^T solve the Schrödinger system:

$$\begin{cases} \frac{\mathrm{d}\mu}{\mathrm{d}\mathfrak{m}_X}(x) = f^T(x)\, \mathbb{E}_{\mathrm{R}}\big[g^T(X_T)|X_0 = x\big], \\ \frac{\mathrm{d}\nu}{\mathrm{d}\mathfrak{m}_X}(y) = g^T(y)\, \mathbb{E}_{\mathrm{R}}\big[f^T(X_0)|X_T = y\big]. \end{cases} \tag{7.2.16}$$

A similar structure result holds for KFSP. Particularly the corresponding fg-decomposition can be deduced from Theorem 2.2.1 (cf. Section 7.2.5). On the contrary, the case KSP requires some extra work. We remark here that for both dual representation of the cost (cf. Proposition 7.2.4 later) and the fg-decomposition the strict convexity of U and its smoothness are not really necessary. A bounded Hessian would suffice. Nevertheless, since the convex case is the one we will be interested in later on, we prefer not to insist on this point.

Proof of Proposition 7.2.3. We only sketch the proof as it is rather standard. We consider the measure $\mathrm{R}^X_{0,T} := (\mathrm{proj}_{x_1}, \mathrm{proj}_{x_2})_{\#}\mathrm{R}_{0,T} = (X_0, X_T)_{\#}\mathrm{R}$ and the minimisation problem,

$$\min_{q\in\Pi(\mu.\nu)} \mathscr{H}\left(q|\mathrm{R}^X_{0,T}\right), \tag{7.2.17}$$

where $\Pi(\mu.\nu)$ is the set of couplings of $\mu, \nu \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$. In view of the heat kernel lower bound in Lemma 7.2.1, we know that for some $C > 0$ and uniformly in x, y it holds

$$\frac{\mathrm{d}\mathrm{R}^X_{0,T}}{\mathrm{d}(\mathfrak{m}_X \otimes \mathfrak{m}_X)}(x,y) \geq \frac{1}{C}e^{-C\left(1+|x|^2+|y|^2\right)}.$$

which in combination with (H3) implies that $\mathscr{H}(\mu\otimes\nu|\mathrm{R}^X_{0,T}) < \infty$. Indeed the bound above implies that for any $T > T_0$, T_0 fixed, there is a constant $C_{d,\alpha,\beta,\gamma,T_0} > 0$ such that

$$\begin{aligned} \mathscr{H}(\mu\otimes\nu|\mathrm{R}^X_{0,T}) &= \mathscr{H}(\mu\otimes\nu|\mathfrak{m}_X\otimes\mathfrak{m}_X) - \int_{\mathbb{R}^{4d}} \log \frac{\mathrm{d}\mathrm{R}^X_{0,T}}{\mathrm{d}(\mathfrak{m}_X\otimes\mathfrak{m}_X)}\, \mathrm{d}\mu\otimes\nu \\ &\overset{(7.2.3)}{\leq} C_{d,\alpha,\beta,\gamma,T_0}\left[1 + \mathscr{H}(\mu|\mathfrak{m}_X) + \mathscr{H}(\nu|\mathfrak{m}_X)\right]. \end{aligned} \tag{7.2.18}$$

Thus, Proposition 2.5 in [Léo14] applies and the above minimisation problem has indeed a unique solution $\pi \in \Pi(\mu, \nu)$. By applying [RT93, Theorem 3], there exist two non-negative measurable functions f^T, g^T on $\mathbb{R}^d$ such that

$$\frac{\mathrm{d}\pi}{\mathrm{dR}^X_{0,T}}(x, y) = f^T(x) g^T(y), \qquad \mathrm{R}^X_{0,T}\text{-a.s.},$$

from which (7.2.16) directly follows. Now, in view of the additive property of the relative entropy (1.A.4), we get for any $\mathrm{P} \in \mathcal{P}(\Omega_{2d})$

$$\mathscr{H}(\mathrm{P}|\mathrm{R}) = \mathscr{H}\left(\mathrm{P}^X_{0,T}|\mathrm{R}^X_{0,T}\right) + \int_{\mathbb{R}^{2d}} \mathscr{H}\left(\mathrm{P}^{x,y}|\mathrm{R}^{x,y}\right) \mathrm{dP}^X_{0,T}(x,y),$$

with $\mathrm{R}^{x,y} = \mathrm{R}(\,\cdot\,|X_0 = x, X_T = y)$ and similarly for $\mathrm{P}^{x,y}$. Therefore, a minimiser to (7.0.1) can be found by defining

$$\mathbf{P}^T(\cdot) = \int_{\mathbb{R}^d \times \mathbb{R}^d} \mathrm{R}(\,\cdot\,|X_0 = x, X_T = y)\, \mathrm{d}\pi(x, y),$$

which satisfies $(X_0, X_T)_{\#}\mathbf{P}^T = \pi$ and $\mathcal{C}_T(\mu,\nu) = \mathscr{H}(\mathbf{P}^T|\mathrm{R}) = \mathscr{H}(\pi|\mathrm{R}^X_{0,T}) < \infty$. In particular, in view of (7.2.18), for all $T > T_0$, T_0 fixed, there is $C_{d,\alpha,\beta,\gamma,T_0} > 0$ such that

$$\mathcal{C}_T(\mu,\nu) \leq C_{d,\alpha,\beta,\gamma,T_0}\left[1 + \mathscr{H}(\mu|\mathfrak{m}_X) + \mathscr{H}(\nu|\mathfrak{m}_X)\right]. \tag{7.2.19}$$

Similarly, for any $q \in \Pi_X(\mu,\nu)$, denoting $q^X = (\mathrm{proj}_{x_1}, \mathrm{proj}_{x_2})_{\#}q$, we have $\mathscr{H}(q|\mathrm{R}_{0,T}) \geq \mathscr{H}(q^X|\mathrm{R}^X_{0,T}) \geq \mathscr{H}(\pi|\mathrm{R}^X_{0,T})$ with equality if and only if $q = \boldsymbol{\pi}^T$ where

$$\boldsymbol{\pi}^T(\cdot) = \int_{\mathbb{R}^d \times \mathbb{R}^d} \mathrm{R}_{0,T}(\,\cdot\,|X_0 = x, X_T = y)\, \mathrm{d}\pi(x, y). \tag{7.2.20}$$

By construction $\boldsymbol{\pi}^T = ((X_0, V_0), (X_T, V_T))_{\#}\mathbf{P}^T$ and $\mathscr{H}(\mathbf{P}^T|\mathrm{R}) = \mathscr{H}(\boldsymbol{\pi}^T|\mathrm{R}_{0,T}) = \mathscr{H}(\pi|\mathrm{R}^X_{0,T}) < \infty$. The solutions are unique by strict convexity of the entropy and the linearity of the constraint. Equation (7.2.20) implies equality of the conditional distributions of $\boldsymbol{\pi}^T$ and $R_{0,T}$ given the space variables. But then, $\mathrm{R}_{0,T}$-a.s. it holds

$$\frac{\mathrm{d}\boldsymbol{\pi}^T}{\mathrm{dR}_{0,T}}(x,v,y,w) = \frac{\mathrm{d}(\mathrm{proj}_{x_1}, \mathrm{proj}_{x_2})_{\#}\boldsymbol{\pi}^T}{\mathrm{d}(\mathrm{proj}_{x_1}, \mathrm{proj}_{x_2})_{\#}\mathrm{R}_{0,T}}(x,y) = \frac{\mathrm{d}\pi}{\mathrm{dR}^X_{0,T}}(x,y) = f^T(x)g^T(y).$$

□

As we did for the classic problem we can define the kinetic Schrödinger potentials as the measurable function $\varphi^T := -\log f^T \in \mathrm{L}^1(\mu)$ and $\psi^T := -\log g^T \in \mathrm{L}^1(\nu)$ and then clearly it holds

$$\frac{\mathrm{d}\boldsymbol{\pi}^T}{\mathrm{dR}_{0,T}}((x,v),(y,w)) = \exp(-\varphi^T(x) - \psi^T(y)) \quad \mathrm{R}_{0,T}\text{-a.e.}\,.$$

Moreover, as already noticed in the classic setting (cf. Proposition 2.2.2) these potentials (actually their opposites $(-\varphi^T)$ and $(-\psi^T)$) are optimiser in a duality representation of the entropic cost, analogous to the Monge-Kantorovich duality (1.2.9).

Proposition 7.2.4. *Grant (H1) and (H3). Then $\mathcal{C}_T(\mu,\nu) < \infty$ and*

$$\mathcal{C}_T(\mu,\nu) = \sup_{\alpha,\beta\in\mathcal{M}_b(\mathbb{R}^d)} \left\{ \int_{\mathbb{R}^d} \alpha \,\mathrm{d}\mu + \int_{\mathbb{R}^d} \beta \,\mathrm{d}\nu - \log \int_{\mathbb{R}^{4d}} e^{\alpha\oplus\beta} \,\mathrm{dR}_{0,T} \right\}.$$

Finally, the supremum is attained at the couple $(-\varphi^T, -\psi^T)$.

Proof. We have already seen in the previous proof that $\mathcal{C}_T(\mu,\nu)$ is finite. Now, since (7.0.5) is equivalent to the minimisation problem (7.2.17), from [Léo01, Proposition 6.1] it follows

$$\begin{aligned}
\mathcal{C}_T(\mu,\nu) &= \sup_{\alpha,\beta\in\mathcal{M}_b(\mathbb{R}^d)} \left\{ \int_{\mathbb{R}^d} \alpha \,\mathrm{d}\mu + \int_{\mathbb{R}^d} \beta \,\mathrm{d}\nu - \int_{\mathbb{R}^{2d}} \left(e^{\alpha\oplus\beta} - 1\right) \mathrm{dR}^X_{0,T} \right\} \\
&\le \sup_{\alpha,\beta\in\mathcal{M}_b(\mathbb{R}^d)} \left\{ \int_{\mathbb{R}^d} \alpha \,\mathrm{d}\mu + \int_{\mathbb{R}^d} \beta \,\mathrm{d}\nu - \log \int_{\mathbb{R}^{4d}} e^{\alpha\oplus\beta} \,\mathrm{dR}_{0,T} \right\} \\
&= \sup_{\alpha,\beta\in\mathcal{M}_b(\mathbb{R}^d)} \left\{ \int_{\mathbb{R}^d} (\alpha\oplus\beta) \,\mathrm{d}\pi - \log \int_{\mathbb{R}^{2d}} e^{\alpha\oplus\beta} \,\mathrm{dR}^X_{0,T} \right\} \\
&\le \sup_{h\in\mathcal{M}_b(\mathbb{R}^{2d})} \left\{ \int_{\mathbb{R}^{2d}} h \,\mathrm{d}\pi - \log \int_{\mathbb{R}^{2d}} e^{h} \,\mathrm{dR}^X_{0,T} \right\} \\
&\overset{(1.A.2)}{=} \mathscr{H}\left(\pi | \mathrm{R}^X_{0,T}\right) = \mathcal{C}_T(\mu,\nu),
\end{aligned}$$

where π is the unique optimiser in (7.2.17). This concludes the proof since $\mathrm{R}^X_{0,T} := (\mathrm{proj}_{x_1}, \mathrm{proj}_{x_2})_\#\mathrm{R}_{0,T}$. □

The Schrödinger system (7.2.16) is particularly useful when f^T and g^T are regular enough. Under (H1) and (H4) they inherit the regularity (smoothness and integrability) of the densities of μ, ν respectively. This follows from the identities

$$\frac{\mathrm{d}\mu}{\mathrm{d}\mathfrak{m}_X} = f^T \int_{\mathbb{R}^d} P_T g^T \,\mathrm{d}\mathfrak{m}_V\,, \qquad \frac{\mathrm{d}\nu}{\mathrm{d}\mathfrak{m}_X} = g^T \int_{\mathbb{R}^d} P^*_T f^T \,\mathrm{d}\mathfrak{m}_V\,,$$

and since $P^*_T f^T$ and $P_T g^T$ are smooth and positive (as a result of the lower bound (7.2.5)). Moreover, arguing exactly as in Lemma 2.1 in [CT21], owing to the lower bound in (7.2.5), and the continuity of p_T, we have that there is $c_{T_0} > 0$, (possibly depending on μ and ν) such that for all $T \ge T_0$

$$\begin{aligned}
\|f^T\|_{\mathrm{L}^\infty(\mathfrak{m})} \|g^T\|_{\mathrm{L}^1(\mathfrak{m})} &\le c_{T_0} \left\| \frac{\mathrm{d}\mu}{\mathrm{d}\mathfrak{m}_X} \right\|_{\mathrm{L}^\infty(\mathfrak{m})}, \\
\|f^T\|_{\mathrm{L}^1(\mathfrak{m})} \|g^T\|_{\mathrm{L}^\infty(\mathfrak{m})} &\le c_{T_0} \left\| \frac{\mathrm{d}\nu}{\mathrm{d}\mathfrak{m}_X} \right\|_{\mathrm{L}^\infty(\mathfrak{m})}.
\end{aligned}$$

These bounds are pivotal to prove that $f^T \to \mathrm{d}\mu/\mathrm{d}\mathfrak{m}_X$ and $g^T \to \mathrm{d}\nu/\mathrm{d}\mathfrak{m}_X$ as $T \to \infty$ in $\mathrm{L}^p(\mathfrak{m})$ for all $p \in [1, \infty)$ akin to what is done in Lemma 3.6 of [CT21].

To ensure that f^T, g^T are in $\mathrm{L}^\infty(\mathfrak{m})$ and with compact support, we work under assumption (H1) and (H4) for the rest of the section. With the help of the forward and adjoint semigroup, and (7.2.15) we can write

$$\begin{cases} \mu_0^T = f^T\, P_T g^T\, \mathfrak{m}\,, \\ \mu_T^T = g^T\, P_T^* f^T\, \mathfrak{m}\,, \end{cases}$$

where we recall that $\mu_t^T = (X_t,\, V_t)_\# \mathbf{P}^T$, with $\mathbf{P}^T$ being optimal for (7.0.1). Furthermore, if we set,

$$f_t^T := P_t^* f^T \qquad \text{and} \qquad g_t^T := P_{T-t} g^T\,,$$

then μ_t^T, $t \in [0, T]$, can be represented as

$$\mathrm{d}\mu_t^T = f_t^T g_t^T \mathrm{d}\mathfrak{m}\,. \tag{7.2.21}$$

It is also immediate to check that it holds

$$\begin{cases} \partial_t f_t^T = L^* f_t^T \\ \partial_t g_t^T = -L g_t^T \end{cases} \quad \text{and} \quad \begin{cases} \partial_t \log f_t^T = L^* \log f_t^T + \Gamma(\log f_t^T) \\ \partial_t \log g_t^T = -L \log g_t^T - \Gamma(\log g_t^T)\,, \end{cases} \tag{7.2.22}$$

where $\Gamma(h) = \gamma\, |\nabla_v h|^2$ is the *carré du champ* operator associated to the generator L.

The fg-decomposition gives us a nice representation formula for the relative entropy along the entropic interpolation $(\mu_t^T)_{t\in[0,T]}$. Indeed, if we introduce the functions

$$h_f^T(t) := \int_{\mathbb{R}^{2d}} \log f_t^T\, \rho_t^T\, \mathrm{d}\mathfrak{m} \quad \text{and} \quad h_b^T(t) := \int_{\mathbb{R}^{2d}} \log g_t^T\, \rho_t^T\, \mathrm{d}\mathfrak{m} \quad \forall t \in [0, T]\,,$$

then it easily follows that

$$\mathscr{H}(\mu_t^T|\mathfrak{m}) = h_f^T(t) + h_b^T(t), \quad \forall t \in [0, T]\,. \tag{7.2.23}$$

Moreover, we have

$$\partial_t h_f^T(t) = -\int_{\mathbb{R}^{2d}} \Gamma(\log f_t^T)\rho_t^T\, \mathrm{d}\mathfrak{m} \quad \text{and} \quad \partial_t h_b^T(t) = \int_{\mathbb{R}^{2d}} \Gamma(\log g_t^T)\rho_t^T\, \mathrm{d}\mathfrak{m}\,. \tag{7.2.24}$$

For a proof of (7.2.24) we refer to Lemma 3.8 in [Con19] where the classical setting is studied, the only difference in the kinetic setting being that the operator that acts on f_t^T should be replaced with L^*, since L is not self-adjoint.

In addition to (7.2.23), the fg-decomposition gives the following representation for the kinetic entropic cost

$$\begin{aligned} \mathcal{C}_T(\mu, \nu) =& \mathscr{H}(\boldsymbol{\pi}^T|\mathrm{R}_{0,T}) = \mathbb{E}_{\mathrm{R}_{0,T}}\left[\rho^T \log \rho^T\right] \\ &= \int_{\mathbb{R}^{2d}} \log f^T\, \rho_0^T\, \mathrm{d}\mathfrak{m} + \int_{\mathbb{R}^{2d}} \log g^T\, \rho_T^T\, \mathrm{d}\mathfrak{m} = h_f^T(0) + h_b^T(T)\,. \end{aligned} \tag{7.2.25}$$

As a byproduct of (7.2.23), (7.2.24), and (7.2.25), we get the identities

$$\begin{aligned} \mathcal{C}_T(\mu,\nu) =& \mathscr{H}(\mu_0^T|\mathfrak{m}) + \int_0^T \int_{\mathbb{R}^{2d}} \Gamma(\log g_t^T)\rho_t^T \,\mathrm{d}\mathfrak{m}\,\mathrm{d}t\,, \\ \mathcal{C}_T(\mu,\nu) =& \mathscr{H}(\mu_T^T|\mathfrak{m}) + \int_0^T \int_{\mathbb{R}^{2d}} \Gamma(\log f_t^T)\rho_t^T \,\mathrm{d}\mathfrak{m}\,\mathrm{d}t\,, \end{aligned} \tag{7.2.26}$$

which corresponds to what we have already shown in Lemma 2.3.1 for classical SP.

A straightforward consequence of the previous identities is the following

Lemma 7.2.5. *Under the assumptions (H1), (H4), for any $t \in [0,T]$ it holds*

$$\begin{aligned} \mathcal{C}_T(\mu,\nu) = \; & \mathscr{H}(\mu_0^T|\mathfrak{m}) + \mathscr{H}(\mu_T^T|\mathfrak{m}) + \int_0^t \int_{\mathbb{R}^{2d}} \Gamma(\log g_s^T)\rho_s^T \,\mathrm{d}\mathfrak{m}\,\mathrm{d}s \\ & + \int_t^T \int_{\mathbb{R}^{2d}} \Gamma(\log f_s^T)\rho_s^T \,\mathrm{d}\mathfrak{m}\,\mathrm{d}s - \mathscr{H}\left(\mu_t^T|\mathfrak{m}\right). \end{aligned} \tag{7.2.27}$$

Proof. From (7.2.26) we can write

$$\mathcal{C}_T(\mu,\nu) = \mathscr{H}(\mu_0^T|\mathfrak{m}) + \int_0^t \int_{\mathbb{R}^{2d}} \Gamma(\log g_s^T)\rho_s^T \,\mathrm{d}\mathfrak{m}\,\mathrm{d}s + \int_t^T \int_{\mathbb{R}^{2d}} \Gamma(\log g_s^T)\rho_s^T \,\mathrm{d}\mathfrak{m}\,\mathrm{d}s.$$

Applying the identities (7.2.24) we obtain that the last summand equals

$$\begin{aligned} &\int_t^T \int_{\mathbb{R}^{2d}} \Gamma(\log g_s^T)\rho_s^T \,\mathrm{d}\mathfrak{m}\,\mathrm{d}s \\ &\qquad = \int_t^T \int_{\mathbb{R}^{2d}} \Gamma(\log f_s^T)\rho_s^T \,\mathrm{d}\mathfrak{m}\,\mathrm{d}s + \int_t^T \partial_s h_b^T(s) + \partial_s h_f^T(s)\,\mathrm{d}s \\ &\qquad = \int_t^T \int_{\mathbb{R}^{2d}} \Gamma(\log f_s^T)\rho_s^T \,\mathrm{d}\mathfrak{m}\,\mathrm{d}s + \int_t^T \partial_s \mathscr{H}(\mu_s^T|\mathfrak{m})\,\mathrm{d}s \\ &\qquad = \int_t^T \int_{\mathbb{R}^{2d}} \Gamma(\log f_s^T)\rho_s^T \,\mathrm{d}\mathfrak{m}\,\mathrm{d}s + \mathscr{H}(\mu_T^T|\mathfrak{m}) - \mathscr{H}\left(\mu_t^T|\mathfrak{m}\right), \end{aligned}$$

and we reach our conclusion. □

7.2.5 The fg-decomposition for KFSP

We now discuss the structural properties for KFSP, as we did in Section 7.2.4 for KSP. Notice that, since we consider fixed the full marginals at time 0 and T, clearly in this case f and g are function of both space and velocity.

Proposition 7.2.6. *Grant (H1), (FH3). Then, for all $T > 0$,* (7.0.6) *and* (7.0.4) *admit unique solutions $\bar{\boldsymbol{\pi}}^T$, $\bar{\mathbf{P}}^T$ with $\bar{\boldsymbol{\pi}}^T = ((X_0,V_0),(X_T,V_T))_{\#}\bar{\mathbf{P}}^T$ and there exist two non-negative measurable functions $\bar{f}^T$, $\bar{g}^T$ on $\mathbb{R}^{2d}$ such that*

$$\bar{\rho}^T(x,v,y,w) := \frac{\mathrm{d}\bar{\boldsymbol{\pi}}^T}{\mathrm{d}\mathrm{R}_{0,T}}(x,v,y,w) = \bar{f}^T(x,v)\bar{g}^T(y,w) \qquad \mathrm{R}_{0,T}\text{-}a.s.$$

and that solve the Schrödinger system

$$\begin{cases} \frac{\mathrm{d}\bar\mu}{\mathrm{d}\mathfrak{m}}(x,v) = \bar f^T(x,v)\,\mathbb{E}_{\mathrm{R}}\big[\bar g^T(X_T,V_T)|X_0 = x,\, V_0 = v\big]\,, \\ \frac{\mathrm{d}\bar\nu}{\mathrm{d}\mathfrak{m}}(y,w) = \bar g^T(y,w)\,\mathbb{E}_{\mathrm{R}}\big[\bar f^T(X_0,V_0)|X_T = y,\, V_T = w\big]\,. \end{cases} \tag{7.2.28}$$

Proof. This follows in a standard way from [RT93] or [Léo14] (or also Theorem 2.2.1). □

Then, we may again define the potentials As we did for the classic problem in (6.0.1) we can define the kinetic Schrödinger potentials as the measurable function $\bar\varphi^T := -\log \bar f^T \in \mathrm{L}^1(\bar\mu)$ and $\bar\psi^T := -\log \bar g^T \in \mathrm{L}^1(\bar\nu)$ and then clearly it holds

$$\frac{\mathrm{d}\bar{\boldsymbol{\pi}}^T}{\mathrm{d}\mathrm{R}_{0,T}}((x,v),(y,w)) = \exp(-\bar\varphi^T(x,v) - \bar\psi^T(y,w)) \quad \mathrm{R}_{0,T}\text{-a.e.}\,,$$

and we may consider again a Kantorovich-type duality (for $(-\bar\varphi^T)$ and $(-\bar\psi^T)$)

Proposition 7.2.7. *Grant (H1) and (FH3). Then, $\mathcal{C}_T^F(\bar\mu,\bar\nu) < \infty$ and*

$$\mathcal{C}_T^F(\bar\mu,\bar\nu) = \sup_{\varphi,\psi\in\mathcal{M}_b(\mathbb{R}^{2d})} \left\{ \int_{\mathbb{R}^{2d}} \varphi\,\mathrm{d}\bar\mu + \int_{\mathbb{R}^{2d}} \psi\,\mathrm{d}\bar\nu - \log \int_{\mathbb{R}^{4d}} e^{\varphi\oplus\psi}\,\mathrm{d}\mathrm{R}_{0,T} \right\}.$$

Finally, the supremum is attained at the couple $(-\bar\varphi^T, -\bar\psi^T)$.

The proof of the above results runs exactly as the one presented for Proposition 2.2.2 and Proposition 7.2.4.

Exactly as in Section 7.2.4, under (H1) and (FH4), $\bar f^T$ and $\bar g^T$ inherit the integrability and regularity of the densities $\mathrm{d}\bar\mu/\mathrm{d}\mathfrak{m}$ and $\mathrm{d}\bar\nu/\mathrm{d}\mathfrak{m}$. In this case this is due to the identities

$$\frac{\mathrm{d}\bar\mu}{\mathrm{d}\mathfrak{m}} = \bar f^T P_T \bar g^T\,, \qquad \frac{\mathrm{d}\bar\nu}{\mathrm{d}\mathfrak{m}} = P_T^* \bar f^T \bar g^T\,,$$

and the fact that $P_T^*\bar f^T$ and $P_T\bar g_T$ are positive and smooth. For the rest of the section we shall assume that (FH4) holds true which guarantees that $\bar f^T$ and $\bar g^T$ belong to $\mathrm{L}^\infty(\mathfrak{m})$ and have compact support.

We define for any $t \in [0,T]$

$$\bar f_t^T := P_t^* \bar f^T \qquad \text{and} \qquad \bar g_t^T := P_{T-t}\bar g^T\,.$$

We recall that $\bar\mu_t^T = (X_t,\, V_t)_\# \bar{\mathbf{P}}^T$, with $\bar{\mathbf{P}}^T$ being the solution to (7.0.4). Then

$$\bar\mu_t^T = \bar f_t^T \bar g_t^T \mathfrak{m}\,.$$

Furthermore, (7.2.28) implies that

$$\bar\mu_0^T = \bar f_0^T \bar g_0^T \mathfrak{m} = \bar\mu \qquad \text{and} \qquad \bar\mu_T^T = \bar f_T^T \bar g_T^T \mathfrak{m} = \bar\nu\,,$$

and it is easy to check that it holds

$$\begin{cases}\partial_t \bar f^T_t = L^* \bar f^T_t \\ \partial_t \bar g^T_t = -L \bar g^T_t\end{cases} \quad \text{and} \quad \begin{cases}\partial_t \log \bar f^T_t = L^* \log \bar f^T_t + \Gamma(\log \bar f^T_t) \\ \partial_t \log \bar g^T_t = -L \log \bar g^T_t - \Gamma(\log \bar g^T_t)\,.\end{cases}$$

Similarly to (7.2.27), under (H1) and (FH4) it holds that, for any $t \in [0, T]$,

$$\begin{aligned}\mathcal{C}^F_T(\bar\mu, \bar\nu) = \; & \mathscr{H}(\bar\mu|\mathfrak{m}) + \mathscr{H}(\bar\nu|\mathfrak{m}) - \mathscr{H}(\bar\mu^T_t|\mathfrak{m}) \\ & + \int_0^t \int_{\mathbb{R}^{2d}} \Gamma(\log \bar g^T_s)\bar\rho^T_s \,\mathrm{d}\mathfrak{m}\,\mathrm{d}s + \int_t^T \int_{\mathbb{R}^{2d}} \Gamma(\log \bar f^T_s)\bar\rho^T_s \,\mathrm{d}\mathfrak{m}\,\mathrm{d}s.\end{aligned} \tag{7.2.29}$$

7.3 Qualitative long-time behaviour

Throughout the whole section we will always assume (H1) and (H3) (respectively (FH3) for KFSP) to be true. Let us just recall here that (H1) implies that $\mathfrak{m} \in \mathcal{P}_2(\mathbb{R}^{2d})$. Note that since $(P_t)_{t\in[0,T]}$ is strongly mixing [DPZ14, Theorem 11.14] for any $\psi, \phi \in C_b(\mathbb{R}^{2d})$ it holds

$$\begin{aligned}&\int_{\mathbb{R}^{2d}} \int_{\mathbb{R}^{2d}} \psi(x,v)\phi(y,w)\mathrm{d}\mathrm{R}_{0,T_n} \\ &= \int_{\mathbb{R}^{2d}} \psi \, P_{T_n}\phi \,\mathrm{d}\mathfrak{m} \xrightarrow{n\to\infty} \int_{\mathbb{R}^{2d}} \int_{\mathbb{R}^{2d}} \psi(x,v)\phi(y,w)\mathrm{d}\mathfrak{m}\otimes\mathrm{d}\mathfrak{m}\,.\end{aligned}$$

From the Portmanteau Theorem, it follows that

$$\mathrm{R}_{0,T_n} \rightharpoonup \mathfrak{m}\otimes\mathfrak{m}\,. \tag{7.3.1}$$

This first weak-convergence result already suggests the turnpike property, or at least a qualitative version of it. Intuitively, this implies that the variational problem KSP (*i.e.*, (7.0.5)) converges, in a sense to be made precise, to the problem

$$\min_{\pi\in\Pi_X(\mu,\nu)} \mathscr{H}(\pi \mid \mathfrak{m}\otimes\mathfrak{m}), \tag{7.3.2}$$

whose optimal solution and optimal value are easily seen to be $(\mu\otimes\mathfrak{m}_V)\otimes(\nu\otimes\mathfrak{m}_V)$ and $\mathscr{H}(\mu|\mathfrak{m}_X) + \mathscr{H}(\nu|\mathfrak{m}_X)$ respectively. From the point of view of the particle system, this means that in the long-time limit, initial and final states of the system become essentially independent of one another. Moreover, the initial and final velocities are well approximated by independent Gaussians, and are independent from the spatial variables. The result below turns this intuition into a solid argument. For a quantitative version of the convergence of the entropic cost towards the sum of the marginal entropies we refer the reader to Theorem 7.5.4 below. For the classical Schrödinger problem, an analogous statement can be found in [CT21].

Our approach relies on a Γ-convergence approach similar to the one used in [CT21] for the classical Schrödinger problem. The main difference with [CT21] is the lack of compactness for the set $\Pi_X(\mu,\nu)$, problem that we address in the next lemma.

Lemma 7.3.1 (Equicoerciveness)**.** *The family of entropic operators*

$$\{\mathscr{H}(\cdot \mid \mathrm{R}_{0,T_n})\colon \Pi_X(\mu,\nu) \to [0,\infty]\}_{n\in\mathbb{N}}$$

is equicoercive, i.e., for any $h \in \mathbb{R}$ there exists a (weakly) compact subset $K_h \subset \Pi_X(\mu,\, \nu)$ such that

$$\left\{q \in \Pi_X\,(\mu,\,\nu)\ \textit{s.t.}\ \mathscr{H}(q \mid \mathrm{R}_{0,T_n}) \le h\right\} \subseteq K_h \quad \forall n \in \mathbb{N}\,.$$

Proof. Since $(\mathrm{R}_{0,T_n})_{n\in\mathbb{N}}$ is tight, a proof of this result is obtained by following the same argument given in [DE97, Lemma 1.4.3c].

Thanks to Prohorov's Theorem it is enough to show that for any $t \in \mathbb{R}$, any sequence $(q_n)_{n\in\mathbb{N}} \subseteq \{q \in \Pi_X\,(\mu,\,\nu)$ s.t. $\mathscr{H}(q \mid \mathrm{R}_{0,T_n}) \le t \quad \forall n \in \mathbb{N}\}$ is tight. Firstly, fix a real number $t \in \mathbb{R}$.

Notice that $(\mathrm{R}_{0,T_n})_{n\in\mathbb{N}}$ is tight since $\mathrm{R}_{0,T_n} \rightharpoonup \mathfrak{m}\otimes\mathfrak{m}$ (cf. (7.3.1)). Hence for any $\epsilon > 0$ there exists a compact subset $K_\epsilon \subset \mathbb{R}^{4d}$ such that

$$\sup_{n\in\mathbb{N}} \mathrm{R}_{0,T_n}(K_\epsilon^C) < \epsilon\,.$$

Now, consider the bounded measurable function $\psi\colon \mathbb{R}^{4d} \to \mathbb{R}$ defined as

$$\psi := \log\left(1+\frac{1}{\epsilon}\right)\mathbf{1}_{K_\epsilon^C}\,.$$

Then, the Donsker-Varadhan formula (1.A.2) tells us that for any $n \in \mathbb{N}$

$$\begin{aligned}
\mathscr{H}(q_n \mid \mathrm{R}_{0,T_n}) &\ge \int_{\mathbb{R}^{4d}} \psi\,\mathrm{d}q_n - \log\int_{R^{4d}} e^{\psi}\mathrm{d}\mathrm{R}_{0,T_n}\\
&= \log\left(1+\frac{1}{\epsilon}\right) q_n(K_\epsilon^C) - \log\left(\int_{K_\epsilon} 1\,\mathrm{d}\mathrm{R}_{0,T_n} + \int_{K_\epsilon^C}\left(1+\frac{1}{\epsilon}\right)\mathrm{d}\mathrm{R}_{0,T_n}\right)\\
&= \log\left(1+\frac{1}{\epsilon}\right) q_n(K_\epsilon^C) - \log\left(\mathrm{R}_{0,T_n}(K_\epsilon) + \mathrm{R}_{0,T_n}(K_\epsilon^C) + \frac{1}{\epsilon}\mathrm{R}_{0,T_n}(K_\epsilon^C)\right)\\
&= \log\left(1+\frac{1}{\epsilon}\right) q_n(K_\epsilon^C) - \log\left(1 + \frac{1}{\epsilon}\mathrm{R}_{0,T_n}(K_\epsilon^C)\right)\\
&\ge \log\left(1+\frac{1}{\epsilon}\right) q_n(K_\epsilon^C) - \log 2\,.
\end{aligned}$$

Therefore, since $\mathscr{H}(q_n \mid \mathrm{R}_{0,T_n}) \le t$, for any $\delta > 0$ it holds

$$\sup_{n\in\mathbb{N}} q_n(K_\epsilon^C) \le \sup_{n\in\mathbb{N}} \frac{\mathscr{H}(q_n \mid \mathrm{R}_{0,T_n}) + \log 2}{\log\left(1+\frac{1}{\epsilon}\right)} \le \frac{t+\log 2}{\log\left(1+\frac{1}{\epsilon}\right)} < \delta\,,$$

where the last inequality holds for $\epsilon = \epsilon_\delta$ small enough, and hence $(q_n)_{n\in\mathbb{N}}$ is tight. $\square$

Theorem 7.3.2 (A Γ-convergence result). *Assume (H1) and (H3). Let $(T_n)_{n\in\mathbb{N}}$ be a sequence of positive real numbers converging to ∞, and for each $n \in \mathbb{N}$ consider the functional $\mathscr{H}(\cdot \mid \mathrm{R}_{0,T_n})$ defined on $\Pi_X(\mu, \nu)$ endowed with the weak topology. Then*

$$\Gamma - \lim_{n\to\infty} \mathscr{H}(\cdot|\mathrm{R}_{0,T_n}) = \mathscr{H}(\cdot|\mathfrak{m}\otimes\mathfrak{m})\,.$$

As a direct consequence, we obtain that

$$\lim_{T\to\infty} \mathcal{C}_T(\mu, \nu) = \mathscr{H}(\mu \mid \mathfrak{m}_X) + \mathscr{H}(\nu \mid \mathfrak{m}_X) < \infty\,, \tag{7.3.3}$$

and that as $T \to \infty$ it holds

$$\boldsymbol{\pi}^T \rightharpoonup (\mu\otimes\mathfrak{m}_V)\otimes(\nu\otimes\mathfrak{m}_V) \in \Pi_X(\mu,\nu) \quad \textit{weakly.} \tag{7.3.4}$$

Proof.

(Γ-convergence lower bound inequality) From (7.3.1) and the lower semi-continuity of the relative entropy we immediately obtain that for any sequence $(q_n)_{n\in\mathbb{N}} \subset \Pi_X(\mu, \nu)$ weakly converging to some $q \in \Pi_X(\mu, \nu)$ it holds

$$\liminf_{n\to\infty} \mathscr{H}(q_n \mid \mathrm{R}_{0,T_n}) \geq \mathscr{H}(q \mid \mathfrak{m}\otimes\mathfrak{m})\,. \tag{7.3.5}$$

(Γ-convergence upper bound inequality) We prove that for any $q \in \Pi_X(\mu, \nu)$ it holds

$$\limsup_{n\to\infty} \mathscr{H}(q \mid \mathrm{R}_{0,T_n}) \leq \mathscr{H}(q \mid \mathfrak{m}\otimes\mathfrak{m})\,. \tag{7.3.6}$$

We may assume $\mathscr{H}(q|\mathfrak{m}\otimes\mathfrak{m}) < \infty$ otherwise the above inequality is trivial. Note that this implies $q \in \mathcal{P}_2(\mathbb{R}^{4d})$ since $\mu,\nu \in \mathcal{P}_2(\mathbb{R}^d)$ (cf. (7.2.3)) while

$$\begin{aligned}\int_{\mathbb{R}^d} |v|^2 \,\mathrm{d}(\mathrm{proj}_{v_1})_\# q &\leq 2\int_{\mathbb{R}^d} |v|^2 \,\mathrm{d}\mathfrak{m}_V + 2\,\mathbf{W}_2((\mathrm{proj}_{v_1})_\# q, \mathfrak{m}_V)^2 \\ &\overset{(7.2.2)}{\lesssim} 1 + \mathscr{H}((\mathrm{proj}_{v_1})_\# q|\mathfrak{m}_V) \leq 1 + \mathscr{H}(q|\mathfrak{m}\otimes\mathfrak{m}) < \infty\,,\end{aligned}$$

and similarly for the measure $(\mathrm{proj}_{v_2})_\# q$. Then we have

$$\mathscr{H}(q \mid \mathrm{R}_{0,T_n}) = \mathscr{H}(q \mid \mathfrak{m}\otimes\mathfrak{m}) - \int_{\mathbb{R}^{2d}\times\mathbb{R}^{2d}} \log \mathrm{p}_{T_n}((x,v),(y,w))\,\mathrm{d}q\,,$$

Thanks to the lower bound given in Lemma 7.2.1 and the fact that $q \in \mathcal{P}_2(\mathbb{R}^{4d})$, we can apply Fatou's Lemma and get

$$\limsup_{n\to\infty} \mathscr{H}(q \mid \mathrm{R}_{0,T_n}) \leq \mathscr{H}(q \mid \mathfrak{m}\otimes\mathfrak{m}) - \int_{\mathbb{R}^{2d}\times\mathbb{R}^{2d}} \liminf_{n\to\infty} \log \mathrm{p}_{T_n}\,\mathrm{d}q\,. \tag{7.3.7}$$

Now, for all $t > 0$ and for all $(x,v),\,(y,w) \in \mathbb{R}^{2d}$

$$\mathsf{p}_{T_n}\big((x,v),\,(y,w)\big) = P_{T_n-t}\left(\mathsf{p}_t(\cdot\,,\,(y,w))\right)(x,v)\,.$$

For any $M > 0$, we introduce the function $\mathsf{p}_t^M(\cdot\,,\,(y,w)) := \mathsf{p}_t(\cdot\,,\,(y,w)) \wedge M \in C_b(\mathbb{R}^{2d})$. Then, since P_T is strongly mixing (cf. [DPZ14, Theorem 11.14]), we get

$$\begin{aligned}&\mathsf{p}_{T_n}\big((x,v),\,(y,w)\big)\\&\qquad \geq P_{T_n-t}\left(\mathsf{p}_t^M(\cdot\,,\,(y,w))\right)(x,v) \xrightarrow{n\to\infty} \int_{\mathbb{R}^{2d}} \mathsf{p}_t^M\big((x,v),(y,w)\big)\mathrm{d}\mathfrak{m}(x,v)\,.\end{aligned}$$

Taking the limit as $M \to \infty$, by dominated convergence we get that

$$\int_{\mathbb{R}^{2d}} \mathsf{p}_t^M\big((x,v),(y,w)\big)\mathrm{d}\mathfrak{m}(x,v) \xrightarrow{M\to\infty} \int_{\mathbb{R}^{2d}} \mathsf{p}_t\big((x,v),(y,w)\big)\mathrm{d}\mathfrak{m}(x,v) = 1\,,$$

where the last equality follows from (7.2.9). Therefore it holds

$$\liminf_{n\to\infty} \log \mathsf{p}_{T_n}\big((x,v),\,(y,w)\big) \geq 0\,, \quad \mathfrak{m}\otimes\mathfrak{m} - a.s.$$

which, together with $q \ll \mathfrak{m}\otimes\mathfrak{m}$ (since $\mathscr{H}(q \mid \mathfrak{m}\otimes\mathfrak{m}) < \infty$), leads to

$$\liminf_{n\to\infty} \log \mathsf{p}_{T_n}\big((x,v),\,(y,w)\big) \geq 0\,, \qquad q\text{-a.s.}$$

Therefore, from (7.3.7) we get inequality (7.3.6). The desired Γ-convergence follows as a byproduct of (7.3.5) and (7.3.6).

As a consequence of the Γ-convergence, we deduce the last two claims as follows. Firstly note that the unique minimiser in (7.3.2) is given the probability measure $\boldsymbol{\pi}^\infty := (\mu\otimes\mathfrak{m}_V)\otimes(\nu\otimes\mathfrak{m}_V)$. Now let us consider $(T_n)_{n\in\mathbb{N}}$ to be any diverging sequence of positive real times. Then, from the optimality of $\boldsymbol{\pi}^{T_n}$ it follows

$$\begin{aligned}\limsup_{n\to\infty} \mathscr{H}(\boldsymbol{\pi}^{T_n}|\mathrm{R}_{0,T_n}) \leq \limsup_{n\to\infty} \mathscr{H}(\boldsymbol{\pi}^{\infty}|\mathrm{R}_{0,T_n}) &\overset{(7.3.6)}{\leq} \mathscr{H}(\boldsymbol{\pi}^{\infty}|\mathfrak{m}\otimes\mathfrak{m})\\&= \mathscr{H}(\mu \mid \mathfrak{m}_X) + \mathscr{H}(\nu \mid \mathfrak{m}_X),\end{aligned}$$

which is finite by our assumptions. Then Lemma 7.3.1 implies that the subsequence $(\boldsymbol{\pi}^{T_n})_{n\in\mathbb{N}}$ is weakly relatively compact. Then, from the Fundamental Theorem of Γ-convergence [Bra06, Theorem 2.10], the uniqueness of the minimiser in (7.3.2) and from the metrizability of the weak convergence on $\mathcal{P}(\mathbb{R}^{4d})$ we deduce (7.3.3) and (7.3.4). □

Even though here we are just interested in the Γ-convergence (as introduced by De Giorgi) on $\Pi_X(\mu,\nu)$ equipped with the weak topology, the previous result is actually stronger: indeed we have actually proven the Mosco convergence of the functional $\mathscr{H}(\cdot \mid \mathrm{R}_{0,T_n})$ since we have considered a constant sequence $q_n = q$ for the upper bound inequality.

Remark 7.3.3. *Equation* (7.3.4) *implies in particular that* $\mu_0^T \rightharpoonup \mu \otimes \mathfrak{m}_V$ *and that* $\mu_T^T \rightharpoonup \nu \otimes \mathfrak{m}_V$. *This convergence is also exponential, as we will show in Theorem 7.4.5.*

The same reasoning applies to the full-setting in KFSP. We omit the proof of the next result since it runs similarly to the one given above in the kinetic setting. The main difference is that in this case the equicoerciveness is not needed since we have the weak compactness of $\Pi(\bar{\mu}, \bar{\nu})$.

Theorem 7.3.4. *Assume (H1) and (FH3). Let* $(T_n)_{n\in\mathbb{N}}$ *be a sequence of positive real numbers converging to* ∞, *and for each* $n \in \mathbb{N}$ *consider the functional* $\mathscr{H}(\cdot \mid \mathrm{R}_{0,T_n})$ *defined on* $\Pi(\bar{\mu}, \bar{\nu})$ *endowed with the weak topology. Then*

$$\Gamma - \lim_{n\to\infty} \mathscr{H}(\cdot|\mathrm{R}_{0,T_n}) = \mathscr{H}(\cdot|\mathfrak{m}\otimes\mathfrak{m}).$$

As a direct consequence, we obtain that

$$\lim_{T\to\infty} \mathcal{C}_T^F(\bar{\mu}, \bar{\nu}) = \mathscr{H}(\bar{\mu}|\mathfrak{m}) + \mathscr{H}(\bar{\nu}|\mathfrak{m}) < \infty, \tag{7.3.8}$$

and that as $T \to \infty$ *it holds*

$$\bar{\boldsymbol{\pi}}^T \rightharpoonup \bar{\mu}\otimes\bar{\nu} \in \Pi(\bar{\mu}, \bar{\nu}) \quad \textit{weakly.} \tag{7.3.9}$$

7.4 Corrector estimates

In the remaining part of this chapter we are going to prove quantitative estimates for (7.3.3), (7.3.4), (7.3.8) and (7.3.9). Throughout we assume (H1), (H2) and (H4) to be true and we will point out whenever the latter can be relaxed to (H3) for KSP. Let us start by defining a few key objects whose behaviour will help us in controlling the convergence rates for the turnpike property. We define the *correctors* as the functions $\varphi^T, \psi^T : [0, T] \to \mathbb{R}$ given by

$$\varphi^T(s) := \int_{\mathbb{R}^{2d}} \left|\nabla \log f_s^T\right|^2_{N^{-1}} \rho_s^T \,\mathrm{d}\mathfrak{m} \quad \text{and} \quad \psi^T(s) := \int_{\mathbb{R}^{2d}} \left|\nabla \log g_s^T\right|^2_{M^{-1}} \rho_s^T \,\mathrm{d}\mathfrak{m},$$

where $M, N \in \mathbb{R}^{2d\times 2d}$ are the matrices appearing in Proposition 7.2.2. Let us also note that by the fg-decomposition it follows $\mathcal{I}(\mu_s^T) \lesssim \varphi^T(s) + \psi^T(s)$. Notice that, besides the change of metric induced by the matrices M, N, the above quantities are the analogous of the correctors considered in Section 3.1 and the next result provides a kinetic version of (3.1.7) in Proposition 3.1.2. More precisely, with the next lemma we show that the contractive properties introduced in Section 7.2.3 translate into an exponentially fast contraction for the correctors.

Lemma 7.4.1. *Under (H1), (H2) and (H4), for any* $0 < t \le s \le T$ *it holds*

$$\varphi^T(s) \le \varphi^T(t) e^{-2\kappa(s-t)} \quad \textit{and} \quad \psi^T(T-s) \le \psi^T(T-t) e^{-2\kappa(s-t)}. \tag{7.4.1}$$

Proof. By definition $f_s^T = P_{s-t}^* f_t^T$ and thus

$$\varphi^T(s) = \int_{\mathbb{R}^{2d}} \left|\nabla \log f_s^T\right|^2_{N^{-1}} \rho_s^T \,\mathrm{d}\mathfrak{m} = \int_{\mathbb{R}^{2d}} \left|\nabla P_{s-t}^* f_t^T\right|^2_{N^{-1}} (P_{s-t}^* f_t^T)^{-1} g_s^T \,\mathrm{d}\mathfrak{m}$$

An application of the gradient estimate (7.2.13) and Cauchy-Schwartz inequality yields

$$\begin{aligned}
\varphi^T(s) &\overset{(7.2.13)}{\leq} e^{-2\kappa(s-t)} \int_{\mathbb{R}^{2d}} \left(P_{s-t}^* \left|\nabla f_t^T\right|_{N^{-1}}\right)^2 (P_{s-t}^* f_t^T)^{-1} P_{T-s} g^T \,\mathrm{d}\mathfrak{m} \\
&\leq e^{-2\kappa(s-t)} \int_{\mathbb{R}^{2d}} P_{s-t}^* \left(\frac{|\nabla f_t^T|^2_{N^{-1}}}{f_t^T}\right) P_{T-s} g^T \,\mathrm{d}\mathfrak{m} \\
&= e^{-2\kappa(s-t)} \int_{\mathbb{R}^{2d}} \left|\nabla \log f_t^T\right|^2_{N^{-1}} \rho_t^T \,\mathrm{d}\mathfrak{m} \leq e^{-2\kappa(s-t)} \varphi^T(t),
\end{aligned}$$

which concludes the proof for the first inequality. The analogous inequality for ψ^T runs as above by using inequality (7.2.12) for the semigroup $(P_t)_{t\in[0,T]}$. □

Proposition 7.4.2. *Grant (H1), (H2) and (H4). There exists $C_{d,\alpha,\beta,\gamma} > 0$ such that for any $0 < \delta \leq 1$ and for any $t \in [\delta, T]$, as soon as $T > \frac{1}{\kappa} \log C_{d,\alpha,\beta,\gamma} + 2\delta$, it holds*

$$\begin{aligned}
\varphi^T(t) &\lesssim e^{-2\kappa t} \left[\mathcal{I}\left(\mu_\delta^T\right) + \mathcal{I}\left(\mu_{T-\delta}^T\right)\right], \\
\psi^T(T-t) &\lesssim e^{-2\kappa t} \left[\mathcal{I}\left(\mu_\delta^T\right) + \mathcal{I}\left(\mu_{T-\delta}^T\right)\right].
\end{aligned} \tag{7.4.2}$$

Proof. Without loss of generalities we may assume $\mathcal{I}\left(\mu_\delta^T\right)$ and $\mathcal{I}\left(\mu_{T-\delta}^T\right)$ to be finite, otherwise the above bounds are trivial. From Lemma 7.4.1 and the fg-decomposition of $\rho_t^T = f_t^T g_t^T$ we know that

$$\begin{aligned}
\varphi^T(T-\delta) &\leq e^{-2\kappa T + 4\kappa\delta} \varphi^T(\delta) \\
&= e^{-2\kappa T + 4\kappa\delta} \int \left|\nabla \log \rho_\delta^T - \nabla \log g_\delta^T\right|^2_{N^{-1}} \mathrm{d}\mu_\delta^T \\
&\lesssim e^{-2\kappa T + 4\kappa\delta} \mathcal{I}\left(\mu_\delta^T\right) + e^{-2\kappa T + 4\kappa\delta} \psi^T(\delta) \\
&\lesssim e^{-2\kappa T + 4\kappa\delta} \mathcal{I}\left(\mu_\delta^T\right) + e^{-4\kappa T + 8\kappa\delta} \psi^T(T-\delta).
\end{aligned}$$

Using the basic inequality $|a-b|^2 \geq a^2/2 - b^2$ we obtain

$$\begin{aligned}
\varphi^T(T-\delta) &= \int_{\mathbb{R}^{2d}} \left|\nabla \log g_{T-\delta}^T - \nabla \log \rho_{T-\delta}^T\right|^2_{N^{-1}} \mathrm{d}\mu_{T-\delta}^T \\
&\gtrsim \psi^T(T-\delta) - 2\mathcal{I}\left(\mu_{T-\delta}^T\right).
\end{aligned}$$

As a result, we get

$$\psi^T(T-\delta) - 2\mathcal{I}\left(\mu_{T-\delta}^T\right) \lesssim e^{-2\kappa T + 4\kappa\delta} \mathcal{I}\left(\mu_\delta^T\right) + e^{-4\kappa T + 8\kappa\delta} \psi^T(T-\delta).$$

Therefore, as soon as $T > \frac{1}{\kappa}\log C_{d,\alpha,\beta,\gamma} + 2\delta$ for some constant $C_{d,\alpha,\beta,\gamma} > 0$, we find

$$\psi^T(T-\delta) \lesssim \mathcal{I}\left(\mu^T_\delta\right) + \mathcal{I}\left(\mu^T_{T-\delta}\right) .$$

Plugging this bound into the contraction estimate (7.4.1) gives the second inequality in (7.4.2) for any $t \in [\delta, T]$. The first inequality is obtained by exchanging the roles of φ and ψ in the above discussion. □

The above results already provides us a uniform in time bound for the Fisher information along Schrödinger bridges.

Proposition 7.4.3. *Assume (H1), (H2) and (H3). Let $0 < \delta \le 1$ be fixed. Then, for all $t \in [\delta, T-\delta]$*

$$\mathcal{I}(\mu^T_t) \lesssim \delta^{-3}\Big(\mathcal{C}_T(\mu,\nu) - \mathscr{H}(\mu^T_t \mid \mathfrak{m})\Big).$$

Proof. Let us first work under (H4). We claim that for any $t \in [\delta, T-\delta]$ it holds

$$\left|\nabla P_{T-t} g^T\right|^2 \lesssim \delta^{-3}\left[P_{T-t}(g^T \log g^T) - (P_{T-t}\, g^T)\log(P_{T-t}\, g^T)\right] P_{T-t}\, g^T . \tag{7.4.3}$$

Indeed by applying Corollary 3.2 in [GW12] to any directional derivative we have

$$\begin{aligned}
&\left|\partial_{x_i} P_{T-t} g^T\right|^2 \\
&\quad\le 4 \inf_{s\in(0,T-t]} \Psi_s(1,0)\left[P_{T-t}(g^T \log g^T) - (P_{T-t}\, g^T)\log(P_{T-t}\, g^T)\right] P_{T-t}\, g^T, \\
&\left|\partial_{v_i} P_{T-t} g^T\right|^2 \\
&\quad\le 4 \inf_{s\in(0,T-t]} \Psi_s(0,1)\left[P_{T-t}(g^T \log g^T) - (P_{T-t}\, g^T)\log(P_{T-t}\, g^T)\right] P_{T-t}\, g^T,
\end{aligned}$$

where $\Psi_s(a,b)$ is defined for any $a, b > 0$ as the quantity

$$\Psi_s(a,b) := \frac{1}{2\gamma} s \left[a\left(\frac{6}{s^2} + \beta + \frac{3\gamma}{2s}\right) + b\left(\frac{4}{s} + \frac{4\beta}{27}s + \gamma\right)\right]^2 .$$

By considering $s = \delta \in (0,1]$ we can bound the above RHS with δ^{-3}, up to a multiplicative constant. Particularly this yields (7.4.3). Similarly one can prove that it holds

$$\left|\nabla P^*_t f^T\right|^2 \lesssim \delta^{-3}\left[P^*_t(f^T \log f^T) - (P^*_t\, f^T)\log(P^*_t\, f^T)\right] P^*_t\, f^T.$$

Therefore because of the fg-decomposition (7.2.21) we obtain

$$\begin{aligned}\mathcal{I}(\mu_t^T) &\leq 2\int_{\mathbb{R}^{2d}} \left[\frac{|\nabla P_{T-t}g^T|^2}{P_{T-t}g^T} P_t^* f^T + \frac{|\nabla P_t^* f^T|^2}{P_t^* f^T} P_{T-t}g^T\right] \mathrm{d}\mathfrak{m} \\ &\lesssim \delta^{-3}\int_{\mathbb{R}^{2d}} \Big\{ \Big[P_{T-t}(g^T \log g^T) - (P_{T-t}\, g^T)\log(P_{T-t}\, g^T)\Big] P_t^* f^T \\ &\qquad + \Big[P_t^*(f^T \log f^T) - (P_t^*\, f^T)\log(P_t^*\, f^T)\Big] P_{T-t}g^T \Big\}\,\mathrm{d}\mathfrak{m}\end{aligned}$$

By integrating by parts and (7.2.21) this last displacement equals

$$\delta^{-3}\left(\int_{\mathbb{R}^d} \log g^T \,\mathrm{d}\nu + \int_{\mathbb{R}^d} \log f^T \,\mathrm{d}\mu - \int_{\mathbb{R}^{2d}} \log \rho_t^T\, \rho_t^T \,\mathrm{d}\mathfrak{m}\right),$$

and the thesis follows in view of (7.2.25).

Now let us just assume (H3). Firstly, define the probability measure q_n^T as the measure whose $\mathrm{R}_{0,T}$-density is given by

$$\frac{\mathrm{d}q_n^T}{\mathrm{dR}_{0,T}} := \left(\rho^T \wedge n\right) \frac{\mathbf{1}_{K_n}}{C_n}\,, \tag{7.4.4}$$

where $(K_n)_{n\in\mathbb{N}}$ is an increasing sequence of compact sets in $\mathbb{R}^{4d}$ and C_n is the normalising constant. Then, by applying Lemma 7.A.1 we know that the marginals $\mu^n := (\mathrm{proj}_{x_1})_\# q_n^T$ and $\nu^n := (\mathrm{proj}_{x_2})_\# q_n^T$ satisfy (H4) and by means of Proposition 7.A.2 and Corollary 7.A.3 it follows that there exists a unique minimiser $\mathbf{P}^{n,T} \in \mathcal{P}(\Omega_{2d})$ for KSP with marginals μ^n, ν^n, and as soon as n diverges it holds

$$\mu_t^{n,T} \rightharpoonup \mu_t^T \qquad \text{and} \qquad \mathcal{C}_T(\mu^n,\nu^n) \to \mathcal{C}_T(\mu,\nu)\,. \tag{7.4.5}$$

Then, the thesis in the general case follows from the one under (H4) and the lower semicontinuity of $\mathcal{I}(\cdot)$ and $\mathscr{H}(\cdot|\mathfrak{m})$. □

The presence of the factor δ^{-3} in the previous result should not be surprising: indeed the more we get close to the extremes $t = 0, T$ the worse we expect this bound to behave. Indeed we do not have any a priori information on the Fisher information of the two prescribed marginals μ, ν which could possibly be infinite. Moreover, since the velocities are not fixed at time $t = 0, T$, even assuming $\mathcal{I}(\mu)$, $\mathcal{I}(\nu)$ to be finite, yet we are not guaranteed a priori bounds on the Fisher information of the full marginals μ_0^T, μ_T^T.

As a byproduct of Proposition 7.4.2 and Proposition 7.4.3 we get

Corollary 7.4.4. *Under (H1), (H2) and (H4), there exists $C_{d,\alpha,\beta,\gamma} > 0$ such that for any $0 < \delta \leq 1$ and $t \in [\delta, T]$, as soon as $T > \frac{1}{\kappa}\log C_{d,\alpha,\beta,\gamma} + 2\delta$, it holds*

$$\varphi^T(t) \lesssim \delta^{-3}\, e^{-2\kappa t}\, \mathcal{C}_T(\mu,\nu) \quad \textit{and} \quad \psi^T(T-t) \lesssim \delta^{-3}\, e^{-2\kappa t}\, \mathcal{C}_T(\mu,\nu)\,. \tag{7.4.6}$$

Let us now draw a few useful consequences from the above correctors estimates. In the first result we consider the long-time behaviour of the marginals of the solution to KSP at times $t = 0, T$; in the second one we give a bound for the entropic cost, uniformly in time.

Theorem 7.4.5. *Under assumptions (H1),(H2) and (H4) there exists a positive constant $C_{d,\alpha,\beta,\gamma}$ such that for any $0 < \delta \leq 1$ and $T > \frac{1}{\kappa}\log C_{d,\alpha,\beta,\gamma} + 2\delta$ it holds*

$$\begin{aligned}\Big|\mathscr{H}(\mu_0^T|\mathfrak{m}) - \mathscr{H}(\mu|\mathfrak{m}_X)\Big| &\leq C_{d,\alpha,\beta,\gamma}\,\delta^{-3}\,\mathcal{C}_T(\mu,\nu)\,e^{-2\kappa T}\,,\\ \Big|\mathscr{H}(\mu_T^T|\mathfrak{m}) - \mathscr{H}(\nu|\mathfrak{m}_X)\Big| &\leq C_{d,\alpha,\beta,\gamma}\,\delta^{-3}\,\mathcal{C}_T(\mu,\nu)\,e^{-2\kappa T}\,.\end{aligned} \tag{7.4.7}$$

Proof. We will prove only the first bound since the second one can be proven similarly. Since $\frac{\mathrm{d}\mu}{\mathrm{d}\mathfrak{m}_X}(\cdot) = \int_{\mathbb{R}^d}\rho_0^T(\cdot,v)\mathrm{d}\mathfrak{m}_V(v)$, the log-Sobolev inequality for the Gaussian measure $\mathfrak{m}_V$ gives

$$\begin{aligned}&\mathscr{H}(\mu_0^T|\mathfrak{m}) - \mathscr{H}(\mu|\mathfrak{m}_X)\\ &= \int_{\mathbb{R}^d}\left[\int_{\mathbb{R}^d}\rho_0^T\log\rho_0^T - \left(\int_{\mathbb{R}^d}\rho_0^T\mathrm{d}\mathfrak{m}_V\right)\log\left(\int_{\mathbb{R}^d}\rho_0^T\mathrm{d}\mathfrak{m}_V\right)\,\mathrm{d}\mathfrak{m}_V\right]\mathrm{d}\mathfrak{m}_X\\ &\leq \int_{\mathbb{R}^d}\left[\int_{\mathbb{R}^d}\left|\nabla_v\sqrt{\rho_0^T}\right|^2\mathrm{d}\mathfrak{m}_V\right]\mathrm{d}\mathfrak{m}_X = \int_{\mathbb{R}^{2d}}\frac{1}{2}\left|\frac{\nabla_v\rho_0^T}{\rho_0^T}\right|^2\rho_0^T\,\mathrm{d}\mathfrak{m}\\ &= \frac{1}{2}\int_{\mathbb{R}^{2d}}\left|\nabla_v\log(f^T P_T g^T)\right|^2\rho_0^T\,\mathrm{d}\mathfrak{m}\\ &= \frac{1}{2}\int_{\mathbb{R}^{2d}}\left|\nabla_v\log f^T + \nabla_v\log P_T g^T\right|^2\rho_0^T\,\mathrm{d}\mathfrak{m}\\ &= \frac{1}{2}\int_{\mathbb{R}^{2d}}\left|\nabla_v\log g_0^T\right|^2\rho_0^T\,\mathrm{d}\mathfrak{m} \lesssim \psi^T(0) \overset{(7.4.6)}{\lesssim} \delta^{-3}\,\mathcal{C}_T(\mu,\nu)\,e^{-2\kappa T},\end{aligned}$$

where the equality in the last line follows from the fact that $f^T = f^T(x)$ does not depend on the velocity variable. Finally, since $(\mathrm{proj}_x)_\#\mu_0^T = \mu$ we know that the left hand side term above is positive. □

The above result can be seen as an entropic exponential convergence estimate for the convergence of the full-marginals at time $t = 0, T$ towards the independent couplings $\mu\otimes\mathfrak{m}_V$ and $\nu\otimes\mathfrak{m}_V$ respectively (as pointed out in Remark 7.3.3).

Since it holds $\mathscr{H}(\mu|\mathfrak{m}_X) = \mathscr{H}\left((X_0)_\#\boldsymbol{\pi}^T|(X_0)_\#\mathrm{R}_{0,T}\right) \leq \mathscr{H}\left(\boldsymbol{\pi}^T|\mathrm{R}_{0,T}\right) = \mathcal{C}_T(\mu,\nu)$, and similarly $\mathscr{H}(\nu|\mathfrak{m}_X) \leq \mathcal{C}_T(\mu,\nu)$, the following lower bound is always true

$$\mathcal{C}_T(\mu,\nu) \geq \frac{\mathscr{H}(\mu|\mathfrak{m}_X) + \mathscr{H}(\nu|\mathfrak{m}_X)}{2}\,.$$

We now give a corresponding upper bound for sufficiently large times.

Lemma 7.4.6. *Under (H1) and (H2) there exists a constant $C_{d,\alpha,\beta,\gamma} > 0$ such that for any $0 < \delta \le 1$ and $T > (\frac{1}{\kappa}\log C_{d,\alpha,\beta,\gamma} + 2\delta) \vee \big(\frac{1}{\kappa}\log\frac{C_{d,\alpha,\beta,\gamma}}{\delta^3}\big)$ it holds*

$$\mathcal{C}_T(\mu,\nu) \le C_{d,\alpha,\beta,\gamma}\left[\mathscr{H}(\mu|\mathfrak{m}_X) + \mathscr{H}(\nu|\mathfrak{m}_X)\right]. \tag{7.4.8}$$

Proof. Firstly, let us assume (H4) to hold. Owing to the bounds $|\nabla_v \log g_s^T|^2 \lesssim |\nabla \log g_s^T|^2_{M^{-1}}$ and $|\nabla_v \log f_s^T|^2 \lesssim |\nabla \log f_s^T|^2_{N^{-1}}$, from (7.2.27) it follows

$$\begin{aligned}\mathcal{C}_T(\mu,\nu) &\le \mathscr{H}(\mu_0^T|\mathfrak{m}) + \mathscr{H}(\mu_T^T|\mathfrak{m}) + \int_0^{\frac{T}{2}} \psi^T(s)\,\mathrm{d}s + \int_{\frac{T}{2}}^T \varphi^T(s)\,\mathrm{d}s\\ &\overset{(7.4.7)}{\lesssim} \mathscr{H}(\mu|\mathfrak{m}_X) + \mathscr{H}(\nu|\mathfrak{m}_X) + 2\,\delta^{-3}\,\mathcal{C}_T(\mu,\nu)\,e^{-2\kappa T} + \int_0^{\frac{T}{2}} \psi^T(s)\,\mathrm{d}s\\ &\qquad + \int_{\frac{T}{2}}^T \varphi^T(s)\,\mathrm{d}s\,.\end{aligned}$$

We first consider $\int_{T/2}^T \varphi^T(s)\,\mathrm{d}s$. For any $s \in [T/2,\, T]$ from Corollary 7.4.4 we have

$$\int_{\frac{T}{2}}^T \varphi^T(s)\,\mathrm{d}s \lesssim \delta^{-3}\,\mathcal{C}_T(\mu,\nu)\int_{\frac{T}{2}}^T e^{-2\kappa s}\,\mathrm{d}s \lesssim \delta^{-3}\,\mathcal{C}_T(\mu,\nu)\left(e^{-\kappa T} - e^{-2\kappa T}\right). \tag{7.4.9}$$

By reasoning in the same way, this time by using the fact that $s \in [0, T/2]$, we get

$$\int_0^{\frac{T}{2}} \psi^T(s)\,\mathrm{d}s \lesssim \delta^{-3}\,\mathcal{C}_T(\mu,\nu)\int_0^{\frac{T}{2}} e^{-2\kappa(T-s)}\,\mathrm{d}s \lesssim \delta^{-3}\,\mathcal{C}_T(\mu,\nu)\left(e^{-\kappa T} - e^{-2\kappa T}\right). \tag{7.4.10}$$

Therefore there exists a positive constant $C_{d,\alpha,\beta,\gamma}$ such that

$$\mathcal{C}_T(\mu,\nu) \le C_{d,\alpha,\beta,\gamma}\left[\mathscr{H}(\mu|\mathfrak{m}_X) + \mathscr{H}(\nu|\mathfrak{m}_X) + \delta^{-3}\,\mathcal{C}_T(\mu,\nu)\left(e^{-\kappa T} - e^{-2\kappa T}\right)\right],$$

which yields our thesis as soon as $T > \frac{1}{\kappa}\log\frac{C_{d,\alpha,\beta,\gamma}}{\delta^3}$ for a well chosen $C_{d,\alpha,\beta,\gamma} > 0$.

Now, let us prove the result under (H2). Firstly, notice that we may assume that μ and ν satisfy (H3), otherwise the bound is trivial. The main idea is defining the probability measures μ_n^M and ν_n^M on $\mathbb{R}^d$, approximating μ and ν, as the measures whose $\mathfrak{m}_X$-densities are given by

$$\frac{\mathrm{d}\mu_n^M}{\mathrm{d}\mathfrak{m}_X} := \left(\frac{\mathrm{d}\mu}{\mathrm{d}\mathfrak{m}_X} \wedge n\right)\frac{\mathbf{1}_{K_n}}{C_n^\mu} \qquad\text{and}\qquad \frac{\mathrm{d}\nu_n^M}{\mathrm{d}\mathfrak{m}_X} := \left(\frac{\mathrm{d}\nu}{\mathrm{d}\mathfrak{m}_X} \wedge n\right)\frac{\mathbf{1}_{K_n}}{C_n^\nu}, \tag{7.4.11}$$

where $(K_n)_{n\in\mathbb{N}}$ is an increasing sequence of compact sets in $\mathbb{R}^d$ and C_n^μ, C_n^ν are the normalising constants. Then, μ_n^M and ν_n^M satisfy (H4) and by means of (7.A.2) it follows

$$\mathscr{H}(\mu_n^M|\mathfrak{m}_X) \overset{n\to\infty}{\longrightarrow} \mathscr{H}(\mu|\mathfrak{m}_X) \qquad\text{and}\qquad \mathscr{H}(\nu_n^M|\mathfrak{m}_X) \overset{n\to\infty}{\longrightarrow} \mathscr{H}(\nu|\mathfrak{m}_X)\,. \tag{7.4.12}$$

Owing to Lemma 7.A.4 and (7.4.8) for the approximated μ_n^M, ν_n^M, we conclude our proof. □

The results given in Theorem 7.4.5 and Lemma 7.4.6 will come at hand while proving the turnpike property in Section 7.5.

Corrector estimates in the kinetic-full setting

In this section we collect results in the kinetic-full setting analogous to the ones already presented above for KSP. We omit the proofs since the arguments are very similar to the one already seen. Throughout we assume (H1), (H2) and (FH4) to be true and we will point out whenever the latter can be relaxed to (FH3)

Therefore, let us define the *correctors* as the functions $\bar{\varphi}^T, \bar{\psi}^T \colon [0,T] \to \mathbb{R}$ given by

$$\bar{\varphi}^T(s) := \int_{\mathbb{R}^{2d}} |\nabla \log \bar{f}_s^T|^2_{N^{-1}} \bar{\rho}_s^T \, \mathrm{d}\mathfrak{m} \quad \text{and} \quad \bar{\psi}^T(s) := \int_{\mathbb{R}^{2d}} |\nabla \log \bar{g}_s^T|^2_{M^{-1}} \bar{\rho}_s^T \, \mathrm{d}\mathfrak{m}\,,$$

where M, $N \in \mathbb{R}^{2d\times 2d}$ are positive definite symmetric matrices as appearing in Proposition 7.2.2. In the next result we collect all the contractive properties satisfied by the above correctors, which correspond to the ones proven for KSP in Lemma 7.4.1, Proposition 7.4.2, Proposition 7.4.3 and Corollary 7.4.4.

Lemma 7.4.7. *Grant (H1), (H2), (FH4) and fix $\delta \in (0,1]$. For any $0 < t \le s \le T$ it holds*

$$\bar{\varphi}^T(s) \le \bar{\varphi}^T(t) e^{-2\kappa\,(s-t)} \quad \textit{and} \quad \bar{\psi}^T(T-s) \le \bar{\psi}^T(T-t) e^{-2\kappa\,(s-t)}\,.$$

Moreover, for any fixed $\delta \in (0,1]$ as soon as $T > \frac{1}{\kappa}\log C_{d,\alpha,\beta,\gamma} + 2\delta$ the followings hold true

$$\begin{aligned}
\bar{\varphi}^T(t) &\lesssim e^{-2\kappa t}\left[\mathcal{I}\left(\bar{\mu}_\delta^T\right) + \mathcal{I}\left(\bar{\mu}_{T-\delta}^T\right)\right] \quad \forall t \in [\delta, T]\,,\\
\bar{\psi}^T(T-t) &\lesssim e^{-2\kappa t}\left[\mathcal{I}\left(\bar{\mu}_\delta^T\right) + \mathcal{I}\left(\bar{\mu}_{T-\delta}^T\right)\right] \quad \forall t \in [\delta, T]\,,\\
\mathcal{I}(\bar{\mu}_t^T) &\lesssim \delta^{-3}\Big(\mathcal{C}_T^F(\bar{\mu},\bar{\nu}) - \mathscr{H}(\bar{\mu}_t^T \mid \mathfrak{m})\Big) \quad \forall t \in [\delta, T-\delta]\,,
\end{aligned}$$

$$\bar{\varphi}^T(t) \lesssim \delta^{-3}\, e^{-2\kappa\, t}\, \mathcal{C}_T^F(\bar{\mu},\bar{\nu}) \;\textit{and}\; \bar{\psi}^T(T-t) \lesssim \delta^{-3}\, e^{-2\kappa\, t}\, \mathcal{C}_T^F(\bar{\mu},\bar{\nu}) \quad \forall t \in [\delta, T]\,. \tag{7.4.13a}$$

Clearly there is no kinetic-full equivalent statement of Theorem 7.4.5 since $\bar{\mu}_0^T = \bar{\mu}$ and $\bar{\mu}_T^T = \bar{\nu}$.

The analogous of Lemma 7.4.6 in the full setting can be shown under (H1) and (H2) following the same approach. Hence, it holds

Lemma 7.4.8. *Under (H1) and (H2) there exists a constant $C_{d,\alpha,\beta,\gamma} > 0$ such that for any $0 < \delta \le 1$ and $T > (\frac{1}{\kappa}\log C_{d,\alpha,\beta,\gamma} + 2\delta) \vee (\frac{1}{\kappa}\log \frac{C_{d,\alpha,\beta,\gamma}}{\delta^3})$ it holds*

$$\mathcal{C}_T^F(\bar{\mu},\bar{\nu}) \le C_{d,\alpha,\beta,\gamma}\left[\mathscr{H}(\bar{\mu}|\mathfrak{m}) + \mathscr{H}(\bar{\nu}|\mathfrak{m})\right].$$

7.5 Exponential turnpike results

One of the main contributions of this chapter are the upcoming quantitative results on the long-time behaviour of Schrödinger bridges, which imply in particular exponential convergence to $\mathfrak{m}$ when looking at timescales of order T and exponential convergence in T to the Langevin dynamics when looking at the Schrödinger bridge over a fixed time-window $[0, t]$.

Below we propose two turnpike results in which distance from equilibrium is measured through the relative entropy $\mathscr{H}(\cdot|\mathfrak{m})$ and the Fisher information $\mathcal{I}(\cdot)$. The use of $\mathscr{H}(\cdot|\mathfrak{m})$ is natural in light of the fact that the costs $\mathcal{C}_T(\mu,\nu)$ and $\mathcal{C}^F_T(\bar{\mu},\bar{\nu})$ are also relative entropies, but computed on different spaces. On the other hand, the bound on $\mathcal{I}(\cdot)$ is reminiscent of the celebrated Bakry-Émery estimates [BÉ85] and the CD gradient estimates (2.1.12).

Theorem 7.5.1 (Entropic turnpike for KSP)**.** *Grant (H1), (H2) and (H3). There exists a positive constant $C_{d,\alpha,\beta,\gamma}$ such that for any $0 < \delta \leq 1$ and $t \in [\delta,\, T-\delta]$, as soon as $T > \frac{1}{\kappa}\log C_{d,\alpha,\beta,\gamma} + 2\delta$, it holds*

$$\mathcal{I}(\mu^T_t) \leq C_{d,\alpha,\beta,\gamma}\, \delta^{-3}\, e^{-2\kappa[t\wedge(T-t)]}\, \mathcal{C}_T(\mu,\nu)\,, \tag{7.5.1}$$

$$\mathscr{H}(\mu^T_t|\mathfrak{m}) \leq C_{d,\alpha,\beta,\gamma}\, \delta^{-3}\, e^{-2\kappa[t\wedge(T-t)]}\, \mathcal{C}_T(\mu,\nu)\,. \tag{7.5.2}$$

Moreover, as soon as $T > (\frac{1}{\kappa}\log C_{d,\alpha,\beta,\gamma} + 2\delta) \vee \frac{1}{\kappa}\log\frac{C_{d,\alpha,\beta,\gamma}}{\delta^3}$, we have

$$\mathscr{H}(\mu^T_t|\mathfrak{m}) \leq C_{d,\alpha,\beta,\gamma}\, \delta^{-3}\, e^{-2\kappa[t\wedge(T-t)]}\left[\mathscr{H}(\mu|\mathfrak{m}_X) + \mathscr{H}(\nu|\mathfrak{m}_X)\right]. \tag{7.5.3}$$

Proof. We start proving the result under (H4). Since $\mathcal{I}(\mu^T_t) \lesssim \varphi^T(t) + \psi^T(t)$, the first inequality (cf. (7.5.1)) is an immediate consequence of Corollary 7.4.4. The relative entropy bound (cf. (7.5.2)) follows from the first one by means of (7.2.4). In order to extend (7.5.1) and (7.5.2) to (H3), it is enough to consider the approximation of the optimiser (cf. (7.4.4) and (7.4.5)) together with the lower semicontinuity of $\mathcal{I}(\cdot)$ and $\mathscr{H}(\cdot|\mathfrak{m})$. Finally, (7.5.3) follows from (7.5.2) by means of Lemma 7.4.6. □

Remark 7.5.2. *The bound on the Fisher information is our strongest result as it implies immediately an entropic bound thanks to the logarithmic Sobolev inequality* (7.2.4)*. Moreover, entropic bounds are stronger than bounds expressed by means of a transport distance such as* $\mathbf{W}_1$ *or* $\mathbf{W}_2$*, since* $\mathfrak{m}$ *satisfies Talagrand's inequality* (7.2.2)*.*

Theorem 7.5.3 (Entropic turnpike for KFSP)**.** *Grant (H1), (H2) and (FH3).There exists a positive constant $C_{d,\alpha,\beta,\gamma}$ such that for any $0 < \delta \leq 1$ and $t \in [\delta,\, T-\delta]$, as soon as $T > \frac{1}{\kappa}\log C_{d,\alpha,\beta,\gamma} + 2\delta$, it holds*

$$\mathcal{I}(\bar{\mu}^T_t) \leq C_{d,\alpha,\beta,\gamma}\, \delta^{-3}\, e^{-2\kappa[t\wedge(T-t)]}\, \mathcal{C}^F_T(\bar{\mu},\bar{\nu})\,,$$

$$\mathscr{H}(\bar{\mu}_t^T|\mathfrak{m}) \leq C_{d,\alpha,\beta,\gamma}\,\delta^{-3}\,e^{-2\kappa[t\wedge(T-t)]}\,\mathcal{C}_T^F(\bar{\mu},\bar{\nu})\,. \tag{7.5.4}$$

Moreover, as soon as $T > (\frac{1}{\kappa}\log C_{d,\alpha,\beta,\gamma} + 2\delta) \vee \frac{1}{\kappa}\log\frac{C_{d,\alpha,\beta,\gamma}}{\delta^3}$, *we have*

$$\mathscr{H}(\bar{\mu}_t^T|\mathfrak{m}) \leq C_{d,\alpha,\beta,\gamma}\,\delta^{-3}\,e^{-2\kappa[t\wedge(T-t)]}\left[\mathscr{H}(\bar{\mu}|\mathfrak{m}) + \mathscr{H}(\bar{\nu}|\mathfrak{m})\right]. \tag{7.5.5}$$

Proof. The proof of the result under (FH4) follows the same reasoning presented in the first part of the proof of Theorem 7.5.1 and for this reason is omitted. An approximating argument akin to the one in the proof of Theorem 7.5.1, this time considering the full marginals $\bar{\mu}^n, \bar{\nu}^n$ and the corresponding KFSP, gives

$$\bar{\mu}_t^{n,T} \rightharpoonup \bar{\mu}_t^T \qquad \text{and} \qquad \mathcal{C}_T^F(\bar{\mu}^n,\bar{\nu}^n) \to \mathcal{C}_T^F(\bar{\mu},\bar{\nu})\,.$$

Therefore the first two bounds follow from the lower semicontinuity of $\mathcal{I}(\cdot)$ and $\mathscr{H}(\cdot|\mathfrak{m})$. Finally, (7.5.5) follows from (7.5.4) by means of Lemma 7.4.8. □

Finally, let us provide quantitative exponential versions of Theorem 7.3.2 and Theorem 7.3.4. The key ingredient in the proof of the exponential estimates is the representation formula for the difference

$$\mathcal{C}_T(\mu,\nu) - \mathscr{H}(\mu|\mathfrak{m}_X) - \mathscr{H}(\nu|\mathfrak{m}_X)$$

that we have established in Lemma 7.2.5 and allows to profit from the turnpike estimates in Theorem 7.5.1 and Theorem 7.5.3.

Theorem 7.5.4. *Grant (H1), (H2) and (H3). Then there exists a positive constant* $C_{d,\alpha,\beta,\gamma}$ *(depending only on* d,α,β *and* γ*) such that for any* $0 < \delta \leq 1$*, as soon as* $T > (\frac{1}{\kappa}\log C_{d,\alpha,\beta,\gamma} + 2\delta) \vee \frac{1}{\kappa}\log\frac{C_{d,\alpha,\beta,\gamma}}{\delta^3}$*, it holds*

$$\begin{aligned}|\mathcal{C}_T(\mu,\nu) - \mathscr{H}(\mu|\mathfrak{m}_X) - &\mathscr{H}(\nu|\mathfrak{m}_X)| \\ &\leq C_{d,\alpha,\beta,\gamma}\,\delta^{-3}\,e^{-\kappa T}\left[\mathscr{H}(\mu|\mathfrak{m}_X) + \mathscr{H}(\nu|\mathfrak{m}_X)\right]\end{aligned} \tag{7.5.6}$$

and as a consequence the following entropic Talagrand inequality holds

$$\mathcal{C}_T(\mu,\nu) \leq \left(1 + C_{d,\alpha,\beta,\gamma}\,\delta^{-3}\,e^{-\kappa T}\right)\left[\mathscr{H}(\mu|\mathfrak{m}_X) + \mathscr{H}(\nu|\mathfrak{m}_X)\right].$$

Proof. Firstly, assume (H4) to hold. By (7.2.27) and owing to the trivial bounds $|\nabla_v \log g_s^T|^2 \lesssim |\nabla \log g_s^T|^2_{M^{-1}}$ and $|\nabla_v \log f_s^T|^2 \lesssim |\nabla \log f_s^T|^2_{N^{-1}}$, we know that

$$\begin{aligned}\left|\mathcal{C}_T(\mu,\nu) - \mathscr{H}(\mu_0^T|\mathfrak{m}) - \mathscr{H}(\mu_T^T|\mathfrak{m})\right| \lesssim \mathscr{H}(\mu_{\frac{T}{2}}^T|\mathfrak{m}) + \int_0^{\frac{T}{2}}\psi^T(s)\,\mathrm{d}s \\ + \int_{\frac{T}{2}}^T \varphi^T(s)\,\mathrm{d}s,\end{aligned}$$

and from (7.4.9), (7.4.10) and the entropic turnpike (7.5.2) it follows

$$\left|\mathcal{C}_T(\mu,\nu) - \mathscr{H}(\mu_0^T|\mathfrak{m}) - \mathscr{H}(\mu_T^T|\mathfrak{m})\right| \lesssim \delta^{-3}\, e^{-\kappa T}\, \mathcal{C}_T(\mu,\nu) + \delta^{-3}\, \mathcal{C}_T(\mu,\nu)\, e^{-\kappa T}.$$

As a byproduct of the above inequality and Theorem 7.4.5 we get

$$|\mathcal{C}_T\left(\mu,\nu\right) - \mathscr{H}\left(\mu|\mathfrak{m}_X\right) - \mathscr{H}\left(\nu|\mathfrak{m}_X\right)| \le C_{d,\alpha,\beta,\gamma}\, \delta^{-3}\, e^{-\kappa T}\, \mathcal{C}_T(\mu,\nu), \qquad (7.5.7)$$

which combined with Lemma 7.4.6 proves (7.5.6) under (H4).

Let us now assume that μ and ν satisfy (H3) only. Firstly, consider the approximating sequence $(\boldsymbol{\pi}_n^T)_{n\in\mathbb{N}}$ of the optimiser (cf. (7.4.4) and (7.4.5)). Then, by means of (7.5.7) under (H4) and the lower semicontinuity of the relative entropy we have

$$\begin{aligned}
&\mathscr{H}\left(\mu|\mathfrak{m}_X\right) + \mathscr{H}\left(\nu|\mathfrak{m}_X\right) - \mathcal{C}_T\left(\mu,\nu\right)\\
&\overset{(7.4.5)}{\le} \liminf_{n\to\infty}\Big[\mathscr{H}\left(\mu^n|\mathfrak{m}_X\right) + \mathscr{H}\left(\nu^n|\mathfrak{m}_X\right) - \mathcal{C}_T\left(\mu^n,\nu^n\right)\Big]\\
&\overset{(7.5.7)}{\lesssim} \delta^{-3}e^{-\kappa T}\liminf_{n\to\infty}\mathcal{C}_T(\mu^n,\nu^n) = \delta^{-3}e^{-\kappa T}\,\mathcal{C}_T(\mu,\nu)\\
&\overset{(7.4.8)}{\lesssim} \delta^{-3}\, e^{-\kappa T}\left[\mathscr{H}\left(\mu|\mathfrak{m}_X\right) + \mathscr{H}\left(\nu|\mathfrak{m}_X\right)\right].
\end{aligned}$$

For the other bound we are going to use the approximation on the marginals (cf. (7.4.11)). Therefore, let us consider μ_n^M, ν_n^M such that (H4) holds. Then, from Lemma 7.A.4 and the convergence of the relative entropies in (7.4.12), we get

$$\begin{aligned}
&\mathcal{C}_T\left(\mu,\nu\right) - \mathscr{H}\left(\mu|\mathfrak{m}_X\right) - \mathscr{H}\left(\nu|\mathfrak{m}_X\right)\\
&\qquad\le \liminf_{n\to\infty}\Big[\mathcal{C}_T(\mu_n^M,\nu_n^M) - \mathscr{H}(\mu_n^M|\mathfrak{m}_X) - \mathscr{H}(\nu_n^M|\mathfrak{m}_X)\Big]\\
&\qquad\overset{(7.5.7)}{\lesssim} \delta^{-3}\, e^{-\kappa T}\, \liminf_{n\to\infty}\mathcal{C}_T\left(\mu_n^M,\nu_n^M\right)\\
&\qquad\overset{(7.4.8)}{\lesssim} \delta^{-3}\, e^{-\kappa T}\, \liminf_{n\to\infty}\left[\mathscr{H}\left(\mu_n^M|\mathfrak{m}_X\right) + \mathscr{H}\left(\nu_n^M|\mathfrak{m}_X\right)\right]\\
&\qquad= \delta^{-3}\, e^{-\kappa T}\left[\mathscr{H}\left(\mu|\mathfrak{m}_X\right) + \mathscr{H}\left(\nu|\mathfrak{m}_X\right)\right].
\end{aligned}$$

□

Theorem 7.5.5. *Grant (H1), (H2) and (FH3). Then there exists a positive constant $C_{d,\alpha,\beta,\gamma}$ such that for any $0<\delta\le 1$, as soon as $T > (\frac{1}{\kappa}\log C_{d,\alpha,\beta,\gamma} + 2\delta) \vee \frac{1}{\kappa}\log\frac{C_{d,\alpha,\beta,\gamma}}{\delta^3}$, it holds*

$$\left|\mathcal{C}_T^F\left(\bar\mu,\bar\nu\right) - \mathscr{H}\left(\bar\mu|\mathfrak{m}\right) - \mathscr{H}\left(\bar\nu|\mathfrak{m}\right)\right| \le C_{d,\alpha,\beta,\gamma}\, \delta^{-3}\, e^{-\kappa T}\left[\mathscr{H}\left(\bar\mu|\mathfrak{m}\right) + \mathscr{H}\left(\bar\nu|\mathfrak{m}\right)\right],$$

and as a consequence the following entropic Talagrand inequality holds

$$\mathcal{C}_T^F(\bar{\mu},\bar{\nu}) \leq \Big(1 + C_{d,\alpha,\beta,\gamma}\,\delta^{-3}\,e^{-\kappa T}\Big)\Big[\mathscr{H}(\bar{\mu}|\mathfrak{m}) + \mathscr{H}(\bar{\nu}|\mathfrak{m})\Big]. \tag{7.5.8}$$

Proof. By means of (7.2.29), the corrector estimates (7.4.13a), the turnpike estimate (7.5.4) and Lemma 7.4.8, at least under (FH4), it follows that for any $0 < \delta \leq 1$, as soon as $T > (\frac{1}{\kappa}\log C_{d,\alpha,\beta,\gamma} + 2\delta) \vee \frac{1}{\kappa}\log\frac{C_{d,\alpha,\beta,\gamma}}{\delta^3}$,

$$\Big|\mathcal{C}_T^F(\bar{\mu},\bar{\nu}) - \mathscr{H}(\bar{\mu}|\mathfrak{m}) - \mathscr{H}(\bar{\nu}|\mathfrak{m})\Big| \leq C_{d,\alpha,\beta,\gamma}\,\delta^{-3}\,e^{-\kappa T}\Big[\mathscr{H}(\bar{\mu}|\mathfrak{m}) + \mathscr{H}(\bar{\nu}|\mathfrak{m})\Big],$$

and from this immediately deduce (7.5.8). The extension to (FH3) is a consequence of a standard approximation argument. □

7.5.1 Wasserstein convergence over a fixed time-window

In this section we show in Theorem 7.5.6 that the entropic interpolations for KSP and KFSP enjoy a turnpike property with respect to the Wasserstein distance.

Notice that a turnpike property in the Wasserstein distance could be deduced from the entropic one (cf. Theorem 7.5.1 and Theorem 7.5.3) by means of the Talagrand inequality (7.2.2). However, below we provide a different proof that is of independent interest for two reasons. Firstly, the inequality below holds for any $t \in [0,T]$, while the entropic turnpike is restricted to the subinterval $[\delta, T-\delta]$. Secondly, the argument in the proof, which uses the optimal control formulation of the Schrödinger problem, will be instrumental for the study of the short-time behaviour of the Schrödinger bridge.

Theorem 7.5.6 (Wasserstein turnpike)**.** *Under hypotheses (H1), (H2) and (H3), there exists a positive constant $C_{d,\alpha,\beta,\gamma}$ such that for any $0 < \delta \leq 1$, as soon as $T > \frac{1}{\kappa}\log C_{d,\alpha,\beta,\gamma} + 2\delta$, for any $t \in [0,T]$ it holds*

$$\mathbf{W}_2(\mu_t^T, \mathfrak{m}) \leq C_{d,\alpha,\beta,\gamma}\,\delta^{-\frac{3}{2}}\,e^{-\kappa\,[t\wedge(T-t)]}\,\sqrt{\mathcal{C}_T(\mu,\nu)}\,.$$

Proof. Let us firstly assume (H4). We we will prove our result for the distorted Wasserstein distance $\mathbf{W}_{M,2}$ induced by the metrics $|\cdot|_M$. Fix $\delta \in (0,1)$ and assume $t \in [0, T-\delta]$. Define $\tilde{\mu}_\cdot^T$ as the marginal flow generated by the uncontrolled process $Z_s^{\mathbf{0},T} := (X_s^{\mathbf{0},T}, V_s^{\mathbf{0},T})_{s\in[0,T]}$ solution of (7.0.2) started at the initial distribution $\mu_0^T \in \mathcal{P}(\mathbb{R}^{2d})$. Then, since $\tilde{\mu}_0^T = \mu_0^T$, it holds

$$\begin{aligned}\mathbf{W}_{M,2}(\mu_t^T,\mathfrak{m}) &\leq \mathbf{W}_{M,2}(\mu_t^T,\tilde{\mu}_t^T) + \mathbf{W}_{M,2}(\tilde{\mu}_t^T,\mathfrak{m})\\ &\overset{(7.2.14)}{\leq} \mathbf{W}_{M,2}(\mu_t^T,\tilde{\mu}_t^T) + e^{-\kappa t}\,\mathbf{W}_{M,2}(\mu_0^T,\mathfrak{m})\,,\end{aligned} \tag{7.5.9}$$

The second term in the right hand side can be handled with the Talagrand inequality:

$$\mathbf{W}_{M,2}(\mu_0^T,\mathfrak{m}) \lesssim \mathbf{W}_2(\mu_0^T,\mathfrak{m}) \overset{(7.2.2)}{\lesssim} \sqrt{\mathscr{H}(\mu_0^T|\mathfrak{m})} \leq \sqrt{\mathcal{C}_T(\mu,\nu)}\,,$$

where the last step holds since $\mathcal{C}_T(\mu,\nu) \geq \mathscr{H}\left((\mathrm{proj}_{x_1})_{\#}\boldsymbol{\pi}^T|(\mathrm{proj}_{x_1})_{\#}\mathrm{R}_{0,T}\right)$.

Let us now focus on $\mathbf{W}_{M,2}(\mu_t^T,\tilde{\mu}_t^T)$. We will use a synchronous coupling between these two measures. Therefore let us introduce the process $Z_s^{\boldsymbol{u},T} := (X_s^{\boldsymbol{u},T},\,V_s^{\boldsymbol{u},T}) \sim \mu_s^T$, *i.e.*, the solution of (7.1.2) (driven by the same Brownian motion for $Z_s^{\mathbf{0},T}$) when considering the control $\boldsymbol{u}_s = 2\gamma\,\nabla_v \log g_s^T(X_s^{\boldsymbol{u},T},V_s^{\boldsymbol{u},T})$. Particularly, from (7.2.22) it follows that $\boldsymbol{u}$ is the optimal control and $Z_s^{\boldsymbol{u},T} \sim \mu_s^T$. For notation's sake set $Z_s^{\boldsymbol{\Delta},T} := Z_s^{\boldsymbol{u},T} - Z_s^{\mathbf{0},T}$. Then it holds

$$\mathrm{d}Z_s^{\boldsymbol{\Delta},T} = \left[b\left(Z_s^{\boldsymbol{u},T}\right) - b\left(Z_s^{\mathbf{0},T}\right)\right]\mathrm{d}s + \begin{pmatrix}0\\ \boldsymbol{u}_s\end{pmatrix}\mathrm{d}s\,,$$

where $b(z)$ denotes the drift of the Langevin dynamics (7.0.2). By Itô's Formula we obtain

$$\begin{aligned}
\mathrm{d}\left|Z_s^{\boldsymbol{\Delta},T}\right|_M^2 &= 2MZ_s^{\boldsymbol{\Delta}}\cdot\left(b(Z_s^{\boldsymbol{u}}) - b(Z_s^{\mathbf{0}})\right)\mathrm{d}s + 2MZ_s^{\boldsymbol{\Delta}}\cdot\begin{pmatrix}0\\ \boldsymbol{u}_s\end{pmatrix}\mathrm{d}s\\
&= 2\int_0^1 Z_s^{\boldsymbol{\Delta},T}\cdot MJ_b\left(rZ_s^{\boldsymbol{u},T} + (1-r)Z_s^{\mathbf{0},T}\right)Z_s^{\boldsymbol{\Delta},T}\mathrm{d}r\mathrm{d}s + 2MZ_s^{\boldsymbol{\Delta},T}\cdot\begin{pmatrix}0\\ \boldsymbol{u}_s\end{pmatrix}\mathrm{d}s\\
&\leq -2\,\kappa\left|Z_s^{\boldsymbol{\Delta},T}\right|_M^2\mathrm{d}s + 2MZ_s^{\boldsymbol{\Delta},T}\cdot\begin{pmatrix}0\\ \boldsymbol{u}_s\end{pmatrix}\mathrm{d}s\,,
\end{aligned}$$

where the last inequality follows from (7.2.11). By taking the expectation, and applying Hölder's inequality we get

$$\frac{\mathrm{d}}{\mathrm{d}s}\,\mathbb{E}_{\mathrm{R}}\left[\left|Z_s^{\boldsymbol{\Delta},T}\right|_M^2\right] \leq -2\,\kappa\,\mathbb{E}_{\mathrm{R}}\left[\left|Z_s^{\boldsymbol{\Delta},T}\right|_M^2\right] + 2\mathbb{E}_{\mathrm{R}}\left[\left|Z_s^{\boldsymbol{\Delta},T}\right|_M^2\right]^{\frac12}\mathbb{E}_{\mathrm{R}}\left[\left|(0,\boldsymbol{u}_s)^T\right|_M^2\right]^{\frac12}.$$

Therefore it holds

$$\frac{\mathrm{d}}{\mathrm{d}s}\sqrt{\mathbb{E}_{\mathrm{R}}\left[|Z_s^{\boldsymbol{\Delta},T}|_M^2\right]} \leq -\kappa\sqrt{\mathbb{E}_{\mathrm{R}}\left[|Z_s^{\boldsymbol{\Delta},T}|_M^2\right]} + \mathbb{E}_{\mathrm{R}}\left[\left|(0,\boldsymbol{u}_s)^T\right|_M^2\right]^{\frac12}.$$

Recalling that the optimal control is given by $\boldsymbol{u}_s = 2\gamma\,\nabla_v\log g_s^T(X_s^{\boldsymbol{u},T},V_s^{\boldsymbol{u},T})$ we obtain that

$$\frac{\mathrm{d}}{\mathrm{d}s}\sqrt{\mathbb{E}_{\mathrm{R}}\left[|Z_s^{\boldsymbol{\Delta},T}|_M^2\right]} \lesssim \left(\int_{\mathbb{R}^{2d}}\left|\nabla_v\log g_s^T\right|^2\rho_s^T\,\mathrm{d}\mathfrak{m}\right)^{\frac12} \lesssim \psi^T(s)^{\frac12}\,.$$

Therefore, by integrating over $s\in[0,\,t]$ we get

$$\sqrt{\mathbb{E}_{\mathrm{R}}\left[|Z_t^{\boldsymbol{\Delta},T}|_M^2\right]} = \int_0^t\frac{\mathrm{d}}{\mathrm{d}s}\sqrt{\mathbb{E}_{\mathrm{R}}\left[|Z_s^{\boldsymbol{\Delta},T}|_M^2\right]}\,\mathrm{d}s \overset{(7.4.6)}{\lesssim} \delta^{-\frac32}\,e^{-\kappa(T-t)}\sqrt{\mathcal{C}_T(\mu,\nu)}\,,$$

and hence it holds

$$\mathbf{W}_{M,2}(\mu_t^T, \tilde{\mu}_t^T) \lesssim \delta^{-\frac{3}{2}}\, e^{-\kappa\,(T-t)}\, \sqrt{\mathcal{C}_T(\mu,\nu)}\,. \tag{7.5.10}$$

Then, from (7.5.9) we deduce that for any $t \in [0, T-\delta]$ it holds

$$\begin{aligned}\mathbf{W}_{M,2}(\mu_t^T, \mathfrak{m}) &\lesssim \delta^{-\frac{3}{2}}\, e^{-\kappa\,(T-t)}\, \sqrt{\mathcal{C}_T(\mu,\nu)} + e^{-\kappa\, t}\, \sqrt{\mathcal{C}_T(\mu,\nu)} \\ &\lesssim \delta^{-\frac{3}{2}}\, e^{-\kappa\,[t\wedge(T-t)]}\, \sqrt{\mathcal{C}_T(\mu,\nu)}\,.\end{aligned}$$

By considering the contraction along P^*, the same argument gives us the same bound for $t \in [\delta, T]$ and therefore on the whole domain $[0, T]$.

In order to relax the assumption to (H3), it is enough to consider once again the approximation of the optimiser (as in the proof of Theorem 7.5.1) together with the lower semicontinuity of the Wasserstein distance. □

The previous argument can also be applied in order to analyse the behaviour of entropic interpolations for a fixed time t, while T grows large. More precisely, we show that the (uncontrolled) Langevin dynamics and the Schrödinger bridge are exponentially close in the long-time regime $T \to \infty$, for all time-windows $[0, t]$. Note that this result cannot be deduced from the turnpike estimates of the former section.

Theorem 7.5.7. *Under hypotheses (H1), (H2) and (H3), there exists a positive constant $C_{d,\alpha,\beta,\gamma}$ such that for any $0 < \delta \le 1$ and $t \in [0, T-\delta]$, as soon as $T > \frac{1}{\kappa}\log C_{d,\alpha,\beta,\gamma} + 2\delta$, it holds*

$$\mathcal{W}_2(\mu_t^T, \mu_t^\infty) \le C_{d,\alpha,\beta,\gamma}\, \delta^{-\frac{3}{2}}\, e^{-\kappa(T-t)}\, \sqrt{\mathcal{C}_T(\mu,\nu)}\,,$$

where μ_t^∞ is the law of $(X_\cdot, V_\cdot)$ satisfying

$$\begin{cases} \mathrm{d}X_t = V_t \mathrm{d}t, \\ \mathrm{d}V_t = -\nabla U(X_t)\mathrm{d}t - \gamma V_t \mathrm{d}t + \sqrt{2\gamma}\, \mathrm{d}B_t, \\ (X_0, V_0) \sim \mu \otimes \mathfrak{m}_V. \end{cases} \tag{7.5.11}$$

Proof. At first, let us assume (H4). We have

$$\begin{aligned}\mathbf{W}_2\left(\mu_t^T, \mu_t^\infty\right) &\le \mathbf{W}_2\left(\mu_t^T, \tilde{\mu}_t^T\right) + \mathbf{W}_2\left(\tilde{\mu}_t^T, \mu_t^\infty\right) \\ &\overset{(7.5.10),(7.2.14)}{\lesssim} \delta^{-\frac{3}{2}}\, e^{-\kappa\,(T-t)}\, \sqrt{\mathcal{C}_T(\mu,\nu)} + e^{-\kappa\, t}\, \mathbf{W}_2\left(\mu_0^T, \mu \otimes \mathfrak{m}_V\right),\end{aligned} \tag{7.5.12}$$

where $\tilde{\mu}_\cdot^T$ is the marginal flow defined in the previous proof, *i.e.*, the flow generated by the uncontrolled process $(X_s^{\mathbf{0},T}, V_s^{\mathbf{0},T})_{s\in[0,T]}$ started at the initial distribution $\mu_0^T \in \mathcal{P}(\mathbb{R}^{2d})$. Using the inequality

$$\mathbf{W}_2\left(\mu_0^T, \mu \otimes \mathfrak{m}_V\right)^2 \le \int_{\mathbb{R}^d} \mathbf{W}_2\left(\mu_0^T(\cdot|x), \mathfrak{m}_V\right)^2 \mathrm{d}\mu(x),$$

applying Talagrand's inequality for $\mathfrak{m}_V$ and using the additive property of relative entropy (1.A.4), we obtain

$$\begin{aligned}
\mathbf{W}_2\left(\mu_0^T, \mu \otimes \mathfrak{m}_V\right)^2 &\leq 2 \int_{\mathbb{R}^d} \mathscr{H}\left(\mu_0^T(\cdot|x)|\mathfrak{m}_V\right) \mathrm{d}\mu(x) = 2\,\mathscr{H}\left(\mu_0^T|\mu \otimes \mathfrak{m}_V\right) \\
&= 2\,\mathscr{H}(\mu_0^T|\mathfrak{m}) - 2\int_{\mathbb{R}^{2d}} \log \frac{\mathrm{d}\,(\mu \otimes \mathfrak{m}_V)}{\mathrm{d}\mathfrak{m}}\, \mathrm{d}\mu_0^T \\
&= 2\,\mathscr{H}(\mu_0^T|\mathfrak{m}) - 2\int_{\mathbb{R}^d} \log \frac{\mathrm{d}\mu}{\mathrm{d}\mathfrak{m}_X}\, \mathrm{d}\mu \\
&= 2\,\mathscr{H}(\mu_0^T|\mathfrak{m}) - 2\,\mathscr{H}(\mu|\mathfrak{m}_X) \overset{(7.4.7)}{\lesssim} \delta^{-3}\, \mathcal{C}_T(\mu,\nu)\, e^{-2\kappa T}\,.
\end{aligned}$$

By combining the above inequalities with (7.5.12) we get our result.

The extension of the result to the weaker (H3) follows from the same approximating argument discussed in the previous proof. □

With a similar reasoning one can prove that the Wasserstein turnpike holds also for KFSP under (FH3). Notice that since in this setting we fix the whole marginals at time 0 and T, it holds $\bar{\mu}_0^T = \bar{\mu}$ and $\bar{\mu}_T^T = \bar{\nu}$ and therefore in this case we do not need a result similar to Theorem 7.4.5. Therefore we have the following

Theorem 7.5.8 (Wasserstein turnpike)**.** *Under hypotheses (H1),(H2) and (FH3), there exists a positive constant $C_{d,\alpha,\beta,\gamma}$ such that for any $0 < \delta \leq 1$, as soon as $T > \frac{1}{\kappa} \log C_{d,\alpha,\beta,\gamma} + 2\delta$, for any $t \in [0,T]$ it holds*

$$\mathbf{W}_2(\bar{\mu}_t^T, \mathfrak{m}) \leq C_{d,\alpha,\beta,\gamma}\, \delta^{-\frac{3}{2}}\, e^{-\kappa\,[t\wedge(T-t)]} \sqrt{\mathcal{C}_T^F\,(\bar{\mu},\bar{\nu})}\,.$$

Similarly, we can easily prove a statement for KFSP equivalent to Theorem 7.5.7 where Assumption (H3) is replaced with (FH3) and where the initial condition in (7.5.11) is just the full fixed marginal $\bar{\mu}$.

Theorem 7.5.9. *Under hypotheses (H1),(H2) and (FH3), there exists a positive constant $C_{d,\alpha,\beta,\gamma}$ such that for any $0 < \delta \leq 1$ and $t \in [0, T-\delta]$, as soon as $T > \frac{1}{\kappa} \log C_{d,\alpha,\beta,\gamma} + 2\delta$, it holds*

$$\mathcal{W}_2(\bar{\mu}_t^T, \bar{\mu}_t^\infty) \leq C_{d,\alpha,\beta,\gamma}\, \delta^{-\frac{3}{2}}\, e^{-k(T-t)} \sqrt{\mathcal{C}_T^F\,(\bar{\mu},\bar{\nu})}$$

where $\bar{\mu}_t^\infty$ is the law of $(X_t,\, V_t)$ satisfying

$$\begin{cases}
\mathrm{d}X_t = V_t \mathrm{d}t, \\
\mathrm{d}V_t = -\nabla U(X_t)\mathrm{d}t - \gamma V_t \mathrm{d}t + \mathrm{d}B_t, \\
(X_0, V_0) \sim \bar{\mu}.
\end{cases}$$

Bibliographical Remarks

The results presented in this chapter come from [CCGR22].

Given that Schrödinger's thought experiment is motivated by statistical mechanics and the physical relevance of the (underdamped) Langevin dynamics and its various applications, the study of the kinetic Schrödinger problem appears to be quite natural. Nevertheless, to the best of our knowledge, it seems that there has been no dedicated study so far, with the exception of [CGP15]. The objective of [CCGR22], from which this chapter is based, is to take some steps forward in this direction, in particular by gaining a quantitative understanding of optimal solutions, namely the Schrödinger bridges.

Proving the turnpike property for Schrödinger bridges in this context is harder than in the classical setting, and we need to work under stronger assumptions on the potential U than its strong convexity. This is not a surprise. Indeed, proving the exponential convergence to equilibrium for the kinetic Fokker-Planck equation is a difficult problem that has been, and still is, intensively studied by means of either a probabilistic or an analytic approach, see [CGMZ19, EGZ19a, Tal02, GLWZ21] for some references on the probabilistic approach. Following the terminology introduced by Villani in his monograph [Vil09], this obstruction is a manifestation of the *hypocoercive* nature of the kinetic Fokker-Planck equation. KSP may indeed be regarded as the prototype of an hypocoercive stochastic control problem. For the moment, we have been able to show the turnpike property under a quasilinearity assumption. The key assumption for obtaining (7.5.2) and (7.5.4) is (H2), asking U to be strongly convex and such that the difference between the smallest and largest eigenvalues of $\nabla^2 U(x)$ is controlled by the friction parameter γ uniformly in x. Assumptions of this type, where the friction parameter has to be in some sense large in comparison with the spectrum of $\nabla^2 U$ are commonly encountered in the literature. In the language of probability, they ensure that the synchronous coupling is contracting for the Langevin dynamics [BGM10, Mon23]. On the other hand, from an analytical standpoint, Assumption (H2) implies *local* gradient bounds for the semigroup generated by the Langevin dynamics [Bau17]. Finally, we recall that the exponential rate κ of Theorems 7.5.1 and 7.5.3 is precisely the one, computed e.g. in [Mon23, BGM10], at which synchronous coupling is contractive for the (uncontrolled) Langevin dynamics.

Although exponential L^2 estimates are known to hold under considerably weaker assumptions (see e.g. [HN04] and [CHSG22, HM19] for singular potentials), and entropic estimates assuming a bounded and positive Hessian have been known for more than a decade [Vil09], it is only recently [GLWZ21] that entropic estimates have been obtained beyond the bounded Hessian case. In light of this, the question of how to improve our results is quite interesting and deserves to be further investigated.

Finally, if we compare our results with what is known in deterministic control we remark that, quite curiously, exponential estimates for the deterministic noiseless version of (7.1.1), obtained removing the Brownian motion form the controlled state equation, do not seem to be covered from existing results, even in the case when μ and ν are Dirac measures.[1] For linear-quadratic problems though, the result is well known, see e.g. [BP20] for precise estimates. Theorem 7.5.1 and Theorem 7.5.3 provide *global* turnpike estimates, that is to say we do not ask μ and ν to be close to $\mathfrak{m}$. We do ask $\mathscr{H}(\mu|\mathfrak{m}_X), \mathscr{H}(\nu|\mathfrak{m}_X) < +\infty$, but this condition is very mild and necessary for the Schrödinger problem to have a finite value. This is in contrast with most exponential turnpike estimates we are aware of in deterministic control (see e.g. [TZ15, Theorem 1]). The passage from local to global estimates seems to be possible [TZ18, Tré23] under some extra assumptions, such as the existence of a storage function, but this comes at the price of losing quite some information on the multiplicative constants appearing in (7.5.2). Moreover, the condition $T > \frac{1}{\kappa} \log C_{d,\alpha,\beta,\gamma} + 2\delta$ of Theorem 7.5.1 should be replaced with a condition of the form $T > T_0$ with T_0 depending on the initial conditions and potentially very large.

[1] For example, if we compare with the reference work [TZ15], the matrix W defined at Eq. (10) therein would not be invertible for the problem under consideration, which thus fails to satisfy the hypothesis of the main turnpike result obtained there.

This page was intentionally left blank.

Appendix 7

7.A From compact support to finite entropy

In this appendix we discuss two types of approximating sequences that we have used in order to extend our main results from (H4) to (H3).

In Section 7.A.1 we deal with the *Approximation of the optimiser* where we are able to prove the convergence of the entropic cost of the approximated problem to the original entropic cost but not the convergence of the associated entropies of the marginals at time $t = 0, T$. On the other hand in Section 7.A.2, we investigate the *Approximation of the marginals*, by approximating directly the marginals and consider the associated Schr̈odinger problems. In this case, we get the convergence of the marginals' relative entropies, but not the one of the entropic cost. The two aforementioned strategies produce complementary bounds which can be applied together in order to relax the assumptions from (H4) to (H3).

We will deal exclusively with the approximations and proofs for KSP and omit those for KFSP, since the latter can be treated in the same way.

7.A.1 Approximating the optimiser

Fix a couple of marginals $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ satisfying (H3). We already know that there exists a unique minimiser $\boldsymbol{\pi}^T \in \Pi_X(\mu, \nu)$ for KSP, with $\mathrm{R}_{0,T}$-density given by ρ^T. Now consider an increasing sequence of rectangular compact sets $(K_n)_{n\in\mathbb{N}}$ in $\mathbb{R}^{4d}$ whose union gives the whole space. For each $n \in \mathbb{N}$ define the probability measure q_n^T as the measure whose $\mathrm{R}_{0,T}$-density is given by

$$\hat{\rho}_n^T = \frac{\mathrm{d}q_n^T}{\mathrm{dR}_{0,T}} := \left(\rho^T \wedge n\right) \frac{\mathbf{1}_{K_n}}{C_n},$$

where $C_n := \int_{K_n} (\rho^T \wedge n)\, \mathrm{dR}_{0,T}$ is the normalising constant. Notice that $C_n \uparrow 1$ by monotone convergence and $\hat{\rho}_n^T \to \rho^T$. For convenience, we fix in this section some $\bar{n} \in \mathbb{N}$ such that $C_n \geq 1/2$ for any $n \geq \bar{n}$.

Lemma 7.A.1. *The following properties hold true.*

(i) The marginals $\mu^n := (\mathrm{proj}_{x_1})_\# q_n^T$ *and* $\nu^n := (\mathrm{proj}_{x_2})_\# q_n^T$ *satisfy (H4).*

(ii) $q_n^T \rightharpoonup \boldsymbol{\pi}^T$.

(iii) $\mathscr{H}\left(q_n^T|\mathrm{R}_{0,T}\right) \to \mathscr{H}\left(\boldsymbol{\pi}^T|\mathrm{R}_{0,T}\right) = \mathcal{C}_T(\mu,\nu)$.

Proof. We start with i). Since q_n^T has compact support, so do its marginals μ^n, ν^n. Moreover if $B \subseteq \mathbb{R}^d$ is a Borel set, then

$$\mu^n(B) = q_n^T(B\times\mathbb{R}^{3d}) \le \frac{n}{C_n}\int_{B\times\mathbb{R}^{3d}} \mathrm{dR}_{0,T} = \frac{n}{C_n}\,\mathrm{R}_{0,T}(B\times\mathbb{R}^{3d}) = \frac{n}{C_n}\,\mathfrak{m}_X(B)$$

and therefore $\|\mathrm{d}\mu^n/\mathrm{d}\mathfrak{m}_X\|_{\mathrm{L}^\infty(\mathfrak{m}_X)} \le \frac{n}{C_n}$. The same reasoning applies also to ν^n.

The weak convergence in (ii) follows from dominated convergence.

Let us prove point (iii). Notice that for each $n \ge \bar n$ it holds

$$\left|\hat\rho_n^T \log\hat\rho_n^T\right| \le \max\left\{e^{-1},\, (\rho^T C_{\bar n}^{-1})\,\log(\rho^T C_{\bar n}^{-1})\right\},$$

and the above RHS is $\mathrm{R}_{0,T}$-integrable since it holds

$$\int_{\mathbb{R}^{4d}} (\rho^T C_{\bar n}^{-1})\,\log(\rho^T C_{\bar n}^{-1})\,\mathrm{dR}_{0,T} = \frac{1}{C_{\bar n}}\,\mathcal{C}_T(\mu,\nu) + \frac{1}{C_{\bar n}}\log\left(\frac{1}{C_{\bar n}}\right) < \infty,$$

which is finite under (H3). From the Dominated Convergence Theorem we get (iii). □

Proposition 7.A.2. *Assume (H1) and (H3) to be true for μ, $\nu \in \mathcal{P}(\mathbb{R}^d)$. Let $\boldsymbol{\pi}^T$ be the unique minimiser in KSP with marginals μ, ν. Suppose we are given a sequence $\left(q_n^T\right)_{n\in\mathbb{N}} \subset \mathcal{P}(\mathbb{R}^{4d})$ such that such that*

(i) $q_n^T \rightharpoonup \boldsymbol{\pi}^T$,

(ii) $\mathscr{H}\left(q_n^T|\mathrm{R}_{0,T}\right) \to \mathscr{H}\left(\boldsymbol{\pi}^T|\mathrm{R}_{0,T}\right)$.

Moreover for each $n \in \mathbb{N}$ consider the marginals q_n^T, that are the probability measures $\mu^n := (\mathrm{proj}_{x_1})_\# q_n^T$ and $\nu^n := (\mathrm{proj}_{x_2})_\# q_n^T$. Then, for each $n \in \mathbb{N}$, there exists a unique minimiser $\boldsymbol{\pi}_n^T \in \Pi_X(\mu^n,\nu^n)$ in KSP with marginals μ^n, ν^n. Moreover it holds

$$\boldsymbol{\pi}_n^T \rightharpoonup \boldsymbol{\pi}^T \qquad \textit{and} \qquad \mathcal{C}_T(\mu^n,\nu^n) \rightharpoonup \mathcal{C}_T(\mu,\nu)\,.$$

Proof. Firstly, (H3) and the convergence of the entropies in the assumptions imply that

$$\mathscr{H}(\mu^n|\mathfrak{m}_X)\,,\mathscr{H}(\nu^n|\mathfrak{m}_X) \le \mathscr{H}\left(q_n^T|\mathrm{R}_{0,T}\right) < C$$

for some positive constant C, uniformly in $n \in \mathbb{N}$. Hence (H3) holds also for μ^n, ν^n. This gives the existence and uniqueness of the minimiser in KSP with marginals μ^n and ν^n for each $n \in \mathbb{N}$.

Then, from (7.2.19) we deduce

$$\sup_{n\in\mathbb{N}} \mathscr{H}\left(\boldsymbol{\pi}_n^T|\mathrm{R}_{0,T}\right) = \sup_{n\in\mathbb{N}} \mathcal{C}_T(\mu^n,\nu^n) \lesssim 1 + \sup_{n\in\mathbb{N}}\left[\mathscr{H}(\mu^n|\mathfrak{m}_X) + \mathscr{H}(\nu^n|\mathfrak{m}_X)\right]$$
$$\lesssim 1 + 2C\,.$$

Since the relative entropy $\mathscr{H}(\cdot|\mathrm{R}_{0,T})$ has compact level sets [DE97, Lemma 1.4.3], there is a subsequence $\left(\boldsymbol{\pi}^T_{n_k}\right)_{k\in\mathbb{N}}$ and a probability measure $\tilde{\boldsymbol{\pi}}^T \in \mathcal{P}(\mathbb{R}^{4d})$ such that $\boldsymbol{\pi}^T_{n_k} \rightharpoonup \tilde{\boldsymbol{\pi}}^T$ weakly. Moreover, from the lower semicontinuity of $\mathscr{H}(\cdot|\mathrm{R}_{0,T})$ and the optimality of $\boldsymbol{\pi}^T_{n_k}$ we get

$$\mathscr{H}\left(\tilde{\boldsymbol{\pi}}^T|\mathrm{R}_{0,T}\right) \leq \liminf_{k\to\infty} \mathscr{H}\left(\boldsymbol{\pi}^T_{n_k}|\mathrm{R}_{0,T}\right) \leq \liminf_{k\to\infty} \mathscr{H}\left(q^T_{n_k}|\mathrm{R}_{0,T}\right) = \mathscr{H}(\boldsymbol{\pi}^T|\mathrm{R}_{0,T}). \tag{7.A.1}$$

Now, we claim that $\tilde{\boldsymbol{\pi}}^T \in \Pi_X(\mu,\nu)$. Indeed we have for any $i = 1,2$

$$\begin{aligned}(\mathrm{proj}_{x_i})_\#\tilde{\boldsymbol{\pi}}^T &= \lim_{k\to\infty}(\mathrm{proj}_{x_i})_\#\boldsymbol{\pi}^T_{n_k} = \lim_{k\to\infty}(\mathrm{proj}_{x_i})_\# q^T_{n_k} \\ &= (\mathrm{proj}_{x_i})_\#\boldsymbol{\pi}^T = \begin{cases}\mu & i=1\\ \nu & i=2,\end{cases}\end{aligned}$$

where the second equality holds because $\boldsymbol{\pi}^T_{n_k}$ and $q^T_{n_k}$ share the same marginals, while the third follows from our hypotheses. Therefore, from the bound (7.A.1) and the optimality of $\boldsymbol{\pi}^T$ as unique minimiser in $\Pi_X(\mu,\nu)$ for KSP, it follows $\tilde{\boldsymbol{\pi}}^T = \boldsymbol{\pi}^T$.

Hence, as $k \to \infty$, it holds $\boldsymbol{\pi}^T_{n_k} \rightharpoonup \boldsymbol{\pi}^T$ and

$$\exists \lim_{k\to\infty} \mathcal{C}_T\left(\mu^{n_k},\nu^{n_k}\right) = \lim_{k\to\infty}\mathscr{H}\left(\boldsymbol{\pi}^T_{n_k}|\mathrm{R}_{0,T}\right) = \mathscr{H}(\boldsymbol{\pi}^T|\mathrm{R}_{0,T}) = \mathcal{C}_T(\mu,\nu)\,.$$

Since in both the limits above the limit objects do not depend on the subsequence and since the weak convergence is metrizable, we get the desired thesis. □

Corollary 7.A.3. *Under the same setting of the previous proposition, if* $\mathbf{P}^{n,T} \in \mathcal{P}(\Omega)$ *denotes the minimiser in* (7.0.1) *for the marginals* μ^n, ν^n, *and if* $\mu^{n,T}_t := (X_t, V_t)_\#\mathbf{P}^{n,T}$, *then for each* $t \in [0, T]$

$$\mathbf{P}^{n,T} \rightharpoonup \mathbf{P}^T \qquad \textit{and} \qquad \mu^{n,T}_t \rightharpoonup \mu^T_t\,.$$

Proof. From the relation between (7.0.1) and KSP, for any $\phi \in C_b(\Omega)$ we have

$$\begin{aligned}\int_\Omega \phi\,\mathrm{d}\mathbf{P}^{n,T} = \int_{\mathbb{R}^{4d}}\left(\int_\Omega \phi\,\mathrm{d}\mathrm{R}^{x,v,y,w}\right)\mathrm{d}\boldsymbol{\pi}^T_n &\to \int_{\mathbb{R}^{4d}}\left(\int_\Omega \phi\,\mathrm{d}\mathrm{R}^{x,v,y,w}\right)\mathrm{d}\boldsymbol{\pi}^T \\ &= \int_\Omega \phi\,\mathrm{d}\boldsymbol{\pi}^T\,,\end{aligned}$$

where $\mathrm{R}^{x,v,y,w}$ denotes the bridge of the reference measure. Let us just justify the middle step. Since ϕ is bounded, so does $\int_\Omega \phi\,\mathrm{d}\mathrm{R}^{x,v,y,w}$. Moreover since the bridge $\mathrm{R}^{x,v,y,w}$ is weakly continuous with respect to its extremes [CB11, Corollary 1], from the continuity of ϕ, it follows the continuity of the function $(x,v,y,w) \mapsto \int_\Omega \phi\,\mathrm{d}\mathrm{R}^{x,v,y,w}$. Hence the above function is bounded and continuous on $\mathbb{R}^{4d}$ and from the weak convergence $\boldsymbol{\pi}^T_n \rightharpoonup \boldsymbol{\pi}^T$ it follows $\mathbf{P}^{n,T} \rightharpoonup \mathbf{P}^T$. The other limit follows by taking the time marginals of $\mathbf{P}^{n,T}$. □

7.A.2 Approximating the marginals

In this section we are going to perform the approximating arguments directly on the fixed marginals. This will not lead to the convergence of the respective kinetic entropic costs, nevertheless it will be useful in proving the bounds where the previous approximating argument fails. The idea is similar to the one performed previously: consider an increasing sequence of compact sets $(K_n)_{n\in\mathbb{N}}$ in $\mathbb{R}^d$ whose union gives the whole space and for any $\mathfrak{q} \in \mathcal{P}(\mathbb{R}^d)$, satisfying (H3) and $n \in \mathbb{N}$ large enough so that $\mathfrak{q}(K_n) > 0$, define the probability measure $\mathfrak{q}_n^M$ as the measure whose $\mathfrak{m}_X$-density is given by

$$\frac{\mathrm{d}\mathfrak{q}_n^M}{\mathrm{d}\mathfrak{m}_X} := \left(\frac{\mathrm{d}\mathfrak{q}}{\mathrm{d}\mathfrak{m}_X} \wedge n\right) \frac{\mathbf{1}_{K_n}}{C_n^{\mathfrak{q}}}\,,$$

where $C_n^{\mathfrak{q}} := \int_{K_n} (\frac{\mathrm{d}\mathfrak{q}}{\mathrm{d}\mathfrak{m}_X} \wedge n)\, \mathrm{d}\mathfrak{m}_X \geq 0$ is the normalising constant. Note that monotone convergence yields $C_n^{\mathfrak{q}} \uparrow 1$. Then it follows that $\mathfrak{q}_n^M$ satisfies (H4), $\mathfrak{q}_n^M \rightharpoonup \mathfrak{q}$ and by mimicking the argument performed in the Lemma 7.A.1 it follows that

$$\mathscr{H}(\mathfrak{q}_n^M|\mathfrak{m}_X) \overset{n\to\infty}{\longrightarrow} \mathscr{H}(\mathfrak{q}|\mathfrak{m}_X)\,. \tag{7.A.2}$$

Lemma 7.A.4. *Fix $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ satisfying (H3). Then, up to restricting ourselves to a subsequence, it holds*

$$\mathcal{C}_T(\mu,\nu) \leq \liminf_{n\to\infty} \mathcal{C}_T(\mu_n^M, \nu_n^M)\,.$$

Proof. Let $\boldsymbol{\pi}_{M,n}^T$ denotes the optimiser for $\mathcal{C}_T(\mu_n^M, \nu_n^M)$. Then we have

$$\begin{aligned}\mathscr{H}(\boldsymbol{\pi}_{M,n}^T|\mathrm{R}_{0,T}) &= \mathcal{C}_T(\mu_n^M,\nu_n^M)\\ &\overset{(7.2.19)}{\lesssim} 1 + \mathscr{H}(\mu_n^M|\mathfrak{m}_X) + \mathscr{H}(\nu_n^M|\mathfrak{m}_X) \overset{n\to\infty}{\longrightarrow} 1 + \mathscr{H}(\mu|\mathfrak{m}_X) + \mathscr{H}(\nu|\mathfrak{m}_X)\,,\end{aligned}$$

which is finite because of (H3). Since $\mathscr{H}(\cdot|\mathrm{R}_{0,T})$ has compact level set, we know that there exists $\boldsymbol{\pi}^\star \in \mathcal{P}(\mathbb{R}^{4d})$ such that $\boldsymbol{\pi}_{M,n}^T \rightharpoonup \boldsymbol{\pi}^\star$, up to considering a subsequence. We claim that $\boldsymbol{\pi}^\star \in \Pi_X(\mu,\nu)$. Indeed we have $(X_0)_\# \boldsymbol{\pi}_{M,n}^T \rightharpoonup (X_0)_\# \boldsymbol{\pi}^\star$ but $(X_0)_\# \boldsymbol{\pi}_{M,n}^T = \mu_n^M \rightharpoonup \mu$ and hence $(X_0)_\# \boldsymbol{\pi}^\star = \mu$. Similarly it holds $(X_T)_\# \boldsymbol{\pi}^\star = \nu$. Therefore we have $\mathcal{C}_T(\mu,\nu) \leq \mathscr{H}(\boldsymbol{\pi}^\star|\mathrm{R}_{0,T})$ and from the lower semicontinuity of $\mathscr{H}(\cdot|\mathrm{R}_{0,T})$ we deduce our thesis. □

7.B Proof of the contraction condition

Proof of (7.2.11) *with $\kappa > 0$ under (H1) and (H2).* Notice that proving (7.2.11) is equivalent to finding a positive definite symmetric matrix $Q \in \mathcal{M}_{2d,2d}^{sym>0}(\mathbb{R})$ and a positive scalar $\kappa > 0$ such that

$$\xi \cdot (J_b(z)Q)\xi \leq -\kappa\, \xi \cdot Q\xi = -\kappa\, |\xi|_Q^2 \qquad \forall \xi \in \mathbb{R}^{2d},\, \forall z \in \mathbb{R}^{2d}\,,$$

and then setting $M = Q^{-1}$. Therefore let us consider the symmetric positive definite matrix

$$Q = \begin{pmatrix} a\mathrm{Id} & -b\mathrm{Id} \\ -b\mathrm{Id} & c\mathrm{Id} \end{pmatrix} \qquad \text{with} \qquad \begin{cases} a, c > 0\,, \\ ac > b^2\,. \end{cases}$$

with $a, b, c \in \mathbb{R}$ to be determined later in such a way such for all $\bar{x},\ \bar{v} \in \mathbb{R}^d \in \mathbb{R}^d$ the matrix $J_b(\bar{x}, \bar{v})\, Q$ is negative definite uniformly in $\bar{x},\ \bar{v}$. This will allow us to conclude that $J_b(\bar{x}, \bar{v})\, Q \le -\kappa\, Q$ for some $\kappa > 0$. For sake of notation we will omit the Jacobian's argument $(\bar{x},\ \bar{v})$ when it is clear from the context. Since

$$J_b\, Q = \begin{pmatrix} -b\mathrm{Id} & c\mathrm{Id} \\ -a\, \nabla^2 U + \gamma b\mathrm{Id} & b\, \nabla^2 U - \gamma c\mathrm{Id} \end{pmatrix},$$

Then for all $x,\ v \in \mathbb{R}^d$ it holds

$$\begin{aligned} \begin{pmatrix} x \\ v \end{pmatrix} \cdot J_b\, Q \begin{pmatrix} x \\ v \end{pmatrix} &= -b\, |x|^2 + c\, x \cdot v - a\, v \cdot \nabla^2 U\, x + \gamma b\, v \cdot x \\ &\qquad + b\, v \cdot \nabla^2 U\, v - \gamma c\, |v|^2\,. \end{aligned}$$

By choosing $a = 1$, and under a sign flip in the *velocity term* $v \mapsto -v$, it is enough to find $b,\ c$ and $\varepsilon > 0$ such that $c > b^2$ and such that for all x, v, uniformly in $\bar{x}, \bar{v}$, it holds

$$b\, |x|^2 + c\, x \cdot v - a\, v \cdot \nabla^2 U\, x + \gamma b\, v \cdot x - b\, v \cdot \nabla^2 U\, v + \gamma c\, |v|^2 \ge \varepsilon\big(|x|^2 + |v|^2\big).$$

A sufficient condition such that the previous inequality holds for some $\varepsilon > 0$ is that for all eigenvalues ℓ of $\nabla^2 U$ (notice $\ell \in [\alpha, \beta]$ because of (H2)) and all $(x, v) \neq (0, 0)$ it holds

$$\begin{aligned} & b\, |x|^2 + (c + \gamma b)\, x \cdot v - \ell\, v \cdot x - b\, \ell\, v \cdot v + \gamma c\, |v|^2 \\ &= b\, |x|^2 + (c + \gamma b - \ell)\, x \cdot v + (\gamma c - b\, \ell)\, |v|^2 > 0. \end{aligned}$$

This last one is satisfied for any non-zero $(x, v) \in \mathbb{R}^{2d}$ as soon as $b > 0$ and

$$4b(\gamma c - b\, \ell) > (c + \gamma b - \ell)^2,$$

i.e., $\ell^2 - 2\lambda\, \ell + \phi^2 < 0$, where $\lambda := c + \gamma b - 2b^2$ and $\phi := c - \gamma b$. Therefore for any eigenvalue $\ell \in [\alpha,\ \beta]$ we have the condition

$$\lambda - \sqrt{\lambda^2 - \phi^2} < \ell < \lambda + \sqrt{\lambda^2 - \phi^2},$$

which is satisfied if

$$\begin{cases} \alpha = \lambda - \sqrt{\lambda^2 - \phi^2} \\ \beta = \lambda + \sqrt{\lambda^2 - \phi^2}\,. \end{cases}$$

This yields to

$$\begin{cases} \lambda = \frac{\beta+\alpha}{2} \\ \phi^2 = \beta\,\alpha \end{cases} \qquad \textit{i.e.,} \qquad \begin{cases} c + \gamma b - 2b^2 = \frac{\beta+\alpha}{2} \\ c - \gamma b = \sqrt{\beta\,\alpha} \end{cases}$$

whose solutions b, c are given by the solutions of

$$\begin{cases} c = b^2 + \frac{1}{4}\left(\sqrt{\beta} + \sqrt{\alpha}\right)^2 \\ b^2 - \gamma b + \frac{1}{4}\left(\sqrt{\beta} - \sqrt{\alpha}\right)^2 = 0 \end{cases}$$

The first equation fixes c and in particular tells us that $c > b^2$ (as requested in order to have $Q \in \mathcal{M}^{sym>0}_{2d,2d}(\mathbb{R})$), while the second equation has a positive solution b if ad only if

$$\gamma^2 - \left(\sqrt{\beta} - \sqrt{\alpha}\right)^2 \geq 0,$$

which we have assumed to be true in (H2). □

Remark 7.B.1. *Let us point out that Equation* (7.0.2) *is just a particular instance of the general class dealt by Monmarché in [Mon23]. Nevertheless the previous result is more sharp than the one presented there in Proposition 5. The idea for the previous proof comes from [Bau17, Theorem 2.12], where the author proves a* Γ-calculus contraction condition*, which turns out to be equivalent to* (7.2.11) *for* $\gamma = 1$.

Looking forward

In this last section we will collect a few possible lines of research that can follow from the results presented in this manuscript.

Convergence of gradients of EOT potentials. In Chapter 3 we have shown the convergence of the gradients of Schrödinger potentials towards the gradients of Kantorovich potentials. The same question can be asked in the more abstract setting of entropic optimal transport for a general cost function $c(\cdot,\cdot)$ (cf. (2.A.1)). More precisely, if φ_ε, ψ_ε denote the entropic potentials as defined at (2.A.2), then it is natural asking whether the result presented in Theorem 3.2.3 can be extended to this different setting. Convergence of entropic potentials to the Kantorovich ones has already been established in [NW22], however the validity of the same result for the gradients is still an open question.

One possible way to tackle this problem would be establishing uniform in time bounds for the gradients' L^p-norms as we did in (3.2.8). This might be accomplished by proving corrector estimates as in Proposition 3.1.2, however mimicking the proof given there seems unfeasible since it relies on a dynamic representation of $\mathcal{C}_T(\mu,\nu)$ as the one given in Lemma 2.3.1. The latter formulation is still missing for general EOT problems. A different approach in order to establish the corrector estimates (and consequently uniform norm bounds) would be reasoning as we did at the end of Section 3.1, where under the condition $\mathsf{CD}(\kappa,\infty)$ we have established the same estimates via the reverse log-Sobolev inequality (3.1.13). This approach seems more feasible even though it requires establishing the validity of the reverse log-Sobolev inequality for a general Markov operator (induced by the convolution with $\exp(-c(x,\cdot)/T)$). The ideas described above are part of a (preliminary) ongoing work with Katharina Eichinger and Luca Tamanini.

Exponential convergence of Sinkhorn's algorithm for EOT. In Chapters 5 and 6 we have established the exponential convergence of Sinkhorn's algorithm for general SP problems with Langevin reference dynamics and log-Lipschitz marginals and respectively for the landmark SP with Brownian motions and marginals that are weakly log-concave. As we have already mentioned in the Bibliographical Remarks to Chapter 6, an iterative fitting algorithm can be in general considered for any EOT problem. Apart from the very recent contribution [Eck23], exponential convergence rates for general EOT problems are still unknown. A first approach in order to prove the convergence of Sinkhorn's algorithm for general EOT problems would be looking for functional inequalities for general Markov operators in order to prove convexity propagation, as we did in Section 6.1. As a corollary, this would eventually lead also to novel functional inequalities for general EOT problems. Some potential references for functional inequalities for general Markov operators would be [Wan13, BG10, BR08, BE21]. Alternatively, a different approach would be trying to directly mimic the backward-in-time convexity propagation approach along Hamilton-Jacobi-Bellman equations (cf. Theorem 6.1.4), this time by considering a stochastic control problem for discrete-time Markov chains and its corresponding dynamic programming principle.

Once the convexity propagation is established, unfortunately, our approach can not be carried over straightforwardly as done in Chapter 6, since there we heavily rely on the link between gradients and conditional probability measures portrayed in (6.0.3), which is peculiar to the quadratic cost case $c(x,y) = |x-y|^2/2$ (that corresponds to the classic Brownian motion SP). Nevertheless, for some specific cases (e.g., $c(x,y) = |x-y|^p$)) some link between gradients and (distorted) Wasserstein distances of conditional measures might still be established.

Exponential convergence of Sinkhorn's algorithm for the Mean-Field SP. In [BCGL20] the authors proposed a SP where instead of considering a cloud of independent Brownian motions, they considered a mean-field interacting system. The entropic minimisation problem then gets harder since both terms in the entropy functional would depend on the optimiser and there is no (algebraically) straightforward connection with EOT. This particularly implies that in order to solve the Mean-Field SP (hereafter MFSP) one cannot simply consider pairs of potentials, but instead corrector processes and forward/backward stochastic differential equations. Nevertheless, Sinkhorn's algorithm for classical SP (more precisely, its primal formulation (2.2.20)) might be emulated also for MFSP, this time directly dealing with diffusion processes and measures on path space instead of densities and potentials. As noticed in [LCST22], MFSP may be considered as an opinion dynamic Mean-Field Game (MFG), where every single individual interacts with the population, which is urged to converge exactly to a desired prescribed opinion (distribution) in finite time. Therefore the study of Sinkhorn's algorithm for MFSP would have a natural application in modelling and solving the collective behaviour of opinion/population dynamics. In [LCST22] the authors suggest an algorithm (referred to as *Deep Generalized Schrödinger Bridge (DGSB)*) which tries to solve MFGs via minimising a total loss function which corresponds to the sum of a loss fitting the correct marginals and a loss fitting the mean-field interaction and which is related to the *Temporal Difference Learning*. Sinkhorn's algorithm should indeed fit the mean-field interaction nature of the considered path measures when minimising the relative entropy with respect to the McKean-Vlasov version of the previous iterate. From a computational point of view it would then be interesting to understand whether the loss associated to this Mean-Field Sinkhorn's algorithm coincides with the one considered in [LCST22] and with the temporal difference loss.

Overdamped limit of the kinetic SP. The study of the overdamped limit for the (underdamped) Langevin dynamics (7.0.2) is a very old research topic, often referred to as the *Smoluchowski–Kramers* limit, initiated with the seminal work of Kramers [Kra40]. Since then many contributions appeared in the literature and let us just mention here a partial (and far from being exhaustive) list of contributions [Nel67, HVW12, GN20, DLP$^+$18]. Particularly, in [DLP$^+$18] the authors provide explicit quantitative convergence rates of the underdamped Langevin dynamics (7.0.2) towards the overdamped Langevin dynamics (2.1.1), in terms of relative entropy and Wasserstein $\mathbf{W}_2$-distance. This means that the reference measure considered in the kinetic SP converges in the overdamped limit to the reference of the classic (Langevin) SP considered in Chapter 2. Therefore, if some sort of stability results hold for SP when varying the underlying reference process (such as [EN22b]), then one might expect the convergence of KSP to SP in the overdamped limit with convergence rates inherited from the results obtained in [DLP$^+$18].

This page was intentionally left blank.

Notation

The set $\mathbb{R}^d$ is endowed with the standard Euclidean metric and we denote by $|\cdot|$ and $\langle \cdot, \cdot \rangle$ the corresponding norm and scalar product. When its clear from the context the scalar product between the vectors v, w will simply be denoted as $v \cdot w$. We denote by $\mathcal{B}(\mathbb{R}^d)$ the Borel σ-field of $\mathbb{R}^d$, by $\mathcal{P}(\mathbb{R}^d)$ the space of probability measures defined on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ and by $\mathcal{P}_2(\mathbb{R}^d)$ the subset of probability measures with finite second moment. Ω denotes the space of continuous trajectories in the time interval $[0, T]$, that is the space $\mathcal{C}([0, T], M)$. In Chapter 7 this space will be denoted with Ω_{2d} since the trajectories will take values in the position-velocity product space $\mathbb{R}^{2d}$. We adopt for L^p-spaces the standard notation and denote L^p-norms as $\|\cdot\|_{\mathrm{L}^p}$. Similarly, we adopt the standard notation for p-Wasserstein distances $\mathbf{W}_p$ for any $p \in [1, +\infty]$. We denote with $W^{1,2}_{loc}$ local-Sobolev spaces.

$a_+ = 0 \vee a$	
$a_- = 0 \vee (-a)$	
$\mathrm{proj}_x(a, b) = a$	projection on first component
$\mathrm{proj}_y(a, b) = b$	projection on second component
$f \oplus g(x, y) := f(x) + g(y)$	
$\mathcal{C}_b(\mathcal{X})$	bounded continuous functions on $\mathcal{X}$
$\mathcal{M}_b(\mathcal{X})$	bounded measurable functions on $\mathcal{X}$
$\|f\|_\infty := \sup_{x \in M} |f(x)|$	supremum norm
$M_k(\mu) = \int \mathsf{d}(z_0, x)^k \mathrm{d}\mu(x)$	k^{th} moment of $\mu \in \mathcal{P}(M)$ ($z_0 = 0$ in $\mathbb{R}^d$)
$\mathrm{Cov}(\mu) := \int x x^\mathsf{T} \mathrm{d}\mu - (\int x \mathrm{d}\mu)(\int x \mathrm{d}\mu)^\mathsf{T}$	covariance matrix of $\mu \in \mathcal{P}_2(\mathbb{R}^d)$
Leb	Lebesgue measure on $\mathbb{R}^d$
$f_\sharp \mu(\cdot) = \mu(f^{-1}(\cdot))$	pushforward measure of μ by f
$\Pi(\mu_1, \mu_2)$	set of couplings between μ_1, $\mu_2 \in \mathcal{P}(M)$
$\mathscr{H}(\lambda_1 \mid \lambda_2) = \int \log(\mathrm{d}\lambda_1 / \mathrm{d}\lambda_2)\, \mathrm{d}\lambda_1$	relative entropy (see Section 1.A)
$\mathrm{Ent}(\lambda_1) = \mathscr{H}(\lambda_1 \mid \mathrm{Leb})$	
$\mathcal{I}(\mu) := \int |\nabla \log(\mathrm{d}\mu / \mathrm{d}\mathfrak{m})|^2 \mathrm{d}\mu$	Fisher information w.r.t. $\mathfrak{m}$ (see (3.2.7))
$\mathbf{P}^T$	Schrödinger bridge (optimiser in (2.1.2))
$\boldsymbol{\pi}^T$	Schrödinger plan (optimiser in (2.1.3))
$\mathcal{C}_T(\mu, \nu) = \mathscr{H}(\mathbf{P}^T \mid \mathrm{R}) = \mathscr{H}(\boldsymbol{\pi}^T \mid \mathrm{R}_{0,T})$	Schrödinger cost

This page was intentionally left blank.

Bibliography

[ADPZ11] Stefan Adams, Nicolas Dirr, Mark A. Peletier, and Johannes Zimmer. From a Large-Deviations Principle to the Wasserstein Gradient Flow: A New Micro-Macro Passage. *Communications in Mathematical Physics*, 307(3):791–815, 2011.

[AGS14a] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. Calculus and heat flow in metric measure spaces and applications to spaces with Ricci bounds from below. *Inventiones mathematicae*, 195(2):289–391, 2014.

[AGS14b] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. Metric measure spaces with Riemannian Ricci curvature bounded from below. *Duke Mathematical Journal*, 163(7):1405–1490, 2014.

[Bau17] Fabrice Baudoin. Bakry-Émery meet Villani. *Journal of Functional Analysis*, 273(7):2275–2291, 2017.

[BCC$^+$15] Jean-David Benamou, Guillaume Carlier, Marco Cuturi, Luca Nenna, and Gabriel Peyré. Iterative Bregman projections for regularized transportation problems. *SIAM Journal on Scientific Computing*, 37(2):A1111–A1138, 2015.

[BCGL20] Julio Backhoff, Giovanni Conforti, Ivan Gentil, and Christian Léonard. The mean field Schrödinger problem: ergodic behavior, entropy estimates and functional inequalities. *Probability Theory and Related Fields*, 178(1):475–530, 2020.

[BÉ85] Dominique Bakry and Michel Émery. Diffusions hypercontractives. In *Séminaire de Probabilités de Strasbourg XIX 1983/84*, pages 177–206. Springer, 1985.

[BE21] Fabrice Baudoin and Nathaniel Eldredge. Transportation inequalities for Markov kernels and their applications. *Electronic Journal of Probability*, 26(none):1 – 30, 2021.

[Ber20] Robert J Berman. The Sinkhorn algorithm, parabolic optimal transport and geometric Monge–Ampère equations. *Numerische Mathematik*, 145(4):771–836, 2020.

[BG10] François Bolley and Ivan Gentil. Phi-entropy inequalities for diffusion semigroups. *Journal de Mathématiques Pures et Appliquées*, 93(5):449–473, 2010.

[BGL13] Dominique Bakry, Ivan Gentil, and Michel Ledoux. *Analysis and geometry of Markov diffusion operators*, volume 348. Springer Science & Business Media, 2013.

[BGM10] François Bolley, Arnaud Guillin, and Florent Malrieu. Trend to equilibrium and particle approximation for a weakly selfconsistent Vlasov-Fokker-Planck equation. *ESAIM: Mathematical Modelling and Numerical Analysis*, 44(5):867–884, 2010.

[BGN22] Espen Bernton, Promit Ghosal, and Marcel Nutz. Entropic Optimal Transport: Geometry and Large Deviations. *Duke Mathematical Journal*, 171(16):3363 – 3400, 2022.

[BK08] Franck Barthe and Alexander V. Kolesnikov. Mass Transport and Variants of the Logarithmic Sobolev Inequality. *Journal of Geometric Analysis*, 18:921–979, 2008.

[BLN94] Jonathan M Borwein, Adrian Stephen Lewis, and Roger Nussbaum. Entropy minimization, DAD problems, and doubly stochastic kernels. *Journal of Functional Analysis*, 123(2):264–307, 1994.

[BP20] Tobias Breiten and Laurent Pfeiffer. On the turnpike property and the receding-horizon method for linear-quadratic optimal control problems. *SIAM Journal on Control and Optimization*, 58(2):1077–1102, 2020.

[BR08] Franck Barthe and Cyril Roberto. Modified Logarithmic Sobolev Inequalities on $\mathbb{R}$. *Potential Analysis*, 29(2):167–193, 2008.

[Bra06] Andrea Braides. A handbook on Γ-convergence. In *Handbook of Differential Equations: Stationary Partial Differential Equations*, volume 3, pages 101–213. Elsevier, 2006.

[Bre91] Yann Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Communications on pure and applied mathematics*, 44(4):375–417, 1991.

[Bre10] Haim Brezis. *Functional Analysis, Sobolev Spaces and Partial Differential Equations*. Springer New York, NY, 2010.

[BV05] François Bolley and Cédric Villani. Weighted Csiszár-Kullback-Pinsker inequalities and applications to transportation inequalities. *Annales de la Faculté des sciences de Toulouse : Mathématiques*, Ser. 6, 14(3):331–352, 2005.

[Car22] Guillaume Carlier. On the Linear Convergence of the Multi-marginal Sinkhorn Algorithm. *SIAM Journal on Optimization*, 32(2):786–794, 2022.

[CB11] Loïc Chaumont and Gerónimo Uribe Bravo. Markovian bridges: Weak continuity and pathwise constructions. *The Annals of Probability*, 39(2):609–647, 2011.

[CCG22] Gauthier Clerc, Giovanni Conforti, and Ivan Gentil. Long-time behaviour of entropic interpolations. *Potential Analysis*, pages 1–31, 2022.

[CCGR22] Alberto Chiarini, Giovanni Conforti, Giacomo Greco, and Zhenjie Ren. Entropic turnpike estimates for the kinetic Schrödinger problem. *Electronic Journal of Probability*, 27:1 – 32, 2022.

[CCGT23] Alberto Chiarini, Giovanni Conforti, Giacomo Greco, and Luca Tamanini. Gradient estimates for the Schrödinger potentials: convergence to the Brenier map and quantitative stability. *Communications in Partial Differential Equations*, 48(6):895–943, 2023.

[CDG23] Giovanni Conforti, Alain Oliviero Durmus, and Giacomo Greco. Quantitative contraction rates for Sinkhorn algorithm: beyond bounded costs and compact marginals. *arXiv preprint arXiv:2304.04451*, 2023.

[CGMZ19] Patrick Cattiaux, Arnaud Guillin, Pierre Monmarché, and Chaoen Zhang. Entropic multipliers method for Langevin diffusion and weighted log Sobolev inequalities. *Journal of Functional Analysis*, 277(11):108288, 2019.

[CGP15] Yongxin Chen, Tryphon Georgiou, and Michele Pavon. Fast cooling for a system of stochastic oscillators. *Journal of Mathematical Physics*, 56(11):113302, 2015.

[CGP16a] Yongxin Chen, Tryphon Georgiou, and Michele Pavon. Entropic and Displacement Interpolation: A Computational Approach Using the Hilbert Metric. *SIAM Journal on Applied Mathematics*, 76(6):2375–2396, 2016.

[CGP16b] Yongxin Chen, Tryphon Georgiou, and Michele Pavon. On the relation between optimal transport and Schrödinger bridges: A stochastic control viewpoint. *Journal of Optimization Theory and Applications*, 169(2):671–691, 2016.

[CGP16c] Yongxin Chen, Tryphon Georgiou, and Michele Pavon. Optimal steering of a linear stochastic system to a final probability distribution, part i. *IEEE Transactions on Automatic Control*, 61(5):1158–1169, 2016.

[CGP16d] Yongxin Chen, Tryphon Georgiou, and Michele Pavon. Optimal steering of a linear stochastic system to a final probability distribution, part ii. *IEEE Transactions on Automatic Control*, 61(5):1170–1180, 2016.

[CHSG22] Evan Camrud, David P. Herzog, Gabriel Stoltz, and Maria Gordina. Weighted L2-contractivity of Langevin dynamics with singular potentials. *Nonlinearity*, 35(2):998–1035, 2022.

[CL20] Guillaume Carlier and Maxime Laborde. A Differential Approach to the Multi-Marginal Schrödinger System. *SIAM Journal on Mathematical Analysis*, 52(1):709–717, 2020.

[CLLP12] Pierre Cardaliaguet, Jean-Michel Lasry, Pierre-Louis Lions, and Alessio Porretta. Long time average of mean field games. *Networks and Heterogeneous Media*, 7(2):279–301, 2012.

[CLLP13] Pierre Cardaliaguet, Jean-Michel Lasry, Pierre-Louis Lions, and Alessio Porretta. Long time average of mean field games with a nonlocal coupling. *SIAM Journal on Control and Optimization*, 51(5):3558–3591, 2013.

[Con19] Giovanni Conforti. A second order equation for Schrödinger bridges with applications to the hot gas experiment and entropic transportation cost. *Probability Theory and Related Fields*, 174(1):1–47, 2019.

[Con23] Giovanni Conforti. Coupling by reflection for controlled diffusion processes: Turnpike property and large time behavior of Hamilton Jacobi Bellman equations. *The Annals of Applied Probability*, 33(6A):4608–4644, 2023.

[Con24] Giovanni Conforti. Weak semiconvexity estimates for Schrödinger potentials and logarithmic Sobolev inequality for Schrödinger bridges. *accepted in Probability Theory and Related Fields*, 2024+.

[CP19] Pierre Cardaliaguet and Alessio Porretta. Long time behavior of the master equation in mean field game theory. *Analysis & PDE*, 12(6):1397–1453, 2019.

[CP23] Sinho Chewi and Aram-Alexandre Pooladian. An entropic generalization of Caffarelli's contraction theorem via covariance inequalities. *Comptes Rendus. Mathématique*, 361:1471–1482, 2023.

[CPT23] Guillaume Carlier, Paul Pegon, and Luca Tamanini. Convergence rate of general entropic optimal transport costs. *Calculus of Variations and Partial Differential Equations*, 62(4):116, 2023.

[CRL+20] Lenaic Chizat, Pierre Roussillon, Flavien Léger, François-Xavier Vialard, and Gabriel Peyré. Faster Wasserstein distance estimation with the Sinkhorn divergence. *Advances in Neural Information Processing Systems*, 33:2257–2269, 2020.

[CT21] Giovanni Conforti and Luca Tamanini. A formula for the time derivative of the entropic cost and applications. *Journal of Functional Analysis*, 280(11):964–1008, 2021.

[DBTHD21] Valentin De Bortoli, James Thornton, Jeremy Heng, and Arnaud Doucet. Diffusion Schrödinger bridge with applications to score-based generative modeling. *Advances in Neural Information Processing Systems*, 34, 2021.

[DdBD24] George Deligiannidis, Valentin de Bortoli, and Arnaud Doucet. Quantitative uniform stability of the iterative proportional fitting procedure. *The Annals of Applied Probability*, 34(1A):501 – 516, 2024.

[DE97] Paul Dupuis and Richard S. Ellis. *A Weak Convergence Approach to the Theory of Large Deviations*. Wiley series in Probability and Statistics. John Wiley & sons, 1997.

[DLP+18] Manh Hong Duong, Agnes Lamacz, Mark A. Peletier, André Schlichting, and Upanshu Sharma. Quantification of coarse-graining error in Langevin and overdamped Langevin dynamics. *Nonlinearity*, 31(10):4517–4566, 2018.

[DM78] Claude Dellacherie and Paul-André Meyer. *Probabilities and Potential*, volume 29 of *North-Holland Mathematics Studies*. North-Holland Publishing Company, 1978.

[DM10] François Delarue and Stéphane Menozzi. Density estimates for a random noise propagating through a chain of differential equations. *Journal of Functional Analysis*, 259(6):1577–1630, 2010.

[DMG20] Simone Di Marino and Augusto Gerolin. An Optimal Transport Approach for the Schrödinger Bridge Problem and Convergence of Sinkhorn Algorithm. *Journal of Scientific Computing*, 85(2):27, 2020.

[DMS15] Jean Dolbeault, Clément Mouhot, and Christian Schmeiser. Hypocoercivity for linear kinetic equations conserving mass. *Transactions of the American Mathematical Society*, 367(6):3807–3828, 2015.

[DPZ14] Giuseppe Da Prato and Jerzy Zabczyk. *Stochastic Equations in Infinite Dimensions*, volume 152 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, 2014.

[DZ10] Amir Dembo and Ofer Zeitouni. *Large Deviations Techniques and Applications*, volume 38 of *Stochastic Modelling and Applied Probability*. Springer Berlin, Heidelberg, 2010.

[Ebe16] Andreas Eberle. Reflection couplings and contraction rates for diffusions. *Probability Theory and Related Fields*, 166(3-4):851–886, 2016.

[Eck23] Stephan Eckstein. Hilbert's projective metric for functions of bounded growth and exponential convergence of Sinkhorn's algorithm. *arXiv preprint arXiv:2311.04041*, 2023.

[EGZ19a] Andreas Eberle, Arnaud Guillin, and Raphael Zimmer. Couplings and quantitative contraction rates for Langevin dynamics. *The Annals of Probability*, 47(4):1982–2010, 2019.

[EGZ19b] Andreas Eberle, Arnaud Guillin, and Raphael Zimmer. Quantitative Harris type theorems for diffusions and McKean-Vlasov processes. *Transactions of the American Mathematical Society*, 37(10):7135–7173, 2019.

[EMR15] Matthias Erbar, Jan Maas, and Michiel Renger. From large deviations to Wasserstein gradient flows in multiple dimensions. *Electronic Communications in Probability*, 20:1–12, 2015.

[EN22a] Stephan Eckstein and Marcel Nutz. Convergence Rates for Regularized Optimal Transport via Quantization. *Mathematics of Operations Research*, forthcoming, 2022.

[EN22b] Stephan Eckstein and Marcel Nutz. Quantitative Stability of Regularized Optimal Transport and Convergence of Sinkhorn's Algorithm. *SIAM Journal on Mathematical Analysis*, 54(6):5922–5948, 2022.

[FG11] Alessio Figalli and Nicola Gigli. Local semiconvexity of Kantorovich potentials on non-compact manifolds. *ESAIM: Control, Optimisation and Calculus of Variations*, 17(3):648–653, 2011.

[Fig07] Alessio Figalli. Existence, uniqueness, and regularity of optimal transport maps. *SIAM Journal on Mathematical Analysis*, 39(1):126–137, 2007.

[FL89] Joel Franklin and Jens Lorenz. On the scaling of multidimensional matrices. *Linear Algebra and its Applications*, 114-115:717–735, 1989.

[Föl88] Hans Föllmer. Random fields and diffusion processes. In *École d'Été de Probabilités de Saint-Flour XV–XVII, 1985–87*, pages 101–203. Springer, 1988.

[FS11] Hans Föllmer and Alexander Schied. *Stochastic Finance: An Introduction in Discrete Time*. W. de Gruyter, Berlin, 3rd edition, 2011.

[GHL04] Sylvestre Gallot, Dominique Hulin, and Jacques Lafontaine. *Riemannian Geometry*. Universitext. Springer Berlin, Heidelberg, 3 edition, 2004.

[GL00] Siegfried Graf and Harald Luschgy. *Foundations of Quantization for Probability Distributions*, volume 1730 of *Lecture Notes in Mathematics*. Springer Berlin, Heidelberg, 2000.

[GLRT20] Ivan Gentil, Christian Léonard, Luigia Ripani, and Luca Tamanini. An entropic interpolation proof of the HWI inequality. *Stochastic Processes and their Applications*, 130(2):907–923, 2020.

[GLWZ21] Arnaud Guillin, Wei Liu, Liming Wu, and Chaoen Zhang. The kinetic Fokker-Planck equation with mean field interaction. *Journal de Mathématiques Pures et Appliquées*, 150:1–23, 2021.

[GN20] Martin Grothaus and Andreas Nonnenmacher. Overdamped limit of generalized stochastic Hamiltonian systems for singular interaction potentials. *Journal of Evolution Equations*, 20(2):577–605, 2020.

[GN22] Promit Ghosal and Marcel Nutz. On the Convergence Rate of Sinkhorn's Algorithm. *arXiv preprint arXiv:2212.06000*, 2022.

[GNB22] Promit Ghosal, Marcel Nutz, and Espen Bernton. Stability of entropic optimal transport and Schrödinger bridges. *Journal of Functional Analysis*, 283(9):109622, 2022.

[GNCD23] Giacomo Greco, Maxence Noble, Giovanni Conforti, and Alain Durmus. Non-asymptotic convergence bounds for Sinkhorn iterates and their gradients: a coupling approach. In Gergely Neu and Lorenzo Rosasco, editors, *Proceedings of Thirty Sixth Conference on Learning Theory*, volume 195 of *Proceedings of Machine Learning Research*, pages 716–746. PMLR, 12–15 Jul 2023.

[Gri09] Alexander Grigor'yan. *Heat kernel and analysis on manifolds*, volume 47. American Mathematical Soc., 2009.

[GT19] Nicola Gigli and Luca Tamanini. Benamou-Brenier and duality formulas for the entropic cost on $RCD^*(K,N)$ spaces. *Probability Theory and Related Fields*, pages 1–34, 2019.

[GT21] Nicola Gigli and Luca Tamanini. Second order differentiation formula on $RCD^*(K, N)$ spaces. *Journal of the European Mathematical Society*, 23(5):1727–1795, 2021.

[GW12] Arnaud Guillin and Feng-Yu Wang. Degenerate Fokker-Planck equations: Bismut formula, gradient estimate and Harnack inequality. *Journal of Differential Equations*, (253):20–40, 2012.

[Ham93] Richard S. Hamilton. A matrix Harnack estimate for the heat equation. *Comm. Anal. Geom.*, 1(1):113–126, 1993.

[HM19] David P. Herzog and Jonathan C. Mattingly. Ergodicity and Lyapunov Functions for Langevin Dynamics with Singular Potentials. *Communications on Pure and Applied Mathematics*, 72:2231–2255, 2019.

[HMS11] Martin Hairer, Jonathan C. Mattingly, and Michael Scheutzow. Asymptotic coupling and a general form of Harris'theorem with applications to stochastic delay equations. *Probability Theory and Related Fields*, 149(1):223–259, 2011.

[HN04] Frédéric Hérau and Francis Nier. Isotropic Hypoellipticity and Trend to Equilibrium for the Fokker-Planck Equation with a High-Degree Potential. *Archive for Rational Mechanics and Analysis*, 171(2):151–218, 2004.

[Hör67] Lars Hörmander. Hypoelliptic second order differential operators. *Acta Mathematica*, 119:147–171, 1967.

[HVW12] Scott Hottovy, Giovanni Volpe, and Jan Wehr. Noise-Induced Drift in Stochastic Differential Equations with Arbitrary Friction and Diffusion in the Smoluchowski-Kramers Limit. *Journal of Statistical Physics*, 146(4):762–773, 2012.

[Hyv05] Aapo Hyvärinen. Estimation of Non-Normalized Statistical Models by Score Matching. *Journal of Machine Learning Research*, 6(24):695–709, 2005.

[JLZ16] Renjin Jiang, Huaiqian Li, and Huichun Zhang. Heat Kernel Bounds on Metric Measure Spaces and Some Applications. *Potential Analysis*, 44:601–627, 2016.

[JZ16] Renjin Jiang and Huichun Zhang. Hamilton's gradient estimates and a monotonicity formula for heat flows on metric measure spaces. *Nonlinear Analysis*, 131:32–47, 2016.

[Kal21] Olav Kallenberg. *Foundations of Modern Probability*, volume 293 of *Probability Theory and Stochastic Modelling*. Springer Cham, 3 edition, 2021.

[Kat24] Kengo Kato. Large deviations for dynamical Schrödinger problems. *arXiv preprint, arXiv:2402.05100*, 2024.

[Kot07] Brett L. Kotschwar. Hamilton's gradient estimate for the heat kernel on complete manifolds. *Proc. Amer. Math. Soc.*, 135(9):3013–3019, 2007.

[Kra40] Hendrik Anthony Kramers. Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica*, 7(4):284–304, 1940.

[LCST22] Guan-Horng Liu, Tianrong Chen, Oswin So, and Evangelos Theodorou. Deep generalized schrödinger bridge. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, editors, *Advances in Neural Information Processing Systems*, volume 35, pages 9374–9388. Curran Associates, Inc., 2022.

[Lég21] Flavien Léger. A gradient descent perspective on Sinkhorn. *Applied Mathematics & Optimization*, 84(2):1843–1855, 2021.

[Léo01] Christian Léonard. Minimization of energy functionals applied to some inverse problems. *Applied Mathematics and Optimization*, 44:273–297, 2001.

[Léo12a] Christian Léonard. From the Schrödinger problem to the Monge–Kantorovich problem. *Journal of Functional Analysis*, 262(4):1879–1920, 2012.

[Léo12b] Christian Léonard. *Girsanov Theory Under a Finite Entropy Condition*, pages 429–465. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012.

[Léo14] Christian Léonard. A survey of the Schrödinger problem and some of its connections with optimal transport. *Discrete and Continuous Dynamical Systems*, 34(4):1533–1574, 2014.

[Mik04] Toshio Mikami. Monge's problem with a quadratic cost by the zero-noise limit of h-path processes. *Probability Theory and Related Fields*, 129:245–260, 2004.

[Mon81] Gaspard Monge. Mémoire sur la théorie des déblais et des remblais. *Histoire de l'Académie Royale des Sciences*, pages 666–704, 1781.

[Mon23] Pierre Monmarché. Almost sure contraction for diffusions on $\mathbb{R}^d$. Application to generalised Langevin diffusions. *Stochastic Processes and their Applications*, 161:316–349, 2023.

[MS23] Hugo Malamut and Maxime Sylvestre. Convergence Rates of the Regularized Optimal Transport: Disentangling Suboptimality and Entropy. *arXiv preprint arXiv:2306.06940*, 2023.

[MSH02] Jonathan C. Mattingly, Andrew M. Stuart, and Desmond J. Higham. Ergodicity for SDEs and approximations: locally Lipschitz vector fields and degenerate noise. *Stochastic Processes and their Applications*, 102(2):185–232, 2002.

[MT93] Sean P. Meyn and R. L. Tweedie. Stability of Markovian Processes III: Foster-Lyapunov Criteria for Continuous-Time Processes. *Advances in Applied Probability*, 25(3):518–548, 1993.

[Nel67] Edward Nelson. *Dynamical theories of Brownian motion*, volume 2. Princeton university press Princeton, 1967.

[Nor97] James R. Norris. Heat kernel asymptotics and the distance function in Lipschitz Riemannian manifolds. *Acta Mathematica*, 179(1):79–103, 1997.

[Nut21] Marcel Nutz. Introduction to Entropic Optimal Transport. `http://www.math.columbia.edu/~mnutz/docs/EOT_lecture_notes.pdf`, 2021.

[NW22] Marcel Nutz and Johannes Wiesel. Entropic optimal transport: convergence of potentials. *Probability Theory and Related Fields*, 184(1):401–424, 2022.

[NW23] Marcel Nutz and Johannes Wiesel. Stability of Schrödinger potentials and convergence of Sinkhorn's algorithm. *The Annals of Probability*, 51(2):699 – 722, 2023.

[Pal19] Soumik Pal. On the difference between entropic cost and the optimal transport cost. *Preprint, arXiv:1905.12206*, 2019.

[PC19] Gabriel Peyré and Marco Cuturi. Computational Optimal Transport. *Foundations and Trends in Machine Learning*, 11(5-6):355–607, 2019.

[Pey18] Rémi Peyre. Comparison between $\mathcal{W}_2$ distance and $\dot{H}^{-1}$ norm, and Localisation of Wasserstein dinstance. *ESAIM: Control, Optimisation and Calculus of Variations*, 24(4):1489–1501, 2018.

[PNW21] Aram-Alexandre Pooladian and Jonathan Niles-Weed. Entropic estimation of optimal transport maps. *arXiv preprint arXiv:2109.12004*, 2021.

[RT93] Ludger Rüschendorf and Wolfgang Thomsen. Note on the Schrödinger equation and I-projections. *Statistics & Probability Letters*, 17(5):369–375, 1993.

[RT96] Gareth O. Roberts and Richard L. Tweedie. Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli*, 2(4):341–363, 1996.

[Rus95] Ludger Ruschendorf. Convergence of the iterative proportional fitting procedure. *The Annals of Statistics*, pages 1160–1174, 1995.

[RY99] Daniel Revuz and Marc Yor. *Continuous Martingales and Brownian Motion*, volume 293 of *Grundlehren der mathematischen Wissenschaften*. Springer Berlin, Heidelberg, 3 edition, 1999.

[Sch31] Erwin Schrödinger. Über die Umkehrung der Naturgesetze. *Sitzungsberichte Preuss. Akad. Wiss. Berlin. Phys. Math.*, 144:144–153, 1931.

[Sch32] Erwin Schrödinger. La théorie relativiste de l'électron et l' interprétation de la mécanique quantique. *Ann. Inst Henri Poincaré*, (2):269 – 310, 1932.

[SDBCD23] Yuyang Shi, Valentin De Bortoli, Andrew Campbell, and Arnaud Doucet. Diffusion Schrödinger bridge matching. *arXiv preprint arXiv:2303.16852*, 2023.

[Sin64] Richard Sinkhorn. A Relationship Between Arbitrary Positive Matrices and Doubly Stochastic Matrices. *The Annals of Mathematical Statistics*, 35(2):876–879, 1964.

[SK67] Richard Sinkhorn and Paul Knopp. Concerning nonnegative matrices and doubly stochastic matrices. *Pacific Journal of Mathematics*, 21(2):343–348, 1967.

[Stu06a] Karl-Theodor Sturm. On the geometry of metric measure spaces. I. *Acta Mathematica*, 196(1):65–131, 2006.

[Stu06b] Karl-Theodor Sturm. On the geometry of metric measure spaces. II. *Acta Mathematica*, 196(1):133–177, 2006.

[Tal02] Denis Talay. Stochastic Hamiltonian systems: exponential convergence to the invariant measure, and discretization by the implicit Euler scheme. *Markov Processes and Related Fields*, 8(2):163–198, 2002.

[Tam17] Luca Tamanini. *Analysis and geometry of RCD spaces via the Schrödinger problem*. PhD thesis, Université Paris Nanterre and SISSA, 2017.

[Tré23] Emmanuel Trélat. Linear turnpike theorem. *Mathematics of Control, Signals, and Systems*, 35(3):685–739, 2023.

[TZ15] Emmanuel Trélat and Enrique Zuazua. The turnpike property in finite-dimensional nonlinear optimal control. *Journal of Differential Equations*, 258(1):81–114, 2015.

[TZ18] Emmanuel Trélat and Can Zhang. Integral and measure-turnpike properties for infinite-dimensional optimal control systems. *Mathematics of Control, Signals, and Systems*, 30(1):1–34, 2018.

[TZZ18] Emmanuel Trélat, Can Zhang, and Enrique Zuazua. Steady-state and periodic exponential turnpike property for optimal control problems in hilbert spaces. *SIAM Journal on Control and Optimization*, 56(2):1222–1252, 2018.

[Vil03] Cédric Villani. *Topics in Optimal Transportation*, volume 58. American Mathematical Society, 2003.

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

[Vil09] Cédric Villani. *Hypocoercivity*, volume 202. American Mathematical Society, 2009.

[Wan94] Feng-Yu Wang. Application of coupling methods to the Neumann eigenvalue problem. *Probability Theory and Related Fields*, 98(3):299–306, 1994.

[Wan11] Feng-Yu Wang. Equivalent semigroup properties for the curvature-dimension condition. *Bulletin des sciences mathematiques*, 135(6-7):803–815, 2011.

[Wan13] Feng-Yu Wang. *Harnack Inequalities for Stochastic Partial Differential Equations*. SpringerBriefs in Mathematics. Springer New York, NY, 1 edition, 2013.

[Zas05] Alexander Zaslavski. *Turnpike properties in the calculus of variations and optimal control*, volume 80. Springer Science & Business Media, 2005.

[Zas19] Alexander Zaslavski. *Turnpike conditions in infinite dimensional optimal control*. Springer, 2019.

Summary

The Schrödinger problem

where analysis meets stochastics

In this thesis we study the most likely behaviour of a cloud of particles subject to the information of its initial and final configuration. Despite the motivation of this problem coming from statistical mechanics, the problem itself lies at the crossroad between Optimal Transport and Stochastic Optimal Control theories, *where analysis meets stochastics*. Indeed this problem can be thought of as finding the best way possible of steering a diffusion process from a starting distribution to a prescribed final one, and at the same time it is equivalent to a regularised version of the Optimal Transport problem, whose aim is moving in the cheapest way possible some goods from an initial configuration to a final one. Particularly, the time-window parameter $T > 0$ in our Schrödinger problem acts as regularising parameter, *i.e.*, the smallest T is and the closest the particles behave according to Optimal Transport theory.

Besides being an interesting theoretical problem that connects two different fields in mathematics, in the last couple of years the Schrödinger problem has found a tremendous use in *generative modelling*, namely *creating data from noise*, which has led to an increasing interest in the study of the behaviour of its solutions and in finding efficient algorithms that allow to rapidly compute the latter. This thesis focuses exactly on these two aspects and the approach we have relied on is always based on the interplay between the analytical and the stochastic point of view.

Below we detail our main contributions.

In Chapter 3, motivated by the control interpretation, we prove the *corrector estimates* which can be interpreted as contraction estimates for the optimal control process that steers the diffusion to the target distribution. Then, we employ these estimates in order to investigate the convergence towards the optimal transport map.

In Chapter 4 we provide quantitative stability estimates for the Schrödinger problem, by relying on the corrector estimates.

In Chapter 5 and Chapter 6 we provide exponential convergence rates for Sinkhorn's algorithm, an iterative scheme that allows to compute the solution of the Schrödinger problem. We have proven this result with two different approaches: one perturbative and the other one non-perturbative.

For the first approach we have studied how Lipschitzianity propagates along Sinkhorn's algorithm. This is deduced from a more general result which proves that Lipschitzianity backward-propagates along Hamilton-Jacobi-Bellman equations, result that we prove via a stochastic optimal control approach.

The non-perturbative approach relies on studying this time how convexity propagates along Sinkhorn's algorithm. From that we have deduced the exponential convergence of the algorithm by meaning of coupling techniques.

Our last contribution, namely Chapter 7, deals with a different instance of the Schrödinger problem, where we consider particles that are described both by their position and velocity and whose density obeys to the kinetic Fokker-Planck equation. We fully characterised this new problem, despite its hypocoercive nature, and study its long time behaviour.

Acknowledgements

Moving to Eindhoven as a PhD candidate meant taking a big step for me, both metaphorically and geographically, even though I did not realise how big that step would have been, back in 2020. As a consequence I can't say that my PhD journey has been a smooth straight path, but rather a never-ending hopping from one state to another (both geographically and metaphorically, *i.e.*, state of mind). Sometimes I have even doubted I would reach the end of this journey, but here we are... and that's mostly because of the help and support I have constantly received.

My deepest gratitude goes to Alberto and Giovanni. At first, I have found in you two parents always ready to support me, patiently explaining to me (basic) mathematics, guiding and inspiring me in the PhD project. You have always found time for me and for our online meetings, which soon became our weekly routine because of Covid-19 at first and then because we were located in three different cities. Now, at the end of my PhD journey, I consider you as older cousins always there with a smart advice and words of guidance. I am looking forward to consider you as coauthors, collaborators and most importantly as friends. So che non vi piace la retorica della *famiglia matematica*, ma questo descrive esattamente come la mia considerazione nei vostri confronti sia evoluta nel corso di questi anni. Lastly, I should thank you for supporting me in my next career choices (e anche per sopportare il mio sali e scendi emotivo al riguardo).

I am proud of saying "I am their first PhD student".

In these past four years I have met different mathematicians, who contributed to my mathematical development. I will try to list them in chronological order of appearance in my life.

Dear Mark, I am extremely grateful to you. You found yourself in a very uncommon situation: dealing with a PhD student that had both his supervisors abroad, and who felt homesick. You granted me the flexibility this situation required, without which I would have not probably reached the end of this adventure. I would like to thank you also for all your comments and feedback on my work and on how I should present it in conferences.

Thank you Oliver for your presence. Having you in the office next to mine meant a lot, especially after Alberto's departure. I have always enjoyed discussing with you and I hope you are going to win your VICI soon enough so that you can invite me to Eindhoven! I am sorry that we have not met often lately (apparently our travel plans are complementary) and that you won't be able to attend my defence, though.

Carissimo Luca, thanks for working on maths with me, for being my personal RCD-guru and for explaining to me every basic maths concept that I was ashamed of asking to Alberto and Giovanni :')

Cher Alain, thank you for having always been available to work with me on tiny details when something wasn't adding up in our proof strategies. I have learned a lot working with you. I am sorry I will not join you in Paris for a PostDoc and I assure you I would have loved doing a PostDoc with you, if only I wasn't feeling that homesick (or if you were working in Italy!). I hope we will be able to work together in the near future.

Of course I am grateful to the members of the defence committee, for reading and assessing my thesis and for accepting to take part into this event (which is important to me). I hope you are going to enjoy this day as well.

Usually at this point in the acknowledgements in thesis written from CASA (our research group) you read the cliché sentence that CASA felt like home (since indeed in italian CASA stands for home). Unfortunately, for me that was not the case and I have often felt guilty about it. Probably what I have left back in Rome meant too much for me compared to what I have found in Eindhoven. Certainly my hopping from one country to another did not help. Nevertheless I have found very nice colleagues whose company cheered my days up. I will just mention two representatives: thank you Michiel for your contagious joy and spark (and remember *la Roma si e il FeyeNO!*) and thank you Antonio for always welcoming me back like San Tommaso. A special thanks goes also to CASA secretaries Diane, Enna and Gea, without whom CASA engine would not work. Particularly, I am grateful to you Diane: you were the first and only contact I had when I started during Covid-times and you have always checked upon me (moreover I have always enjoyed speaking with you in italian, that moments indeed I felt like being at home).

Within CASA, I should not forget my office mates: Anastasiia (who welcomed me at the very beginning in our office, since Jasper used to *work from home* a bit too often), Oxana (who shared my point of view on rigour and difficulty required from university teaching), Jasper (the *Dutch guy, older than my supervisors, and that I have never seen in the office* who became a friend, a person I could rely upon, emotional/career support and lastly my paranymph), Joop and Luke (unfortunately you started while I was in Paris and close to the end of my journey, I always enjoy your company in the office). Let me mention also two office-squatter whose company I have enjoyed: Evie (she joined the dark side, though) and Robert (he is a physicist, though).

I should not forget to thank my other paranymph, Kathi! You have been like an older cousin in maths (again the rhetoric of maths-family stuff, sorry!) and I have enjoyed your company during my stay in Paris and Padova. We should still work a bit on your italian visto che devi ottenere una posizione in matematica in Italia nel prossimo futuro!

A special thank to all the teachers and professors I ever had in my education because I would not be here without you. Indeed *siamo come nani sulle spalle di giganti* and you were the giants on whose shoulder I have been standing, looking forward to the future. (Tra l'altro questa frase mi è stata ovviamente detta per la prima volta e spiegata proprio da una professoressa al liceo)

I feel I should also thank Eindhoven Airport (solely for its existence, this tiny overcrowded airport is a mess), Leonardo da Vinci Rome-Fiumicino Airport (for being a wonderful, organised and huge airport where I have always enjoyed spending time), Ryanair and Wizzair since they provided me with an almost-cheap invisible bridge between Eindhoven and Ostia.

Grazie anche alla mia famiglia per sostenermi sempre, e un ringraziamento speciale va in particolare ai miei genitori sempre disposti a farmi da taxi da/per l'aeroporto, spesso ad orari improponibili.

Grazie ai miei amici e ai miei fratelli di comunità, sempre felici di riabbracciarmi. Oltre al bel clima, al cibo e al mare, voi più di tutto siete stati per me un pull factor decisivo per voler tornare a casa. La cosa più sorprendente è che il mio rapporto con alcuni di voi si sia deepened proprio durante questi miei anni di vita all'estero.

Grazie a chi ha fatto parte della mia vita, anche se ora non è più come una volta.

Grazie anche alla Zanzara di Radio24, a Cruciani e al Tigre per allietare le mie giornate e fornire un ottimo argomento di conversazione alle conferenze. A quanto pare la Zanzara va un sacco tra i giovani matematici italiani!

Ringrazio Dio per la storia che sta facendo con me, benché magari qualche strada a mio avviso sia stata un po' storta...
ma vabbe mi fido, dopotutto *le mie vie non sono le tue vie*.

This page was intentionally left blank.

Curriculum Vitae

Giacomo Greco was born on November 9th, 1996, in Rome, Italy. He grew up in Rome where he obtained his Diploma di Maturità Scientifica cum laude at Liceo Scientifico Statale "Antonio Labriola". Then, he studied Mathematics at Sapienza University of Rome. He completed his bachelor degree cum laude with the thesis "Homological algebra of unbounded complexes" in July 2018, and later graduated cum laude in Mathematics in July 2020 with the master's thesis "Signal-to-noise ratio thresholds in principal component analysis" under the supervision of Prof. A. Faggionato.

In September 2020 he moved to Eindhoven as a PhD student in M. A. Peletier group, under the supervision of Dr. A. Chiarini (University of Padova) and Prof. Giovanni Conforti (École Polytechnique - University of Padova). His daily work took place at the Centre for Analysis, Scientific computing and Applications (CASA) of the Department of Mathematics and Computer Science, with frequent visits to Padova and Paris. Particularly, in 2022 he won a PhD Travel Grant that allowed him to spend 3 consecutive months working at the École Polytechnique as Visiting PhD Student.

His PhD position was funded by the NWO Research Project 613.009 "Analysis meets Stochastics: Scaling limits in complex systems", his research was also partially funded by Nuffic in the framework of the Van Gogh Programme and with the title "The kinetic Schrödinger Problem", by the French ANR (National Research Agency) in the framework of the research project SPOT (ANR-20-CE40-0014), and by NDNS+ (NDNS/2023.004. PhD Travel Grant).

The results obtained during his PhD track are described in this thesis