EINDHOVEN UNIVERSITY OF TECHNOLOGY

Quantitative contraction rates for Sinkhorn algorithm

Citation for published version (APA):
Conforti, G., Durmus, A., & Greco, G. (2023). Quantitative contraction rates for Sinkhorn algorithm: beyond bounded costs and compact marginals. *arXiv*, *2023*, Article 2304.04451.
https://doi.org/10.48550/arXiv.2304.04451

DOI:
10.48550/arXiv.2304.04451

Document status and date:
Published: 10/04/2023

Document Version:
Publisher’s PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

Take down policy
If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.

Download date: 28. Mar. 2025

arXiv:2304.04451v1 [math.PR] 10 Apr 2023

QUANTITATIVE CONTRACTION RATES FOR SINKHORN ALGORITHM: BEYOND BOUNDED COSTS AND COMPACT MARGINALS

GIOVANNI CONFORTI, ALAIN DURMUS, AND GIACOMO GRECO

Abstract. We show non-asymptotic geometric convergence of Sinkhorn iterates to the Schrödinger potentials, solutions of the quadratic Entropic Optimal Transport problem on $\mathbb{R}^d$. Our results hold under mild assumptions on the marginal inputs: in particular, we only assume that they admit an asymptotically positive log-concavity profile, covering as special cases log-concave distributions and bounded smooth perturbations of quadratic potentials. More precisely, we provide exponential L^1 and pointwise convergence of the iterates (resp. their gradient and Hessian) to Schrödinger potentials (resp. their gradient and Hessian) for large enough values of the regularization parameter. As a corollary, we establish exponential convergence of Sinkhorn plans and bridges w.r.t. a symmetric relative entropy. Up to the authors' knowledge, these are the first results which establish geometric convergence of Sinkhorn algorithm in a general setting without assuming bounded cost functions or compactly supported marginals. Our results are proven following a probabilistic approach that rests on integrated semiconvexity estimates for Sinkhorn iterates that are of independent interest.

Contents

1. Introduction 2
2. The Sinkhorn algorithm and its geometric convergence 4
2.1. Assumptions 4
2.2. Strategy and main results 6
2.3. Literature review 13
3. Integrated semiconvexity propagation along Sinkhorn 14
4. Wasserstein distance w.r.t a measure with log-concave profile 21
5. Proof of the main results 25
5.1. Exponential convergence of the gradients along Sinkhorn algorithm 25
5.2. Exponential convergence of dual and primal Sinkhorn algorithm 28
5.3. Exponential convergence of the Hessians along Sinkhorn algorithm 31
Appendix A. Dynamical Schrödinger problem and Sinkhorn bridges 39
Appendix B. Convergence along the iterates marginals 40
Appendix C. Explicit rates of convergence 41
C.1. Explicit rates for marginals strictly log-concave 44
Appendix D. Technical results and moment bounds 46
References 48

GC acknowledges funding from the grant SPOT (ANR-20-CE40- 0014).

AD acknowledges support from the Lagrange Mathematics and Computing Research Center.

GG thanks École Polytechnique for its hospitality, where this research has been carried out, and NDNS+ for funding his visit there (NDNS/2023.004 PhD Travel Grant). GG is supported from the NWO Research Project 613.009.111 "Analysis meets Stochastics: Scaling limits in complex systems".

1. Introduction

Given two distributions $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$, we consider in this paper the problem of finding a solution to the entropic optimal transport problem:

$$\text{(EOT)} \qquad \text{minimize} \int \frac{|x-y|^2}{2}\,\mathrm{d}\pi + \varepsilon\,\mathscr{H}(\pi|\mu\otimes\nu) \text{ under the constraint } \pi\in\Pi(\mu,\nu)\,,$$

$\mathscr{H}(\cdot|\cdot)$ denotes relative entropy, $\Pi(\mu,\nu)$ is the set of couplings of μ and ν and $\varepsilon > 0$ has to be understood as a regularization parameter: the case $\varepsilon = 0$ is indeed the Monge-Kantorovich problem. EOT has recently drawn the attention of the machine learning community in light of the fact that it provides with a more convex and numerically more tractable proxy of the optimal transport problem [26]. On the other hand, the investigation of EOT is much older: its beginning can be traced back to a question posed by Schrödinger [70] about the most likely behavior of a cloud of diffusive particles conditionally on the observation of its configuration at an initial and final times. This is why EOT, when rewritten as the problem of minimizing relative entropy against the stationary Wiener measure (see SP and Appendix A below), is also known as the Schrödinger problem, see [52]. Nowadays, entropic optimal transport stands at the crossroads of different research areas in both pure applied mathematics including stochastic optimal control [16, 17, 57], geometry of optimal transport and functional inequalities [4, 24, 21, 39, 42, 43, 63], abstract metric spaces [45, 46, 58], numerics for PDEs and gradient flows [6, 7, 5, 64], as well as limit theorems in statistics [29, 54]. EOT finds a wealth applications in machine learning see [2, 3, 41] and the book [65] for a compilation of references, generative modeling [28, 27], computer vision [32, 73] and signal processing [49]. We refer to [52, 59] for surveys on SP and EOT. As hinted above, EOT recovers the Monge-Kantorovich problem in the $\varepsilon \to 0$ limit: the investigation of the connections between these two variational problems has intensified over the last years, see [1, 8, 14, 19, 20, 25, 38, 51, 56, 62, 60, 66] for a sample of relevant contributions. Most of the appeal of SP in applications can be attributed to the fact that its solutions can be efficiently computed in a very elegant and fast way via Sinkhorn algorithm, as pointed out in [26]. To introduce Sinkhorn's algorithm, that is the object of study of this work, we begin by recalling that under mild assumptions on the marginals μ,ν (see for instance [19, Proposition 2.2]), EOT admits a unique minimizer $\pi^\star_T \in \Pi(\mu,\nu)$, referred to as the Schrödinger plan, and there exist two functions $\varphi^\star \in \mathrm{L}^1(\mu)$ and $\psi^\star \in \mathrm{L}^1(\nu)$, called the Schrödinger potentials, such that the representation

$$\pi^\star_T(\mathrm{d}x\mathrm{d}y) = (2\pi T)^{-d/2}\exp\Big(-\frac{|x-y|^2}{2T} - \varphi^\star(x) - \psi^\star(y)\Big)\mathrm{d}x\,\mathrm{d}y \tag{1}$$

holds. Moreover, the couple $(\varphi^\star,\psi^\star)$ is unique up to constant translations $a \mapsto (\varphi^\star + a, \psi^\star - a)$. If we suppose that the marginals admit densities of the form

$$\mu(\mathrm{d}x) = \exp(-U_\mu(x))\mathrm{d}x\,, \qquad \nu(\mathrm{d}x) = \exp(-U_\nu(x))\mathrm{d}x\,, \tag{2}$$

then, imposing that the probability measure (1) belongs to $\Pi(\mu,\nu)$ yields the Schrödinger system

$$\begin{cases}\varphi^\star = U_\mu + \log P_T \exp(-\psi^\star)\\ \psi^\star = U_\nu + \log P_T\exp(-\varphi^\star)\,.\end{cases} \tag{3}$$

Starting from a given initialization $\psi^0\colon \mathbb{R}^d \to \mathbb{R}$, Sinkhorn's algorithm solves the Schrödinger system as a fixed point problem

$$\begin{cases} \varphi^{n+1} = U_\mu + \log P_T \exp(-\psi^n) \\ \psi^{n+1} = U_\nu + \log P_T \exp(-\varphi^{n+1}), \end{cases} \tag{4}$$

where $(P_t)_{t\geq 0}$ is the Markov semigroup generated by the standard d-dimensional Brownian motion. Perhaps the most remarkable theoretical property of Sinkhorn algorithm is its exponential convergence in the number of iterations. This fact has been rigorously established in the discrete setting and in the continuous setting assuming either boundedness of the cost or compactness of the marginals. However, the techniques developed to tackle problems with bounded costs cannot be generalized to the unbounded setting and, to the best of our knowledge and understanding, there is no result that covers unbounded costs and marginals with unbounded support, see Section 2.3 for a detailed literature overview. In this article we start filling this important conceptual gap by showing convergence of Sinkhorn iterates and its derivatives in the landmark example of the quadratic cost for a large class of unbounded marginal distributions and large enough values of the regularization parameter. To be more precise, we only require marginals to have an asymptotically positive integrated log-concavity profile, in a sense to be made precise at **A**2 below.

Organisation. The rest of the paper is organized as follows. We start Section 2 introducing Sinkhorn algorithm, then we state and comment our main assumptions in Section 2.1 and eventually present our main results and proof strategy in Section 2.2. Section 2.3 is devoted to a review of the existing literature on the convergence of Sinkhorn algorithm. In Section 3 we prove integrated semiconvexity estimates for Sinkhorn iterates and Schrödinger potential. After that, in Section 4 we show how (synchronous and reflection) couplings techniques can be employed in order to bound the Wasserstein distance with respect to a measure with an asymptotically positive log-concave profile. This result is stated in a general setting as it is of independent interest. In Section 5 we carry out the proof of the main results assembling together the results of the previous sections. In Appendix A we discuss the original formulation of Schrödinger problem as an entropy minimization problem on path space providing versions of our results that apply to dynamical Schrödinger bridges, whereas in Appendix B we give some more intuition on Sinkhorn algorithm and prove additional integrated convergence results. Appendix C provides with explicit expressions for several quantities that we use in coupling arguments: in particular Remark 35 contains an explicit formula for the contraction rates appearing in our main results. Lastly, we gather at Appendix D some technical results and moment bounds employed in Section 5.

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. 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. For any probability measure $\mu \in \mathcal{P}(\mathbb{R}^d)$ we denote with $M_k(\mu)$ its k^{th} moment and define its covariance matrix as $\mathrm{Cov}(\mu) \coloneqq \int xx^\mathsf{T}\mathrm{d}\mu - (\int x\mathrm{d}\mu)(\int x\mathrm{d}\mu)^\mathsf{T}$. For two distributions μ_1, μ_2 on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, $\Pi(\mu_1,\mu_2)$ denotes the set of couplings between μ_1 and μ_2, *i.e.*, $\xi \in \Pi(\mu_1,\mu_2)$ if and only if ξ is a probability measure on $\mathbb{R}^d\times\mathbb{R}^d$ and $\xi(\mathsf{A}\times\mathbb{R}^d) = \mu_1(\mathsf{A})$ and $\xi(\mathbb{R}^d\times\mathsf{A}) = \mu_2(\mathsf{A})$. On $\mathbb{R}^d\times\mathbb{R}^d$ we consider the projection operators $\mathrm{proj}_x(a,b) = a$ and $\mathrm{proj}_y(a,b) = b$ and we denote their pushforwards as $(\mathrm{proj}_x)_\#$ and $(\mathrm{proj}_y)_\#$. Denote by $\mathrm{L}^1(\mu_1)$ the set of function integrable with respect to μ_1. We define the relative entropy (or Kullback Leibler divergence) between μ_1 and μ_2 as $\mathscr{H}(\mu_1 \mid \mu_2) = \int_{\mathbb{R}^d} \log(\mathrm{d}\mu_1/\mathrm{d}\mu_2)\mathrm{d}\mu_1$ if $\mu_1 \ll \mu_2$ and $\mathscr{H}(\mu_1 \mid \mu_2) = +\infty$ otherwise. We denote by Leb the Lebesgue measure and define $\mathrm{Ent}(\mu_1) = \int \log(\mathrm{d}\mu_1/\mathrm{dLeb})\mathrm{dLeb}$ if $\mu_1 \ll \mathrm{Leb}$ and $+\infty$ otherwise.

2. The Sinkhorn algorithm and its geometric convergence

As specified above, Sinkhorn algorithm (4) is an iterative procedure to approximate optimal dual variables, known as Schrödinger potentials, in the entropic optimal transport problem. To gain a deeper understandig and build intuition on what the algorithm does, it is worth giving some details about the Schrödinger problem, that is an equivalent formulation of EOT inspired by Schrödinger's seminal work [70] about the behavior of Brownian particles conditionally to observations, see [52, Sec. 6] for an enlightening discussion on the statistical mechanics motivation behind Schrödinger's question. Given two probability measures μ, ν on $\mathbb{R}^d$ and an horizon $T>0$, SP consists in solving the optimization problem

$$\text{minimize } \mathscr{H}(\pi|\mathrm{R}_{0,T}) \text{ under the constraint } \pi\in\Pi(\mu,\nu)\ , \tag{SP}$$

where $\Pi(\mu,\nu)$ denotes the set of of couplings between μ and ν and $\mathscr{H}(\pi|\mathrm{R}_{0,T})$ is the relative entropy between the coupling π and $\mathrm{R}_{0,T}$ defined by

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

Indeed if $\mathrm{Ent}(\mu_1)$ denotes the relative entropy with respect to the Lebesgue measure, setting $\varepsilon=T$ the identity

$$T\,\mathscr{H}(\pi|\mathrm{R}_{0,T})-T\,\mathrm{Ent}(\mu)-T\,\mathrm{Ent}(\nu)=\int\frac{|x-y|^2}{2}\,\mathrm{d}\pi+T\,\mathscr{H}(\pi|\mu\otimes\nu)\qquad\text{for any }\pi\in\Pi(\mu,\nu)$$

immediately implies the equivalence between SP and EOT. In particular, the regularization parameter ε in EOT has to be understood as the length T of the time interval in between two observations of a particle system in SP. At a high level, it has been pointed out [6] that Sinkhorn is a special case of Bregman's iterated projection algorithm in that it alternates between computing an entropic projection on the set of probability measures with first marginal μ and computing an entropic projection on the set of measures with second marginal ν. More precisely, when ψ^n is given and we compute the next iterate φ^{n+1}, we are implicitly prescribing the couple (φ^{n+1},ψ^n) to fit the first marginal constraint, *i.e.*, we are imposing the first marginal of the probability measure

$$\pi^{n+1,n}(\mathrm{d}x\mathrm{d}y)\propto\exp\Big(-\frac{|x-y|^2}{2T}-\varphi^{n+1}(x)-\psi^n(y)\Big) \tag{5}$$

to be exactly μ. At the next iteration, when computing ψ^{n+1}, the probability measure

$$\pi^{n+1,n+1}(\mathrm{d}x\mathrm{d}y)\propto\exp\Big(-\frac{|x-y|^2}{2T}-\varphi^{n+1}(x)-\psi^{n+1}(y)\Big) \tag{6}$$

violates the first marginal constraint but satisfies the second one: its second marginal is equal to ν. We will refer to the above couplings as the Sinkhorn plans. Introducing Sinkhorn plans makes the connection with the iterated Bregman projection algorithm even more transparent. Indeed one can check (see [59, Sec. 5]) that Sinkhorn algorithm is equivalent to minimizing at each step relative entropy w.r.t. the previous plan subject to a one-sided marginal constraint, *i.e.*,

$$\pi^{n+1,n}:=\arg\min\nolimits_{\Pi(\mu,\star)}\mathscr{H}(\cdot|\pi^{n,n})\,,\qquad\pi^{n+1,n+1}:=\arg\min\nolimits_{\Pi(\star,\nu)}\mathscr{H}(\cdot|\pi^{n+1,n})\,, \tag{7}$$

where $\Pi(\mu,\star)$ (resp. $\Pi(\star,\nu)$) is the set of probability measures on $\mathbb{R}^{2d}$ such that the first marginal is μ (resp. the second marginal is ν). For later convenience let us define the adjusted marginals produced along Sinkhorn as the probability measures $\mu^n:=(\mathrm{proj}_x)_{\#}\pi^{n,n}$ and $\nu^n:=(\mathrm{proj}_y)_{\#}\pi^{n+1,n}$.

2.1. Assumptions.

Informal description. To state the assumptions we impose on marginal distributions, we introduce the integrated convexity profile $\kappa_U : \mathbb{R}_+^* \to \mathbb{R}$ of a function $U : \mathbb{R}^d \to \mathbb{R}$ 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 μ represented as a Gibbs measure as in (2), we refer to κ_{U_μ} as to the integrated log-concavity profile of μ. The function κ_U is often employed to quantify ergodicity of stochastic differential equations whose drift field is $-\nabla U$, see [34]. The term integrated convexity profile we coined here is motivated by the observation that

$$\kappa_U(r) \geq \alpha \Leftrightarrow \int_0^r \langle \nabla^2 U(x+hv)v, v\rangle \mathrm{d}h \geq \alpha r \quad \forall\, x, v \in \mathbb{R}^d \quad \text{s.t. } |v| = 1.$$

Therefore $\kappa_U(r) \geq \alpha$ is equivalent to say that, U behaves like an α-convex function for points at distance r. In particular, U is α-convex if and only if $\inf_{r>0} \kappa_U(r) \geq \alpha$. The integrated concavity profile of U is defined in a similar way as $\ell_U = -\kappa_{-U}$. Our main results apply when the marginals μ, ν satisfy the following property

$$\begin{aligned} \liminf_{r\to+\infty} \kappa_{U_\mu}(r) > 0, &\quad \liminf_{r\to+\infty} \kappa_{U_\nu}(r) > 0,\\ \liminf_{r\to 0} r\kappa_{U_\mu}(r) = 0, &\quad \liminf_{r\to 0} r\kappa_{U_\nu}(r) = 0. \end{aligned} \tag{8}$$

Convergence rates can be improved if we have additional information on the integrated concavity profile such as

$$\limsup_{r\to+\infty} \ell_{U_\nu}(r) < \infty, \quad \liminf_{r\to+\infty} \ell_{U_\nu}(r) < \infty. \tag{9}$$

or some weaker versions of the above (see the precise statement of **A**2 below). However, we stress that the (9) and variants thereof are not necessary conditions. To obtain the sharpest possible convergence rates, we will choose a particular parametric description of the set (8) and of its intersection with (9) through two set of functions, called $\mathcal{G}$ and $\mathcal{G}_\kappa$ below. Though it may not appear as the most natural at first sight, these choices will become clear in light of proofs. Indeed, the sets $\mathcal{G}$ and $\mathcal{G}_\kappa$ enjoy some natural and powerful invariance properties (see Theorem 14 and Lemma 21) under the mapping

$$g \mapsto -\log P_T \exp(-g)(\cdot),$$

around which the proof of the lower bounds on the integrated convexity profiles of potentials (see Section 3) are built. In turn, these estimates are at the core of our whole proof architecture.

Precise statement. We begin by introducing the two sets of functions mentioned above.

$$\mathcal{G} := \left\{ g \in \mathcal{C}^2((0,+\infty),\, \mathbb{R}_+) \quad \text{s.t.} \quad \begin{aligned} &(r \mapsto r\, g(r)) \text{ is non-decreasing},\\ &\lim_{r\downarrow 0} r\, g(r) = 0\ ,\\ &g''(r) + r^{-1}\, g'(r) - r^{-2}\, g(r) \leq 0\ . \end{aligned} \right\}$$

and its subset

$$\mathcal{G}_\kappa := \left\{ g \in \mathcal{G} \text{ bounded and s.t.} \quad \lim_{r\downarrow 0} g(r) = 0\ , \quad g' \geq 0 \quad \text{and} \quad 2g'' + g\, g' \leq 0 \right\} \subset \mathcal{G}\ .$$

The above classes of functions are non-empty, for instance the hyperbolic tangent function $4\tanh$ is a non-trivial example of function belonging to $\mathcal{G}_\kappa$. Our main results require the following assumptions.

A 1. *The two distributions $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$, have finite relative entropy with respect to the Lebesgue measure* Leb*,* $\mathrm{Ent}(\mu)$, $\mathrm{Ent}(\nu) < +\infty$*, and satisfy* (2) *with log-densities* U_μ, $U_\nu \in \mathcal{C}^2(\mathbb{R}^d)$.

A 2.

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

$$\kappa_{U_\nu}(r) \geq \alpha_\nu - r^{-1}\, g_\nu^\kappa(r) \quad \text{and} \quad \ell_{U_\mu}(r) \leq \beta_\mu + r^{-1}\, g_\mu^\ell(r)\ ,$$

for some $g_\nu^\kappa \in \mathcal{G}_\kappa$ and $g_\mu^\ell \in \mathcal{G}$;

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

$$\kappa_{U_\mu}(r) \geq \alpha_\mu - r^{-1}\, g_\mu^\kappa(r) \quad \text{and} \quad \ell_{U_\nu}(r) \leq \beta_\nu + r^{-1}\, g_\nu^\ell(r)\ ,$$

for some $g_\mu^\kappa \in \mathcal{G}_\kappa$ and $g_\nu^\ell \in \mathcal{G}$.

Let us remark here that in the above assumptions we are allowed to consider $\beta = +\infty$. When $\beta = +\infty$, we adopt the convention $\beta^{-1} = 0$ and $(a,\, a + \beta^{-1}] \coloneqq \{a\}$.

Remark 1 (Lipschitz perturbations of convex potentials)**.** *If U_ν rewrites as the sum of a strongly convex potential and a Lipschitz potential with first derivative that is p-Hölder continuous for some $p > 0$ or with second derivative bounded below, then item (i) in **A** 2 holds. Note that this does not necessarily imply that U_ν is a bounded perturbation of a convex function or that U_ν is convex outside a ball.*

Remark 2 (Asymptotically positive convexity profile)**.** *Let us remark here that the assumption on the lower-bound for κ_U is satisfied by a large class of potentials U. For instance if $U\colon \mathbb{R}^d \to \mathbb{R}$ satisfies*

$$\kappa_U(r) \geq \begin{cases} \alpha & \text{if } r > R \\ \alpha - L' & \text{if } r \leq R\,, \end{cases}$$

for some positive parameters α, $L' > 0$ and radius $R > 0$, then [23, Proposition 5.1] *implies that*

$$\kappa_U(r) \geq \alpha - r^{-1}\, g_L^{\kappa_U}(r)$$

where $g_L^{\kappa_U} \in \mathcal{G}_\kappa$ is given by

$$g_L^{\kappa_U}(r) \coloneqq (2L)^{1/2}\, \tanh\!\left(\frac{r}{4}\,(2L)^{1/2}\right) \quad \text{with} \quad L \coloneqq \inf\{\bar{L} :\ R^{-1}\, g_{\bar{L}}^{\kappa_U}(R) \geq L'\}\ .$$

We will prove in Section 3 that the previous set of assumptions guarantees integrated semiconvexity not only for the Schrödinger potentials but also for Sinkhorn iterates, extending the results of [23] to Sinkhorn iterates and under milder assumptions. This transfer of integrated log-concavity from marginals into integrated semiconvexity of potentials is at the core of our proof strategy. Moreover, [23] considers a more restrictive situation where $\mathcal{G}_\kappa$ contains only rescaled versions of the hyperbolic tangent function and the only element of $\mathcal{G}$ is the null function.

2.2. **Strategy and main results.**

Proof strategy. The starting point of our discussion is the following well known result, see [67, Proposition 2] or [18] for instance.

Proposition 3. *Assume **A** 1 and **A** 2-(i). Then for any $h \in \{\psi^\star, \psi^n\}$ it holds*

$$
\begin{aligned}
\nabla \log P_T \exp(-h)(x) &= T^{-1} \int (y-x)\, \pi_T^{x,h}(\mathrm{d}y)\,, \\
\nabla^2 \log P_T \exp(-h)(x) &= -T^{-1}\,\mathrm{Id} + T^{-2}\,\mathrm{Cov}(\pi_T^{x,h})\,,
\end{aligned} \tag{10}
$$

*where $(x, A) \mapsto \pi_T^{x,h}(A)$ is the Markov kernel on $\mathbb{R}^d \times \mathcal{B}(\mathbb{R}^d)$ whose transition density w.r.t. Lebesgue measure is proportional to $\exp(-h(y) - |x-y|^2/(2T))$. The same identity holds true for $h \in \{\varphi^\star, \varphi^n\}$ assuming **A** 1 and **A** 2-(ii).*

The proof of this technical result is postponed to Appendix D. To give a concise sketch of our proof strategy, we record here two observations. The first is that for any $x \in \mathbb{R}^d$ the conditional distribution

$$
\pi_T^{x,h}(\mathrm{d}y) \propto \exp\Big(-\frac{|y-x|^2}{2T} - h(y)\Big)\mathrm{d}y \tag{11}
$$

may be viewed as the invariant probability measure for the SDE

$$
\mathrm{d}Y_t = -\Big(\frac{Y_t - x}{2T} + \frac{1}{2}\nabla h(Y_t)\Big)\mathrm{d}t + \mathrm{d}B_t\,. \tag{12}
$$

The second one is that, setting $\mu^n = \mathrm{proj}_x \pi^{n,n}$, the Schrödinger and Sinkhorn plans can be written as

$$
\pi_T^\star(\mathrm{d}x, \mathrm{d}y) = \mu(\mathrm{d}x) \otimes \pi_T^{x,\psi^\star}(\mathrm{d}y) \quad \text{and} \quad \pi^{n,n}(\mathrm{d}x, \mathrm{d}y) = \mu^n(\mathrm{d}x) \otimes \pi_T^{x,\psi^n}(\mathrm{d}y)\,, \tag{13}
$$

and a direct computation shows that

$$
\int \pi_T^{x,\psi^\star}(\mathrm{d}y)\, \mu(\mathrm{d}x) = \nu(\mathrm{d}y). \qquad \int \pi_T^{x,\psi^n}(\mathrm{d}y)\, \mu^n(\mathrm{d}x) = \nu(\mathrm{d}y)\,. \tag{14}
$$

Let us now proceed to give a minimal description of our proof strategy. From (4) and the formula (10) with choices $h = \psi^n, \psi^\star$ we deduce the upper bound

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

Combining together the lower bounds on the integrated convexity profiles κ_{ψ^n} (see Section 3), the representation of $\pi_T^{x,\psi^n}, \pi_T^{x,\psi^\star}$ as invariant measures for the SDE (12) and coupling techniques (see Section 4), we obtain at Corollary 24 that the key estimate

$$
\mathcal{W}_1(\pi_T^{x,\psi^n}, \pi_T^{x,\psi^\star}) \leq \gamma_n^\nu \int |\nabla\psi^n - \nabla\psi|(y)\, \pi_T^{x,\psi^\star}(\mathrm{d}y)
$$

holds uniformly on $x \in \mathbb{R}^d$ for some $\gamma_n^\nu > 0$. Integrating the above w.r.t. μ and invoking (14) yields

$$
\int |\nabla\varphi^{n+1} - \nabla\varphi^\star|(x)\, \mu(\mathrm{d}x) \leq \frac{\gamma_n^\nu}{T} \int |\nabla\psi^n - \nabla\psi|(y)\nu(\mathrm{d}y). \tag{15}
$$

Repeating the same argument but exchanging the roles of $\psi^n, \psi^\star$ and $\varphi^n, \varphi^\star$ we obtain

$$
\int |\nabla\psi^n - \nabla\psi^\star|(y)\, \nu(\mathrm{d}y) \leq \frac{\gamma_{n-1}^\mu}{T} \int |\nabla\varphi^n - \nabla\varphi|(x)\mu(\mathrm{d}x) \tag{16}
$$

for some $\gamma^{\mu}_{n-1} > 0$. Concatenating (16) with (15) allows to establish the sought exponential convergence provided $T^{-2}\gamma^{\nu}_n\gamma^{\mu}_{n-1} < 1$ for n large enough.

Statement of the main results. Our first contribution is the L^1-convergence of Sinkhorn iterates.

Theorem 4. *Assume the validity of* ***A**1* *and* ***A**2*. Then the estimates*

$$
\begin{aligned}
\int |\nabla\varphi^n - \nabla\varphi^\star|\mathrm{d}\mu &\leq \frac{T}{\gamma^{\mu}_{n-1}} \prod_{k=0}^{n-1} \frac{\gamma^{\mu}_k\,\gamma^{\nu}_k}{T^2} \int |\nabla\psi^0 - \nabla\psi^\star|\mathrm{d}\nu\,,\\
\int |\nabla\psi^n - \nabla\psi^\star|\mathrm{d}\nu &\leq \prod_{k=0}^{n-1} \frac{\gamma^{\mu}_k\,\gamma^{\nu}_k}{T^2} \int |\nabla\psi^0 - \nabla\psi^\star|\mathrm{d}\nu\,,
\end{aligned}
\tag{17}
$$

hold uniformly on $n \geq 1$, where $(\gamma^{\mu}_k)_{k\in\mathbb{N}}$ and $(\gamma^{\nu}_k)_{k\in\mathbb{N}}$ are monotone decreasing sequences given in Remark 35, depending respectively on α_μ, β_ν, g^{κ}_μ, g^{ℓ}_ν, T and on α_ν, β_μ, g^{κ}_ν, g^{ℓ}_μ, T, and converging to some finite γ^{μ}_∞ and γ^{ν}_∞. 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 $n \in \mathbb{N}^*$,

$$
\begin{aligned}
\int |\nabla\varphi^n - \nabla\varphi^\star|\mathrm{d}\mu \leq& C\lambda^n \int |\nabla\psi^0 - \nabla\psi^\star|\mathrm{d}\nu\,,\\
\int |\nabla\psi^n - \nabla\psi^\star|\mathrm{d}\nu \leq& C\lambda^n \int |\nabla\psi^0 - \nabla\psi^\star|\mathrm{d}\nu\,.
\end{aligned}
$$

Remark 5. *It can be deduced from the fully explicit form of $\gamma^{\mu}_\infty, \gamma^{\nu}_\infty$ that the condition for exponential convergence $T^{-2}\gamma^{\mu}_\infty\gamma^{\nu}_\infty < 1$ is always satisfied for large enough values of T. The minimal value of T needed for this can also be estimated, though calculations are a bit involved. When marginals are strongly log-concave, the situation becomes easier and we carry out calculations explicitly at Example 6 below.*

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 result is postponed to Appendix B. Let us now give explicit bounds on the minimal value of T starting from which exponential convergence holds in the case when marginals are strongly log-concave. In this setting, the coefficients $\gamma^{\mu}_\infty, \gamma^{\nu}_\infty$ can be expressed in a very simple way, as it had already been noticed in [18].

Example 6 (Log-concave marginals)**.** *In the (strictly) log-concave regime, i.e., for $g^{\kappa}_\mu = g^{\kappa}_\nu = g^{\ell}_\mu = g^{\ell}_\nu \equiv 0$ in* ***A**2*, the above (explicit) rates of convergence are given by*

$$
\begin{cases}
\gamma^{\mu}_0 \coloneqq \alpha_\mu^{-1}\\
\gamma^{\mu}_{k+1} \coloneqq (\alpha_\mu + (T^2\,\beta_\nu + \gamma^{\mu}_k)^{-1})^{-1}
\end{cases}
\qquad \text{and} \qquad
\begin{cases}
\gamma^{\nu}_0 \coloneqq \alpha_\nu^{-1}\\
\gamma^{\nu}_{k+1} \coloneqq (\alpha_\nu + (T^2\,\beta_\mu + \gamma^{\nu}_k)^{-1})^{-1}
\end{cases}
$$

The above sequences are monotone increasing and converging respectively to the limit rates

$$
\gamma^{\mu}_\infty \coloneqq 2\left(\alpha_\mu + \sqrt{\alpha_\mu^2 + 4\alpha_\mu/(T^2\beta_\nu)}\right)^{-1} \qquad \text{and} \qquad \gamma^{\nu}_\infty \coloneqq 2\left(\alpha_\nu + \sqrt{\alpha_\nu^2 + 4\alpha_\nu/(T^2\beta_\mu)}\right)^{-1}\,.
$$

As a consequence, we deduce that for any $\theta \in (0,\infty)$ it holds

$$
\begin{aligned}
T > \theta\,\alpha_\mu^{-1} - \theta^{-1}\beta_\nu^{-1} \quad &\Leftrightarrow \quad T > \theta\,\gamma^{\mu}_\infty\,,\\
T > \theta^{-1}\alpha_\nu^{-1} - \theta\,\beta_\mu^{-1} \quad &\Leftrightarrow \quad T > \theta^{-1}\,\gamma^{\nu}_\infty\,,
\end{aligned}
\tag{18}
$$

and therefore if

$$T > \inf_{\theta\in(0,\infty)} \max\{\theta\,\alpha_\mu^{-1} - \theta^{-1}\beta_\nu^{-1},\, \theta^{-1}\alpha_\nu^{-1} - \theta\,\beta_\mu^{-1}\} = \frac{\alpha_\mu^{-1}\alpha_\nu^{-1} - \beta_\mu^{-1}\beta_\nu^{-1}}{\sqrt{(\alpha_\mu^{-1}+\beta_\mu^{-1})(\alpha_\nu^{-1}+\beta_\nu^{-1})}} = \frac{\beta_\mu\beta_\nu - \alpha_\mu\alpha_\nu}{\sqrt{\alpha_\mu\,\beta_\mu\,\alpha_\nu\,\beta_\nu\,(\alpha_\mu+\beta_\mu)(\alpha_\nu+\beta_\nu)}}\,,$$

then we are guaranteed that $T^2 > \gamma_\infty^\mu\,\gamma_\infty^\nu$*, and hence the exponential convergence of Sinkhorn algorithm. For completeness, we sketch the details in Appendix C.1.*

Our coupling approach can also be employed in order to prove a pointwise convergence result. Obviously, in order to to achieve pointwise convergence the assumptions we impose on the regularization parameter T are more stringent than the ones we need for L^1-convergence. Note that the next result requires the additional assumption (19) concerning the initialization of the algorithm. As we explain below there is a natural choice for ψ^0 that guarantees that (19) holds.

Theorem 7. *Assume the validity of* ***A****1 and* ***A****2 and that there exist two positive constants* A, $B>0$ *such that for any* $x\in\mathbb{R}^d$

$$|\nabla\psi^0 - \nabla\psi^\star|(x) \le A\,|x| + B\,. \tag{19}$$

Then the estimates

$$\begin{aligned} |\nabla\varphi^n - \nabla\varphi^\star|(x) &\le \frac{1}{\hat\gamma_{n-1}^\mu}\prod_{k=0}^{n-1}\hat\gamma_k^\mu\,\hat\gamma_k^\nu(A|x|+B)\\ |\nabla\psi^n - \nabla\psi^\star|(x) &\le \prod_{k=0}^{n-1}\hat\gamma_k^\mu\,\hat\gamma_k^\nu\,(A|x|+B) \end{aligned} \tag{20}$$

hold uniformly on $n\ge1$ *and* $x\in\mathbb{R}^d$*, where the rates* $\hat\gamma_k^\mu$ *and* $\hat\gamma_k^\nu$ *are equal to*

$$\hat\gamma_k^\mu := T^{-1}\,\gamma_k^\mu\;\max\left\{(T\,\alpha_{\varphi^\star}+1)^{-1}\,,\left(1+\frac{A}{B}\frac{1+\|g_\mu^\kappa\|_\infty+|\nabla\varphi^\star(0)|}{\alpha_{\varphi^\star}+T^{-1}}\right)\right\},$$

and

$$\hat\gamma_k^\nu := T^{-1}\,\gamma_k^\nu\;\max\left\{(T\,\alpha_{\psi^\star}+1)^{-1}\,,\left(1+\frac{A}{B}\frac{1+\|g_\nu^\kappa\|_\infty+|\nabla\psi^\star(0)|}{\alpha_{\psi^\star}+T^{-1}}\right)\right\},$$

with γ_k^μ, γ_k^ν *defined in Remark 35, whereas* $\alpha_{\psi^\star}\in(\alpha_\nu - T^{-1}, \alpha_\nu - T^{-1} + (\beta_\mu T^2)^{-1}]$ *and* $\alpha_{\psi^\star}\in(\alpha_\nu - T^{-1}, \alpha_\nu - T^{-1} + (\beta_\mu T^2)^{-1}]$ *are the scalar lower-bounds for the convexity profiles of* $\varphi^\star$ *and* $\psi^\star$ *respectively, as in Theorem 14. In particular, if* T *is large enough, e.g., if*

$$T > \max\left\{\alpha_\mu^{-1}\,,\;\alpha_\nu^{-1}\,,\;\gamma_\infty^\mu + \frac{\gamma_\infty^\mu A}{\alpha_\mu B}(1+\|g_\mu^\kappa\|_\infty+|\nabla\varphi^\star(0)|)\,,\;\gamma_\infty^\nu + \frac{\gamma_\infty^\nu A}{\alpha_\nu B}(1+\|g_\nu^\kappa\|_\infty+|\nabla\psi^\star(0)|)\right\}, \tag{21}$$

the asymptotic geometric rate $\hat\gamma_\infty^\mu\,\hat\gamma_\infty^\nu$ *is strictly less than one, and as a result, for any* $\hat\gamma_\infty^\mu\,\hat\gamma_\infty^\nu < \lambda < 1$*, there exists* $C\ge0$ *(independent of* $x\in\mathbb{R}^d$*) such that for any* $n\in\mathbb{N}^*$*,*

$$|\nabla\varphi^n - \nabla\varphi^\star|(x) \le C\,\lambda^n(A|x|+B) \quad\text{and}\quad |\nabla\psi^n - \nabla\psi^\star|(x) \le C\,\lambda^n\,(A|x|+B)\,.$$

Example 8 (Log-concave marginals (continued)). *In the (strictly) log-concave regime, the above (explicit) rates of convergence read as*

$$\hat{\gamma}^{\mu}_k \coloneqq T^{-1}\,\gamma^{\mu}_k\ \max\Big\{T^{-1}\,\gamma^{\mu}_\infty\,,\,\Big(1+\gamma^{\mu}_\infty\,\frac{A(1+|\nabla\varphi^\star(0)|)}{B}\Big)\Big\}\,,$$
$$\hat{\gamma}^{\nu}_k \coloneqq T^{-1}\,\gamma^{\nu}_k\ \max\Big\{T^{-1}\,\gamma^{\nu}_\infty\,,\,\Big(1+\gamma^{\nu}_\infty\,\frac{A(1+|\nabla\psi^\star(0)|)}{B}\Big)\Big\}\,,$$

with γ^{μ}_∞ and γ^{ν}_∞ as in Example 6. We postpone the details to Appendix C.1.

We stress that the previous theorem holds for any smooth initialization $\psi^0 \in \mathcal{C}^1(\mathbb{R}^d)$ satisfying (19). A common choice would be starting at $\psi^0 = U_\nu$, which corresponds in terms of EOT and entropic potentials as having a null-initialization for Sinkhorn algorithm.[1] This choice agrees with the usual normalization imposed to the Sinkhorn iterates when studying its convergence [31, 13, 30, 12]. Let us also point out that if one starts the Sinkhorn algorithm one step before with the null initialization $\varphi^0 \coloneqq 0$, then at the first iteration we immediately get the desired $\psi^0 = U_\nu$. Under this choice, from (3) we deduce that

$$\psi^0 - \psi^\star = -\log P_T \exp(-\varphi^\star)$$

hence

$$|\nabla\psi^0-\nabla\psi^\star|(x) = \frac{1}{T}\left|\int (y-x)\,\pi_T^{x,\varphi^\star}(\mathrm{d}y)\right| \le \frac{1}{T}\int |y-x|\,\pi_T^{x,\varphi^\star}(\mathrm{d}y) \le \frac{|x|}{T}+\frac{1}{T}\int |y|\,\mathrm{d}\pi_T^{x,\varphi^\star}(y)\,. \tag{22}$$

The latter combined with the bound we are going to give later in (63) shows that the initialization $\psi^0 = U_\nu$ automatically satisfies the linear growth condition of Theorem 7. At the same time, integrating (22) w.r.t. ν and using (13) (exchanging the roles between first and second marginal) yields to

$$\int |\nabla\psi^0 - \nabla\psi^\star|\,\mathrm{d}\nu \le T^{-1}\int |x-y|\,\pi_T^{x,\varphi^\star}(\mathrm{d}y)\,\nu(\mathrm{d}x) = T^{-1}\int |x-y|\,\mathrm{d}\pi^\star_T$$

which allows us to state (17) in terms of the first moments of the marginals μ and ν. Summarizing the above discussion, if we start from $\psi^0 = U_\nu$, the previous results read as

Corollary 9. *Assume the validity of **A** 1 and **A** 2. If we set the initial Sinkhorn iterate equal to $\psi^0 = U_\nu$ (or equivalently if we start with $\varphi^0 = 0$), then the linear-growth condition (19) is satisfied with*

$$A = T^{-1}\,\frac{T\,\alpha_{\varphi^\star}+2}{T\,\alpha_{\varphi^\star}+1} \qquad \text{and} \qquad B = \frac{1+\|g^\kappa_\mu\|_\infty + |\nabla\varphi^\star(0)|}{T\,\alpha_{\varphi^\star}+1}\,. \tag{23}$$

Moreover, under the same initialization choice, the integrated bounds (17) read as

$$\int \nabla\varphi^n - \nabla\varphi^\star|(x)\mu(dx) \le \frac{1}{\gamma^{\mu}_{n-1}}\prod_{k=0}^{n-1}\frac{\gamma^{\mu}_k\,\gamma^{\nu}_k}{T^2}\,(M_1(\mu)+M_1(\nu))\,,$$
$$\int |\nabla\psi^n - \nabla\psi^\star|(y)\nu(dy) \le \prod_{k=0}^{n-1}\frac{\gamma^{\mu}_k\,\gamma^{\nu}_k}{T^2}\,(M_1(\mu)+M_1(\nu))\,.$$

[1]Indeed let us just stress here that if $\varphi^\star$ and $\psi^\star$ denote the Schrödinger potentials, then the functions $-T\log P_T\exp(-\psi^\star)$ and $-T\log P_T\exp(-\varphi^\star)$ correspond to the couple of entropic potentials commonly considered in EOT literature.

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 7.

Example 10 (Log-concave marginals (continued))**.** *In the (strictly) log-concave regime, if $\psi^0 = U_\nu$ (i.e., $\varphi^0 = 0$) the convergence rates appearing in Theorem 7 read as*

$$\hat\gamma_k^\mu \coloneqq T^{-1}\,\gamma_k^\mu\ \max\left\{T^{-1}\,\gamma_\infty^\mu\ ,\ 2+T^{-1}\,\gamma_\infty^\mu\right\} = T^{-1}\,\gamma_k^\mu\ (2+T^{-1}\,\gamma_\infty^\mu)\ ,$$
$$\hat\gamma_k^\nu \coloneqq T^{-1}\,\gamma_k^\nu\ \max\left\{T^{-1}\,\gamma_\infty^\nu\ ,\left(1+\left(T^{-1}\,\gamma_\infty^\nu+\frac{\gamma_\infty^\nu}{\gamma_\infty^\mu}\right)\frac{1+|\nabla\psi^\star(0)|}{1+|\nabla\varphi^\star(0)|}\right)\right\}\ ,$$

with γ_∞^μ and γ_∞^ν as in Example 6. Notice that if for instance we assume the validity of (18) for $\theta = 1$, then the asymptotic rates reads as

$$\begin{aligned}\hat\gamma_\infty^\mu &= T^{-1}\,\gamma_\infty^\mu\ (2+T^{-1}\,\gamma_\infty^\mu) < 3\,T^{-1}\,\gamma_\infty^\mu\ ,\\ \hat\gamma_\infty^\nu &= T^{-1}\,\gamma_\infty^\nu\left(1+\left(T^{-1}\,\gamma_\infty^\nu+\frac{\gamma_\infty^\nu}{\gamma_\infty^\mu}\right)\frac{1+|\nabla\psi^\star(0)|}{1+|\nabla\varphi^\star(0)|}\right) < T^{-1}\,\gamma_\infty^\nu\left(1+M\,\frac{1+|\nabla\psi^\star(0)|}{1+|\nabla\varphi^\star(0)|}\right)\ ,\\ &\text{with } M = 1+\sup_{s\ge0}\frac{\alpha_\mu\,s+\sqrt{\alpha_\mu^2\,s^2+4\alpha_\mu/\beta_\nu}}{\alpha_\nu\,s+\sqrt{\alpha_\nu^2\,s^2+4\alpha_\nu/\beta_\mu}} < +\infty\ .\end{aligned} \tag{24}$$

Therefore Sinkhorn algorithm converges exponentially fast if for instance it holds

$$T > \max\left\{3\,\alpha_\mu^{-1}-3^{-1}\beta_\nu^{-1},\ \left(1+M\,\frac{1+|\nabla\psi^\star(0)|}{1+|\nabla\varphi^\star(0)|}\right)\alpha_\nu^{-1}-\left(1+M\,\frac{1+|\nabla\psi^\star(0)|}{1+|\nabla\varphi^\star(0)|}\right)^{-1}\beta_\mu^{-1}\right\}\ . \tag{25}$$

We postpone the details to Appendix C.1.

It is possible to infer convergence of Sinkhorn iterates $(\varphi^n)_{n\in\mathbb{N}}$ and $(\psi^n)_{n\in\mathbb{N}}$. from the convergence of gradients. Since Schrödinger potentials are unique up to a trivial additive shift, we have to impose a normalization. In particular, we consider the symmetric normalization

$$\int\varphi^\star\,\mathrm{d}\mu+\mathrm{Ent}(\mu) = \int\psi^\star\,\mathrm{d}\nu+\mathrm{Ent}(\nu) = \frac12\left(\mathrm{Ent}(\mu)+\mathrm{Ent}(\nu)-\mathscr{H}(\pi_T^\star|\mathrm{R}_{0,T})\right)\ . \tag{26}$$

This normalization has already been used while showing convergence of rescaled Schrödinger potentials and thier gradients to Kantorovich potentials and the Brenier map respectively [60, 19]. In what concerns iterates, we work with the following normalization

$$\varphi^{\diamond n} = \varphi^n-\left(\int\varphi^n\mathrm{d}\mu-\int\varphi^\star\mathrm{d}\mu\right)\ ,\qquad \psi^{\diamond n} = \psi^n-\left(\int\psi^n\mathrm{d}\nu-\int\psi^\star\mathrm{d}\nu\right), \tag{27}$$

This choice guarantees that

$$\int\varphi^{\diamond n}\mathrm{d}\mu+\mathrm{Ent}(\mu) = \int\varphi^\star\mathrm{d}\mu+\mathrm{Ent}(\mu) = \int\psi^{\diamond n}\mathrm{d}\nu+\mathrm{Ent}(\nu) = \int\psi^\star\mathrm{d}\nu+\mathrm{Ent}(\nu)\ .$$

Theorem 11. *Assume that **A**1, **A**2 and (19) hold. Then for any $n \geq 1$ and $x \in \mathbb{R}^d$ it holds*

$$|\varphi^{\diamond n}-\varphi^\star|(x) \le \frac{1}{\hat\gamma_{n-1}^\mu}\prod_{k=0}^{n-1}\hat\gamma_k^\mu\,\hat\gamma_k^\nu\left[A\,|x|^2+(A\,M_1(\mu)+B)\,|x|+B\,M_1(\mu)+2\,A\,M_2(\mu)\right]\ ,$$
$$|\psi^{\diamond n}-\psi^\star|(x) \le \prod_{k=0}^{n-1}\hat\gamma_k^\mu\,\hat\gamma_k^\nu\left[A\,|x|^2+(A\,M_1(\nu)+B)\,|x|+B\,M_1(\nu)+2\,A\,M_2(\nu)\right]\ .$$

Hence, for T large enough (e.g., (21)), for any $\hat\gamma^\mu_\infty\,\hat\gamma^\nu_\infty < \lambda < 1$ (independent of $x \in \mathbb{R}^d$), there exists $C_x \geq 0$ such that for any $n \in \mathbb{N}^*$ it holds

$$|\varphi^{\diamond n} - \varphi^\star|(x) \leq\; C_x\, \lambda^n \quad \text{ and } \quad |\psi^{\diamond n} - \psi^\star|(x) \leq\; C_x\, \lambda^n \;.$$

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 (23).

Let us mention here that a straightforward adaptation of the proof of Theorem 11 implies that this result also holds true under a pointwise normalization (e.g., $\psi^\star(0) = \psi^{\diamond n}(0) = U_\nu(0)$) or for the symmetric zero-mean option considered in [31, 13, 30, 12].

We now move on to consider exponential convergence of Sinkhorn algorithm on the primal side, *i.e.*, for the Sinkhorn plans $(\pi^{n,n})_{n\in\mathbb{N}}$ and $(\pi^{n+1,n})_{n\in\mathbb{N}}$ and for the adjusted marginals $(\mu^n)_{n\in\mathbb{N}}$ and $(\nu^n)_{n\in\mathbb{N}}$, using a symmetrized version of the relative entropy; let us explain here why this choice is the most adapted to Sinkhorn's setting. As observed in [19, 44], measuring distances between plans with this divergence leads to tractable expressions.

Theorem 12 (Exponential convergence of Sinkhorn on the primal side)**.** *Assume that **A** 1,**A** 2 and (19) hold. Then, for any $n \geq 1$ it holds*

$$\mathscr{H}(\pi^{n,n}|\pi^\star_T) + \mathscr{H}(\pi^\star_T|\pi^{n,n}) \leq\; \frac{D(A,B,\mu)}{\hat\gamma^\mu_{n-1}} \prod_{k=0}^{n-1} \hat\gamma^\mu_k\,\hat\gamma^\nu_k \;,$$

$$\mathscr{H}(\pi^{n+1,n}|\pi^\star_T) + \mathscr{H}(\pi^\star_T|\pi^{n+1,n}) \leq D(A,B,\nu) \prod_{k=0}^{n-1} \hat\gamma^\mu_k\,\hat\gamma^\nu_k \;,$$

and

$$\mathscr{H}(\pi^{n,n}|\pi^{n,n-1}) + \mathscr{H}(\pi^{n,n-1}|\pi^{n,n}) = \mathscr{H}(\mu^n|\mu) + \mathscr{H}(\mu|\mu^n) \leq\; \frac{D(A,B,\mu)}{\hat\gamma^\mu_{n-1}} \prod_{k=0}^{n-1} \hat\gamma^\mu_k\,\hat\gamma^\nu_k \;,$$

$$\mathscr{H}(\pi^{n+1,n}|\pi^{n,n}) + \mathscr{H}(\pi^{n,n}|\pi^{n+1,n}) = \mathscr{H}(\nu^n|\nu) + \mathscr{H}(\nu|\nu^n) \leq D(A,B,\nu) \prod_{k=0}^{n-1} \hat\gamma^\mu_k\,\hat\gamma^\nu_k \;,$$

where the multiplicative constants are explicitly given at (67) and (68).

As a consequence, for T large enough (e.g. (21)), for any $\hat\gamma^\mu_\infty\,\hat\gamma^\nu_\infty < \lambda < 1$, there exists $C \geq 0$ such that for any $n \in \mathbb{N}^*$ it holds

$$\mathscr{H}(\pi^{n,n}|\pi^\star_T) + \mathscr{H}(\pi^\star_T|\pi^{n,n}) + \mathscr{H}(\pi^{n+1,n}|\pi^\star_T) + \mathscr{H}(\pi^\star_T|\pi^{n+1,n}) \leq\; C\,\lambda^n \;,$$

and

$$\mathscr{H}(\mu^n|\mu) + \mathscr{H}(\mu|\mu^n) + \mathscr{H}(\nu^n|\nu) + \mathscr{H}(\nu|\nu^n) \leq\; C\,\lambda^n \;.$$

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 (23).

In Remark 29 we show how the multiplicative constants $D(A,B,\cdot)$ can be further improved. The benefit of considering symmetric relative entropies in Theorem 12 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 11.We conclude the presentation of our main results by showing that the pointwise convergence of $(\nabla\varphi^n)_{n\in\mathbb{N}}$ 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, *i.e.*,

$$\|A\|_{\mathrm{F}} \coloneqq \sqrt{\mathrm{Tr}(AA^{\mathsf{T}})} \quad \forall\, A \in \mathbb{R}^{d\times d}\,.$$

The proof of convergence for the Hessians follows a similar scheme than the one for gradients replacing the representation formula for the derivative of Sinkhorn potentials with its second order counterpart that follows from the general relation

$$\nabla^2 \log P_T \exp(-h)(x) = -T^{-1}\,\mathrm{Id} + T^{-2}\,\mathrm{Cov}(\pi_T^{x,h})\,, \tag{28}$$

The above is proven as a part of Proposition 3.

Theorem 13. *Assume that* $\boldsymbol{A}$*1,*$\boldsymbol{A}$*2 and* (19) *hold. Then, for any* $n \geq 1$ *it holds*

$$\|\nabla^2\varphi^{n+1} - \nabla^2\varphi^\star\|_{\mathrm{F}}(x) \leq C(x,T,\nu,A,B)\prod_{k=0}^{n-1}\hat{\gamma}_k^\mu\,\hat{\gamma}_k^\nu\,,$$

$$\|\nabla^2\psi^{n+1} - \nabla^2\psi^\star\|_{\mathrm{F}}(x) \leq \frac{C(x,T,\mu,A,B)}{\hat{\gamma}_n^\mu}\prod_{k=0}^{n}\hat{\gamma}_k^\mu\,\hat{\gamma}_k^\nu\,,$$

where $C(x,T,\nu,A,B)$ *and* $C(x,T,\mu,A,B)$ *are given at* (82)*. As a consequence, for* T *large enough (e.g.* (21)*), for any* $\hat{\gamma}_\infty^\mu\,\hat{\gamma}_\infty^\nu < \lambda < 1$ *(independent of* $x\in\mathbb{R}^d$*), there exists* $C_x \geq 0$ *such that for any* $n \in \mathbb{N}^*$ *it holds*

$$\|\nabla^2\varphi^{n+1} - \nabla^2\varphi^\star\|_{\mathrm{F}}(x) \leq\ C_x\,\lambda^n \quad \text{ and } \quad \|\nabla^2\psi^{n+1} - \nabla^2\psi^\star\|_{\mathrm{F}}(x) \leq\ C_x\,\lambda^n\,.$$

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* (23)*.*

2.3. **Literature review.** Even though the what is nowadays called Sinkhorn's algorithm has been introduced over a century ago [74], its study has accelerated over the last years after the seminal work [26] made it a popular tool in statistical machine learning applications and beyond.

When marginal distributions are discrete probability measures, convergence of the algorithm is very well understood: we refer to the book [65] for an extensive overview. In this setting, where the algorithm is also known as the Iterative Proportional Fitting Procedure (IPFP), convergence has been proven for the first time by Sinkhorn in [71] and Sinkhorn and Knopp [72]. In [40, 11] IPFP iterations are shown to be equivalent to a sequence of iterations of a contraction in the Hilbert projective metrics. Thus, the analysis of the algorithm boils down to studying a fixed-point problem, and exponential convergence is deduced from Birkhoff's positive mapping Theorem [9]. The combined use of Birkhoff's Theorem and Hilbert metrics is also present in the works [15, 30]. In [15], the authors prove exponential convergence of the algorithm for continuous spaces and compact marginals. In [30] quantitative stability estimates for IPFP on compact metric spaces are established. When considering non compact spaces with possibly unbounded costs and marginals, the techniques developed to handle the discrete setting break down and new ideas have emerged. For multimarginal problems and bounded costs and marginals, (or equivalently compact spaces) Carlier and Laborde [13] have shown well-posedness of the Schrödinger system and smooth dependence of Schrödinger potentials on the marginal inputs. Di Marino and Gerolin [31] managed to show qualitative convergence of Sinkhorn iterates to Schrödinger potentials in L^p-norms with a calculus of variations approach: their results require bounded costs but apply to multimarginal

problems. These results have been subsequently improved by Carlier [12] who establishes exponential convergence rates. When it comes to results that allow for unbounded costs, Ruschendorf has shown qualitative convergences of iterates in relative entropy and total variation for Sinkhorn plans in [69]. Nutz and Wiesel have recently established [61] qualitative convergence both on the primal and dual sides under weak hypotheses. Concerning convergence rates, Léger [50] has given an interpretation of Sinkhorn's algorithm as a block coordinate descent on the dual problem that allows to obtain convegence of marginal distributions in relative entropy at speed n^{-1} under minimal assumptions. Nutz and Eckstein have shown [37] polynomial rates of convergence in Wasserstein assuming, among other things, that marginals admit exponential moments. Lastly, Nutz and Ghosal [44] obtain polynomial rates of convergence for optimal plans with respect to a symmetric relative entropy improving previously obtained results in what concerns the dependence of the estimates on the size of the regularization parameter.
The main contribution of this paper is to establish exponential convergence bounds for the gradients and Hessians of Sinkhorn 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 holds for unbounded costs and marginals. They also are among the very few results that yield convergence of derivatives of potentials. The other results about the convergence of gradients are those following from [29, Lemma 4.8] or those recently established by the authors and collaborators [47]. In both cases, boundedness of the cost is assumed. Our proof methods are probabilistic, and differ from other proposed methodologies in that they rely on one-sided integrated semiconvexity estimates for potentials along Sinkhorn iterates. These estimates are by themselves a new result, that has potentially several further implications. Though the current approach and the one we proposed in [47] are both inspired by coupling methods and stochastic control, there is a fundamental difference. In [47] exponential convergence is achieved through Lipschitz estimates on potentials. These are two-sided estimates and therefore we are still in a perturbative framework. 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. Lispchitz estimates for potentials cannot be expected and transferring integrated log-concavity from marginals into integrated semiconvexity of potentials is a delicate task, that requires a fine study of the properties of Hamilton-Jacobi-Bellman equations.

3. Integrated semiconvexity propagation along Sinkhorn

In this section we extend the results of [23] in order to study how lower-bounds for integrated convexity profile propagates along Sinkhorn iterations. Before proceeding further, let us point out here that in the results of this section it is enough assuming

A'2.

(i) There exist $\alpha_\nu \in (0,+\infty)$ *and* $\beta_\mu \in (0,+\infty]$ *such that*

$$\kappa_{U_\nu}(r) \geq \alpha_\nu - r^{-1}\,\tilde{g}_\nu(r) \quad \text{and} \quad \ell_{U_\mu}(r) \leq \beta_\mu + r^{-1}\,g_\mu(r)\ ,$$

with $g_\mu \in \mathcal{G}$ *and* $\tilde{g}_\nu \in \tilde{\mathcal{G}}$*;*

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

$$\kappa_{U_\mu}(r) \geq \alpha_\mu - r^{-1}\,\tilde{g}_\mu(r) \quad \text{and} \quad \ell_{U_\nu}(r) \leq \beta_\nu + r^{-1}\,g_\nu(r)\ ,$$

with $g_\nu \in \mathcal{G}$ *and* $\tilde{g}_\mu \in \tilde{\mathcal{G}}$*,* *where*

$$\tilde{\mathcal{G}} := \left\{ g \in \mathcal{G} \quad \textit{bounded and s.t.} \quad 2g'' + g\, g' \leq 0 \right\} \subseteq \mathcal{G}\ .$$

Notice that the above assumption allows for concave functions having non-null limit value in the origin (whereas the elements of $\mathcal{G}_\kappa$ are sublinear in a neighborhood of the origin), that $\mathcal{G}_\kappa \subseteq \tilde{\mathcal{G}}$ and hence that **A**2 implies **A'**2.

Let us introduce for any fixed $\beta > 0$ and any $g \in \mathcal{G}$ and $\tilde{g} \in \tilde{\mathcal{G}}$, the following functions for $\alpha,\ s,\ u \geq 0$

$$\begin{aligned} F^{g,\tilde{g}}_{\beta}(\alpha, s) &= \beta\, s + \frac{s}{T(1+T\alpha)} + s^{1/2}\, g(s^{1/2}) + \frac{s^{1/2}\, \tilde{g}(s^{1/2})}{(1+T\alpha)^2}\ , \\ G^{g,\tilde{g}}_{\beta}(\alpha, u) &= \inf\{ s \geq 0 :\ F^{g,\tilde{g}}_{\beta}(\alpha, s) \geq u\}\ , \end{aligned} \tag{29}$$

with the convention $G^{g,\tilde{g}}_{\beta}(\alpha, u) \equiv 0$ whenever $\beta = +\infty$.

Then, the main result of this section can be stated as follows.

Theorem 14. *Under* ***A*** *1 and* ***A'*** *2-(i), and if*

$$\kappa_{\psi^0}(r) \geq \alpha_\nu - T^{-1} - r^{-1}\tilde{g}_\nu(r) \tag{30}$$

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*

$$\ell_{\varphi^n}(r) \leq r^{-2}\, F^{g_\mu,\tilde{g}_\nu}_{\beta_\mu}(\alpha_{\nu,n}, r) - T^{-1} \quad \textit{and} \quad \kappa_{\psi^n}(r) \geq \alpha_{\nu,n} - r^{-1}\tilde{g}_\nu(r)\ , \tag{31}$$

with $F^{g^\ell_\mu, g^\kappa_\nu}_{\beta_\mu}$ *defined as in* (29)*. Moreover, the sequence can be explicitly built by setting*

$$\begin{cases} \alpha_{\nu,0} := \alpha_\nu - T^{-1}\ , \\ \alpha_{\nu,n+1} := \alpha_\nu - T^{-1} + \frac{G^{g_\mu,\tilde{g}_\nu}_{\beta_\mu}(\alpha_{\nu,n},2)}{2\,T^2}\ , \quad n \in \mathbb{N}\ , \end{cases} \tag{32}$$

$G^{g_\mu,\tilde{g}_\nu}_{\beta_\mu}$ *given in* (29)*. 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* (32) *and for any* $r > 0$*,*

$$\ell_{\varphi^\star}(r) \leq r^{-2}\, F^{g_\mu,\tilde{g}_\nu}_{\beta_\mu}(\alpha_{\psi^\star}, r) - T^{-1} \quad \textit{and} \quad \kappa_{\psi^\star}(r) > \alpha_{\psi^\star} - r^{-1}\tilde{g}_\nu(r)\ , \tag{33}$$

where $\varphi^\star$ *and* $\psi^\star$ *are the Schrödinger potentials introduced in* (1)*.*

Similarly, under ***A*** *1 and* ***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*

$$\ell_{\psi^n}(r) \leq r^{-2}\, F^{g_\nu,\tilde{g}_\mu}_{\beta_\nu}(\alpha_{\mu,n}, r) - T^{-1} \quad \textit{and} \quad \kappa_{\varphi^n}(r) \geq \alpha_{\mu,n} - r^{-1}\tilde{g}_\mu(r)\ , \tag{34}$$

with $F^{\beta_\nu, g^\ell_\nu, g^\kappa_\mu}_{T}$ *defined as in* (29) *and*

$$\begin{cases} \alpha_{\mu,1} := \alpha_\mu - T^{-1}\ , \\ \alpha_{\mu,n+1} := \alpha_\mu - T^{-1} + \frac{G^{g_\nu,\tilde{g}_\mu}_{\beta_\nu}(\alpha_{\mu,n},2)}{2\,T^2}\ , \quad n \in \mathbb{N}\ , \end{cases} \tag{35}$$

with $G^{g_\nu,\tilde{g}_\mu}_{\beta_\nu}$ *defined as in* (29)*. 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* (35) *and for any* $r > 0$*,*

$$\ell_{\psi^\star}(r) \leq r^{-2}\, F^{g_\nu,\tilde{g}_\mu}_{\beta_\nu}(\alpha_{\varphi^\star}, r) - T^{-1} \quad \textit{and} \quad \kappa_{\varphi^\star}(r) > \alpha_{\varphi^\star} - r^{-1}\tilde{g}_\mu(r)\ . \tag{36}$$

Remark 15. *The above result is an extension of* [23, Theorem 1.2] *where the author just provides the limit-bounds* (33) *and* (36)*. In the above result we show that the iterative proof given there can be actually employed when proving the estimates* (31) *and* (34) *along Sinkhorn algorithm.*

Let us also mention that our result encompasses [18, Theorem 4] *when considering* $\tilde g_\nu \equiv 0$ *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 just mention here that we will show in Remark 22 that Assumption (30) on the initial iterate 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 existence and uniqueness for the strong solution of the SDE

$$\mathrm{d}Y_t = -\Big(\frac{Y_t - x}{2T} + \frac{1}{2}\nabla h(Y_t)\Big)\mathrm{d}t + \mathrm{d}B_t\,. \tag{37}$$

for any choice of $h = \psi^\star,\ \psi^n,\ \varphi^\star,\ \varphi^n$. Indeed from Theorem 14 we immediately deduce

Corollary 16. *Under the assumptions of Theorem 14 it holds for any* $y \in \mathbb{R}^d$

$$\Big\langle \nabla\psi^\star(y),\, \frac{y}{|y|}\Big\rangle \geq \alpha_{\psi^\star}\,|y| - \tilde g_\nu(|y|) - |\nabla\psi^\star(0)|\,, \quad \Big\langle \nabla\psi^n(y),\, \frac{y}{|y|}\Big\rangle \geq \alpha^{\nu,n}\,|y| - \tilde g_\nu(|y|) - |\nabla\psi^n(0)|\,,$$
$$\Big\langle \nabla\varphi^\star(y),\, \frac{y}{|y|}\Big\rangle \geq \alpha_{\varphi^\star}\,|y| - \tilde g_\mu(|y|) - |\nabla\psi^\star(0)|\,, \quad \Big\langle \nabla\varphi^n(y),\, \frac{y}{|y|}\Big\rangle \geq \alpha^{\mu,n}\,|y| - \tilde g_\mu(|y|) - |\nabla\varphi^n(0)|\,.$$

As a consequence for any even $p \geq 2$ *the potential* $V_p(y) = 1 + |y|^p$ *is a Lyapunov function for* (37) *with* $h = \psi^\star,\ \psi^n,\ \varphi^\star,\ \varphi^n$ *and these SDEs admit existence and uniqueness of strong solutions.*

Proof. The lower-bounds displayed above are a direct consequence of Theorem 14. Let $x \in \mathbb{R}^d$ be fixed. We will only consider now the case $h = \psi^\star$. The other cases follow the same line. Since $\tilde g_\nu \in \tilde{\mathcal{G}}$ is bounded, it holds for any $y \in \mathbb{R}^d$

$$-\frac{1}{2}\,\big\langle T^{-1}(y-x) + \nabla\psi^\star(y),\, y\big\rangle \leq -\frac{\alpha_{\psi^\star} + T^{-1}}{2}\,|y|^2 + \frac{\|\tilde g_\nu\|_\infty + |\nabla h(0)| + T^{-1}|x|}{2}|y|\,,$$

and hence there exist $\gamma > 0$ and $R > 0$ such that

$$-\frac{1}{2}\,\big\langle T^{-1}(y-x) + \nabla\psi^\star(y),\, y\big\rangle \leq -\gamma\,|y|^2 \quad \forall\,|y| \geq R\,.$$

At this stage, [53, 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 (37). 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 $y \in \mathbb{R}^d$

$$\mathcal{L}_{\psi^\star} V_p(y) \leq -A_{\psi^\star} V_p(y) + B_{\psi^\star}\,, \tag{38}$$

where above $\mathcal{L}_{\psi^\star} \coloneqq \Delta/2 - \frac{1}{2}\langle T^{-1}(y-x) + \nabla\psi^\star(y),\, \nabla\rangle$ denotes the generator associated to the SDE (37). Finally, existence and uniqueness of strong solutions of (37) follows from [68, Theorem 2.1] (see also [55, Section 2]). □

The proof of Theorem 14 will be based on a propagation of integrated-convexity along Hamilton-Jacobi-Bellman (HJB) equations observed in [23], based on coupling by reflection techniques, which reads as follows

Theorem 17 (Theorem 2.1 in [23])**.** *For any fixed function $\tilde g \in \tilde{\mathcal{G}}$, consider the class of functions*

$$\mathcal{F}_{\tilde g} \coloneqq \{h \in \mathcal{C}^1(\mathbb{R}^d) :\ \kappa_h(r) \geq -r^{-1}\,\tilde g(r) \quad \forall\, r>0\}\,.$$

Then, the class $\mathcal{F}_{\tilde g}$ is stable under the action of the HJB flow, i.e.,

$$h \in \mathcal{F}_{\tilde g} \ \Rightarrow\ -\log P_{T-t}\exp(-h) \in \mathcal{F}_{\tilde g} \quad \forall\, 0 \leq t \leq T\,.$$

We omit the proof of the above result since it runs exactly as stated in [23]. There it is proven for the specific choice $\tilde g(r) = \tanh(r/2)$, however the proof works as well for any function $\tilde g \in \tilde{\mathcal{G}}$ since it only requires $\tilde g$ to satisfy the differential inequality

$$2(\tilde g)'' + \tilde g\,(\tilde g)' \leq 0\,,$$

which is an equality in the special case considered there $\tilde g(r) = \tanh(r)$.

As a first consequence of the previous theorem we may immediately deduce the following integrated propagation *convexity-to-concavity* result.

Lemma 18. *If **A'**2-(i) holds true, $\alpha_{\nu,n} > -T^{-1}$ and if for any $r>0$*

$$\kappa_{\psi^n}(r) \geq \alpha_{\nu,n} - r^{-1}\,\tilde g_\nu(r)\,, \tag{39}$$

then

$$\ell_{\varphi^{n+1}}(r) \leq \beta_\mu + g_\mu(r) - \frac{\alpha_{\nu,n}}{1+T\alpha_{\nu,n}} + \frac{r^{-1}\,\tilde g_\nu(r)}{(1+T\alpha_{\nu,n})^2} = -T^{-1} + r^{-2}\,F^{g_\mu,\tilde g_\nu}_{\beta_\mu}(\alpha_{\nu,n}, r^2)\,.$$

*Similarly if **A'**2-(ii) holds true, $\alpha_{\mu,n} > -T^{-1}$ and if for any $r>0$*

$$\kappa_{\varphi^n}(r) \geq \alpha_{\mu,n} - r^{-1}\,\tilde g_\mu(r)\,,$$

then

$$\ell_{\psi^n}(r) \leq \beta_\nu + r^{-1}\,g_\nu(r) - \frac{\alpha_{\mu,n}}{1+T\alpha_{\mu,n}} + \frac{r^{-1}\,\tilde g_\mu(r)}{(1+T\alpha_{\mu,n})^2} = -T^{-1} + r^{-2}\,F^{g_\nu,\tilde g_\mu}_{\beta_\nu}(\alpha_{\mu,n}, r^2)\,.$$

Proof. Let us firstly notice that our assumption (39) is equivalent to stating that

$$\hat\psi^n \coloneqq \psi^n - \frac{\alpha_{\nu,n}}{2}\,|\cdot|^2 \in \mathcal{F}_{\tilde g_\nu}\,,$$

and therefore Theorem 17 implies that

$$-\log P_T \exp(-\hat\psi^n) \in \mathcal{F}_{\tilde g_\nu}\,. \tag{40}$$

By recalling that φ^{n+1} is defined via (4), in order to conclude it is enough noticing now that

$$\begin{aligned}
-\log P_T\exp(-\psi^n)(x) - \frac{d}{2}\log(2\pi T) &= -\log\int \exp\Big(-\frac{|x-y|^2}{2T} - \frac{\alpha_{\nu,n}}{2}|y|^2 - \hat\psi^n(y)\Big)\,\mathrm{d}y\\
&= \frac{\alpha_{\nu,n}\,|x|^2}{2(1+T\alpha_{\nu,n})} - \log\int\exp\Big(-\frac{1+T\alpha_{\nu,n}}{2T}\,|y-(1+T\alpha_{\nu,n})^{-1}x|^2 - \hat\psi^n(y)\Big)\,\mathrm{d}y\\
&= \frac{\alpha_{\nu,n}\,|x|^2}{2(1+T\alpha_{\nu,n})} - \log P_{T/(1+T\alpha_{\nu,n})}\exp(-\hat\psi^n)((1+T\alpha_{\nu,n})^{-1}x) - \frac{d}{2}\log\frac{2\pi T}{1+T\alpha_{\nu,n}}\,,
\end{aligned}$$

and combining it with (40) and **A'**2-(i).

The second part of the statement, under **A'**2-(ii), can be proven similarly. □

Lemma 19. *Fix $\beta \in (0,+\infty]$ and two functions $g \in \mathcal{G}$ and $\tilde g \in \tilde{\mathcal{G}}$. Then the following properties hold true.*

(1) For any $\alpha > -T^{-1}$ the function $s \mapsto F_\beta^{g,\tilde g}(\alpha, s)$ is concave and increasing on $[0, +\infty)$.
(2) The function $\alpha \mapsto G_\beta^{g,\tilde g}(\alpha, 2)$ is positive and non-decreasing over $(-T^{-1}, +\infty)$.
(3) For any given $a_0 > 0$, the fixed-point problem

$$\alpha = a_0 - T^{-1} + \frac{G_\beta^{g,\tilde g}(\alpha, 2)}{2\,T^2} \tag{41}$$

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 fixed-points solutions.

Proof.

(1) Since $r \mapsto r\,\tilde g(r)$ and $r \mapsto r\,g(r)$ are non-decreasing and $\alpha > -T^{-1}$, an explicit computations shows that $s \mapsto F_\beta^{g,\tilde g}(\alpha, s)$ is an increasing function. The concavity of the latter function follows from the properties of $g,\, \tilde g \in \mathcal{G}$, since for $h = g,\, \tilde g$ 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 $G_\beta^{g,\tilde g}(\cdot, 2)$ is continuous and assume that is not a positive function, which implies that there exists some $\alpha > -T^{-1}$ such that $G_\beta^{g,\tilde g}(\alpha, 2) = 0$ and hence by definition that there exists a sequence $(s_n)_{n\in\mathbb{N}}$ converging to zero and such that $F_\beta^{g,\tilde g}(\alpha, s) \geq 2$, which is clearly impossible since $\lim_{s\downarrow 0} F_\beta^{g,\tilde g}(\alpha, s) = 0$. Hence $G_\beta^{g,\tilde g}(\cdot, 2)$ is a positive function. The monotonicity of $G_T^{\beta_\mu, g_\mu, \tilde g_\nu}(\cdot, 2)$ follows from the fact that $F_\beta^{g,\tilde g}(\alpha, s)$ is increasing in s and decreasing in $\alpha \in (-T^{-1}, +\infty)$, which implies for any $\alpha' \geq \alpha$ and $u \geq 0$

$$\{s :\ F_\beta^{g,\tilde g}(\alpha', s) \geq u\} \subseteq \{s :\ F_\beta^{g,\tilde g}(\alpha, s) \geq u\}\,.$$

(3) Consider the map associated to the fixed-point problem (32), *i.e.*, the continuous function $H \colon [a_0 - T^{-1}, +\infty) \to \mathbb{R}$ defined for $\alpha \in (-T^{-1}, +\infty)$ as

$$H(\alpha) \coloneqq \alpha - a_0 + T^{-1} - \frac{G_\beta^{g,\tilde g}(\alpha, 2)}{2\,T^2}\,.$$

Let us now prove that

$$H(a_0 - T^{-1}) < 0 \quad \text{and} \quad \lim_{\alpha\to+\infty} H(\alpha) = +\infty\,. \tag{42}$$

The first inequality follows from a direct computation, showing that $G_\beta^{g,\tilde g}(a_0 - T^{-1}, 2) > 0$. In order to prove the second statement it is enough noticing that $G_\beta^{g,\tilde g}(\cdot, 2)$ is bounded, which is immediate since $g,\, \tilde g \geq 0$ implies for any $\alpha > -T^{-1}$ and $s > 0$ that $F_\beta^{g,\tilde g}(\alpha, s) \geq \beta\, s$ and hence that

$$G_\beta^{g,\tilde g}(\alpha, 2) \leq 2/\beta\,. \tag{43}$$

From (42) and the continuity of $G_\beta^{g,\tilde g}(\cdot, 2)$ 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 (41). As a consequence (43) 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 $G_\beta^{g,\tilde g}(\cdot, 2)$ would have implied $H(a_0 - T^{-1}) = 0$, clearly in contrast with (42).

□

As a corollary of the previous lemma we have already proven the following

Corollary 20. *If* ***A'****2-(i) holds true, then there exists at least one solution* $\alpha_\nu^\star$ *on* $(\alpha_\nu - T^{-1}, \alpha_\nu - T^{-1} + (\beta_\mu T^2)^{-1}]$ *to the fixed point associated to (32). 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, then there exists at least one solution* $\alpha_\mu^\star$ *on* $(\alpha_\mu - T^{-1}, \alpha_\mu - T^{-1} + (\beta_\nu T^2)^{-1}]$ *to the fixed point associated to (35). 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 18 and it shows that we do also have an integrated propagation *concavity-to-convexity*.

Lemma 21. *If* ***A'****2-(i) holds true,* $\alpha_{\nu,n} > -T^{-1}$ *and if*

$$\ell_{\varphi^{n+1}}(r) \leq -T^{-1} + r^{-2}\, F^{g_\mu,\tilde{g}_\nu}_{\beta_\mu}(\alpha_{\nu,n}, r^2)\ , \tag{44}$$

then $\alpha_{\nu,n+1} > -T^{-1}$ *and for any* $r > 0$

$$\kappa_{\psi^{n+1}}(r) \geq \alpha_{\nu,n+1} - r^{-1}\, \tilde{g}_\nu(r)\ .$$

Similarly if ***A'****2-(ii) holds true,* $\alpha_{\mu,n} > -T^{-1}$ *and if*

$$\ell_{\psi^n}(r) \leq -T^{-1} + r^{-2}\, F^{g_\nu,\tilde{g}_\mu}_{\beta_\nu}(\alpha_{\mu,n}, r^2)\ ,$$

then $\alpha_{\mu,n+1} > -T^{-1}$ *and for any* $r > 0$

$$\kappa_{\varphi^{n+1}}(r) \geq \alpha_{\mu,n+1} - r^{-1}\, \tilde{g}_\mu(r)\ .$$

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

$$\hat{\psi}^{n+1}(y) \coloneqq T\psi^{n+1}(y) - TU_\nu(y) + \frac{|y|^2}{2}\ . \tag{45}$$

Then (28) implies that its Hessian is given for any $y \in \mathbb{R}^d$ by

$$\nabla^2 \hat{\psi}^{n+1}(y) = \frac{1}{T}\operatorname{Cov}(\pi_T^{y,\varphi^{n+1}})\ , \tag{46}$$

where we recall from the definition of $\pi_T^{y,\varphi^{n+1}}$ (11) 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 $V^{y,n+1} \coloneqq -\log(\mathrm{d}\pi_T^{y,\varphi^{n+1}}/\mathrm{dLeb})$, then our assumption implies

$$\ell_{V^{y,n+1}}(r) \leq r^{-2}\, F^{g_\mu,\tilde{g}_\nu}_{\beta_\mu}(\alpha_{\nu,n}, r^2) \quad \forall\, r > 0\ . \tag{47}$$

In order to prove the desired bound for $\kappa_{\psi^{n+1}}$, we will previously show a lower bound for the Hessian $\nabla^2\hat{\psi}^{n+1}$, *i.e.*, a lower bound for the covariance matrix (46). In view of that, let us consider for any fixed $y \in \mathbb{R}^d$, the variance $\operatorname{Var}_{X\sim\pi_T^{y,\varphi^{n+1}}}(X_1)$ where X_i denotes the i^{th} scalar component of the random vector $X \sim \pi_T^{y,\varphi^{n+1}}$. Next, observe that

$$\operatorname{Var}_{X\sim\pi_T^{y,\varphi^{n+1}}}(X_1) \geq \mathbb{E}_{X\sim\pi_T^{y,\varphi^{n+1}}}[\operatorname{Var}_{X\sim\pi_T^{y,\varphi^{n+1}}}(X_1|X_2,\,\dots,\,X_d)]\ , \tag{48}$$

and notice that for any given $z=(z_2,\ldots,z_d)\in\mathbb{R}^{d-1}$ it holds

$$\mathrm{Var}_{X\sim\pi_T^{y,\varphi^{n+1}}}(X_1|X_2=z_d,\ldots,X_d=z_d)=\frac{1}{2}\int_{\mathbb{R}^2}|x-\hat{x}|^2\,\pi_T^{y,\varphi^{n+1}}(\mathrm{d}x|z)\,\pi_T^{y,\varphi^{n+1}}(\mathrm{d}\hat{x}|z)$$

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 $V^{y,z}(x)\coloneqq V^{y,n+1}(x,z)$ we then have, uniformly in $z\in\mathbb{R}^{d-1}$,

$$\begin{aligned}
1=&\frac{1}{2}\int(\partial_x V^{y,z}(x)-\partial_x V^{y,z}(\hat{x}))(x-\hat{x})\,\pi_T^{y,\varphi^{n+1}}(\mathrm{d}x|z)\,\pi_T^{y,\varphi^{n+1}}(\mathrm{d}\hat{x}|z)\\
=&\frac{1}{2}\int\langle\nabla V^{y,n+1}(x,z)-\nabla V^{y,n+1}(\hat{x}),\ (x,z)-(\hat{x},z)\rangle\,\pi_T^{y,\varphi^{n+1}}(\mathrm{d}x|z)\,\pi_T^{y,\varphi^{n+1}}(\mathrm{d}\hat{x}|z)\\
\overset{(47)}{\le}&\frac{1}{2}\int F^{g_\mu,\tilde{g}_\nu}_{\beta_\mu}(\alpha_{\nu,n},|x-\hat{x}|^2)\,\pi_T^{y,\varphi^{n+1}}(\mathrm{d}x|z)\,\pi_T^{y,\varphi^{n+1}}(\mathrm{d}\hat{x}|z)\\
\le&F^{g_\mu,\tilde{g}_\nu}_{\beta_\mu}\left(\alpha_{\nu,n},2\,\mathrm{Var}_{X\sim\pi_T^{y,\varphi^{n+1}}}(X_1|X_2=z_d,\ldots,X_d=z_d)\right),
\end{aligned}$$

where the last step follows from the concavity of $s\mapsto F^{g_\mu,\tilde{g}_\nu}_{\beta_\mu}(\alpha_{\nu,n},s)$ (cf. Lemma 19) and Jensen's inequality. By combining the above with (48), since since $\alpha_{\nu,n}>-T^{-1}$ and $F^{g_\mu,\tilde{g}_\nu}_{\beta_\mu}(\alpha_{\nu,n},\cdot)$ is increasing (cf. Lemma 19), we deduce

$$\mathrm{Var}_{X\sim\pi_T^{y,\varphi^{n+1}}}(X_1)\ge\frac{1}{2}\,G^{g_\mu,\tilde{g}_\nu}_{\beta_\mu}(\alpha_{\nu,n},2)\,.$$

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 (44) too. The previous bound and this observation leads to

$$\mathrm{Var}_{X\sim\pi_T^{y,\varphi^{n+1}}}(\langle v,X\rangle)\ge\frac{1}{2}\,G^{g_\mu,\tilde{g}_\nu}_{\beta_\mu}(\alpha_{\nu,n},2)\quad\forall\,y,\,v\in\mathbb{R}^d\text{ s.t. }|v|=1\,,$$

and hence from (46) we finally deduce

$$\langle v,\,\nabla^2\hat{\psi}^{n+1}(y)\,v\rangle\ge G^{g_\mu,\tilde{g}_\nu}_{\beta_\mu}(\alpha_{\nu,n},2)\,\frac{|v|^2}{2T}\quad\forall\,y,\,v\in\mathbb{R}^d\,.$$

By recalling (45) and (32), the above bound concludes our proof. □

Remark 22 (A first trivial lower bound)**.** *Under **A'**2, the above discussion already provides a first trivial lower bound for κ_{ψ^n}. Indeed (46) tells us that $\hat{\psi}^{n+1}$ is a convex function for any $n\ge0$, which combined with (45) yields to*

$$\kappa_{\psi^{n+1}}(r)\ge\alpha_\nu-T^{-1}-r^{-1}\tilde{g}_\nu(r)\quad\forall\,n\ge0\,.$$

Therefore in Theorem 14 we could consider any initialization ψ^0 without prescriptions on its behaviour and run an iteration of Sinkhorn in order to get $\alpha_{\nu,1}\ge\alpha_\nu-T^{-1}$. At this point once can proceed again with the proof of Theorem 14 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

$$\kappa_{\varphi^{n+1}}\ge\alpha_\mu-T^{-1}-r^{-1}\tilde{g}_\mu(r)\quad\forall\,n\ge0\,,$$

which for $n=0$ gives for free the base step of the iteration (35).

Finally, let us remark here that since the potentials couple $(\varphi^\star, \psi^\star)$ *can be thought as a constant sequence of Sinkhorn iterates, the above discussion proves also that*

$$\kappa_{\psi^\star}(r) \geq \alpha_\nu - T^{-1} - r^{-1}\tilde{g}_\nu(r) \quad \textit{and} \quad \kappa_{\varphi^\star} \geq \alpha_\mu - T^{-1} - r^{-1}\tilde{g}_\mu(r)\ .$$

Given the above lemmata, we are finally able to prove how lower-bounds for integrated convexity profiles propagate along Sinkhorn.

Proof of Theorem 14. Let us start showing the first statement Hence let us consider the sequence $(\alpha_{\nu,n})_{n\in\mathbb{N}}$ defined in (32). We will prove our statement by induction. The case $n=0$ is met under the assumption $\kappa_{\psi^0}(r) \geq \alpha_\nu - T^1 - r^{-1}\tilde{g}_\nu(r)$ The inductive step follows by applying consecutively Lemma 21 and Lemma 18. As a direct consequence of item *(ii)* in Lemma 19 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 $G^{g_\mu,\tilde{g}_\nu}_{\beta_\mu}$ is continuous and $\alpha_\nu - T^{-1}$ is not an accumulation point for the set of solutions of (41) (cf. item *(iii)* in Lemma 19), 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 $\alpha^\star_\psi$. The upper bound on $\alpha_{\nu,n}$ comes for free from (32) and the upper bound (43). The proof of (33) is obtained in the same way by considering the (constant) Sinkhorn iterates $(\varphi^\star, \psi^\star)$ with the same sequence of $(\alpha_{\nu,n})_{n\in\mathbb{N}}$.

The proof second half of the 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 22. □

4. Wasserstein distance w.r.t a measure with 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* $g^\kappa_\mathfrak{p} \in \mathcal{G}_\kappa$ *such that*

$$\kappa_{U_\mathfrak{p}}(r) \geq \alpha_\mathfrak{p} - r^{-1}\,g^\kappa_\mathfrak{p}(r) \quad \forall r > 0\ .$$

Notice that the properties prescribed in the definition of $\mathcal{G}_\kappa$ in particular imply that its elements are sublinear (*i.e.*, $\sup_{r>0} g(r)/r < +\infty$) concave functions on $(0,+\infty)$. Indeed, any $g \in \mathcal{G}_\kappa$ 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 $g \in \mathcal{G}_\kappa$ follows.

Let us also emphasize here that the convexity of integrated profile assumption is stronger than the coercivity, since the former implies

$$-\frac{1}{2}\langle \nabla U_\mathfrak{p}(x), x\rangle \leq -\frac{\alpha_\mathfrak{p}}{2}\,|x|^2 + \frac{|\nabla U_\mathfrak{p}(0)| + \|g^\kappa_\mathfrak{p}\|_\infty}{2}|x|\ ,$$

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

$$\begin{cases} \mathrm{d}X_t = -\frac{1}{2}\nabla U_{\mathfrak{p}}(X_t)\,\mathrm{d}t + \mathrm{d}B_t\ , \\ \mathrm{d}Y_t = -\frac{1}{2}\nabla U_{\mathfrak{q}}(Y_t)\,\mathrm{d}t + \mathrm{d}B_t\ , \end{cases} \tag{49}$$

which admit unique strong solutions, in view of the coercivity of the corresponding drifts and owing to [68, 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 (49).

Theorem 23. *Assume **AO1**. Then it holds*

$$\mathcal{W}_1(\mathfrak{p},\ \mathfrak{q}) \leq (2\ C_{\mathfrak{p}}\ \lambda_{\mathfrak{p}})^{-1} \int |\nabla U_{\mathfrak{p}} - \nabla U_{\mathfrak{q}}|\mathrm{d}\mathfrak{q} = (2\ C_{\mathfrak{p}}\ \lambda_{\mathfrak{p}})^{-1} \int \left|\nabla \log \frac{\mathrm{d}\mathfrak{p}}{\mathrm{d}\mathfrak{q}}\right| \mathrm{d}\mathfrak{q}\ ,$$

where $\lambda_{\mathfrak{p}} > 0$ and $C_{\mathfrak{p}} > 0$ are respectively the ergodicity exponential rate and the equivalence constant associated to the Langevin SDE driven by $-\nabla U_{\mathfrak{p}}$, obtained via the coupling by reflection construction portrayed in Appendix C.

Proof. In order to bound the above Wasserstein distance we will consider the Wasserstein distance (cf. Corollary 34)

$$\mathcal{W}_{f_{\mathfrak{p}}}(\mathfrak{p},\ \mathfrak{q}) := \inf_{\pi\in\Pi(\mathfrak{p},\mathfrak{q})} \mathbb{E}_{(X,Y)\sim\pi}\left[f_{\mathfrak{p}}(|X-Y|)\right]. \tag{50}$$

induced by the concave function $f_{\mathfrak{p}}$ built in Proposition 33, associated to $\kappa_{\mathfrak{p}}(r) := \alpha_{\mathfrak{p}} - r^{-1} g_{\mathfrak{p}}^{\kappa}(r)$. Let us also introduce the rate $\lambda_{\mathfrak{p}} > 0$ and the constant $C_{\mathfrak{p}} > 0$ therein built, associated to such $f_{\mathfrak{p}}$. Recall that it holds the following equivalence of Wasserstein distances

$$C_{\mathfrak{p}}\ \mathcal{W}_1(\mathfrak{p},\ \mathfrak{q}) \leq \mathcal{W}_{f_{\mathfrak{p}}}(\mathfrak{p},\ \mathfrak{q}) \leq \mathcal{W}_1(\mathfrak{p},\ \mathfrak{q})\ , \tag{51}$$

and the differential inequality

$$2\ f_{\mathfrak{p}}{}''(r) - \frac{r\, f_{\mathfrak{p}}'(r)}{2}\,\kappa_{\mathfrak{p}}(r) \leq -\lambda_{\mathfrak{p}}\ f_{\mathfrak{p}}(r) \quad \forall\, r \in [0,+\infty)\ . \tag{52}$$

We will bound (50) via the triangular inequality by considering for any $t \geq 0$

$$\mathcal{W}_{f_{\mathfrak{p}}}(\mathfrak{p},\ \mathfrak{q}) \leq \mathcal{W}_{f_{\mathfrak{p}}}(\mathfrak{p},\ \mathfrak{q}\,P_t^{\mathfrak{p}}) + \mathcal{W}_{f_{\mathfrak{p}}}(\mathfrak{q}\,P_t^{\mathfrak{p}},\ \mathfrak{q}) = \mathcal{W}_{f_{\mathfrak{p}}}(\mathfrak{p}\,P_t^{\mathfrak{p}},\ \mathfrak{q}\,P_t^{\mathfrak{p}}) + \mathcal{W}_{f_{\mathfrak{p}}}(\mathfrak{q}\,P_t^{\mathfrak{p}},\ \mathfrak{q})\ . \tag{53}$$

In order to bound the first Wasserstein distance appearing in the upper bound (53), namely $\mathcal{W}_{f_{\mathfrak{p}}}(\mathfrak{p}\,P_t^{\mathfrak{p}},\ \mathfrak{q}\,P_t^{\mathfrak{p}})$, fix an initial coupling $(X_0, X_0^{\mathfrak{q}}) \sim \pi^\star$ distributed according to the optimal coupling for $\mathcal{W}_{f_{\mathfrak{p}}}(\mathfrak{p},\ \mathfrak{q})$ 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

$$\mathrm{d}\hat{B}_t := (\mathrm{Id} - 2\, e_t\, e_t^{\mathsf{T}}\, \mathbf{1}_{\{t<\tau\}})\,\mathrm{d}B_t \qquad \text{where} \quad e_t := \begin{cases} \frac{Z_t}{|Z_t|} & \text{when } r_t > 0\,, \\ u & \text{when } r_t = 0\,. \end{cases}$$

where $Z_t \coloneqq X_t - X_t^{\mathfrak{q}}$, $r_t \coloneqq |Z_t|$ and $u \in \mathbb{R}^d$ is a fixed (arbitrary) unit-vector. By Lévy's characterization, $(\hat{B}_t)_{t\geq 0}$ 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 $\mathrm{d}W_t \coloneqq e_t^{\intercal}\mathrm{d}B_t$ 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 Lemma to the concave function $f_{\mathfrak{p}}$ gives

$$\begin{aligned} \mathrm{d}f_{\mathfrak{p}}(r_t) &= -\frac{f_{\mathfrak{p}}^{'}(r_t)}{2}\,\langle e_t,\; \nabla U_{\mathfrak{p}}(X_t) - \nabla U_{\mathfrak{p}}(X_t^{\mathfrak{q}})\rangle\,\mathrm{d}t + 2\, f_{\mathfrak{p}}^{''}(r_t)\,\mathrm{d}t + 2\, f_{\mathfrak{p}}^{'}(r_t)\,\mathrm{d}W_t \\ &\leq \left(2\, f_{\mathfrak{p}}^{''}(r_t) - \frac{r_t\, f_{\mathfrak{p}}^{'}(r_t)}{2}\kappa_{\mathfrak{p}}(r_t)\right)\mathrm{d}t + 2\, f_{\mathfrak{p}}^{'}(r_t)\,\mathrm{d}W_t \overset{(52)}{\leq} -\lambda_{\mathfrak{p}}\, f_{\mathfrak{p}}(r_t)\,\mathrm{d}t + 2\, f_{\mathfrak{p}}^{'}(r_t)\,\mathrm{d}W_t \;. \end{aligned}$$

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 Gronwall Lemma we have finally proven that for any $t \geq 0$

$$\mathcal{W}_{f_{\mathfrak{p}}}(\mathfrak{p}\, P_t^{\mathfrak{p}},\, \mathfrak{q}\, P_t^{\mathfrak{p}}) \leq \mathbb{E}[f_{\mathfrak{p}}(|X_t - X_t^{\mathfrak{q}}|)] \leq e^{-\lambda_{\mathfrak{p}}\, t}\, \mathbb{E}[f_{\mathfrak{p}}(|X_0 - X_0^{\mathfrak{q}}|)] = e^{-\lambda_{\mathfrak{p}}\, t}\, \mathcal{W}_{f_{\mathfrak{p}}}(\mathfrak{p},\, \mathfrak{q})\,. \tag{54}$$

Let us now provide a bound for the second Wasserstein distance appearing in the upper bound (53), namely $\mathcal{W}_{f_{\mathfrak{p}}}(\mathfrak{q}\, P_t^{\mathfrak{p}},\, \mathfrak{q})$, by relying on 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 $\bar{r}_t \coloneqq |X_t^{\mathfrak{q}} - Y_t|$ 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 $\rho_\delta(r) \coloneqq \sqrt{r+\delta}$. Then it holds

$$\begin{aligned} \mathrm{d}\rho_\delta(\bar{r}_t^2) &= -\,(2\,\rho_\delta(\bar{r}_t^2))^{-1}\langle X_t^{\mathfrak{q}} - Y_t,\; \nabla U_{\mathfrak{p}}(X_t^{\mathfrak{q}}) - \nabla U_{\mathfrak{q}}(Y_t)\rangle\,\mathrm{d}t \\ &= -\,(2\,\rho_\delta(\bar{r}_t^2))^{-1}\langle X_t^{\mathfrak{q}} - Y_t,\; \nabla U_{\mathfrak{p}}(X_t^{\mathfrak{q}}) - \nabla U_{\mathfrak{p}}(Y_t)\rangle\,\mathrm{d}t \\ &\qquad\qquad - (2\,\rho_\delta(\bar{r}_t^2))^{-1}\langle X_t^{\mathfrak{q}} - Y_t,\; \nabla U_{\mathfrak{p}}(Y_t) - \nabla U_{\mathfrak{q}}(Y_t)\rangle\,\mathrm{d}t \\ &\leq -\, 2^{-1}\,\frac{\bar{r}_t^2}{\rho_\delta(\bar{r}_t^2)}\,\kappa_{\mathfrak{p}}(\bar{r}_t)\,\mathrm{d}t + 2^{-1}\,\frac{\bar{r}_t}{\rho_\delta(\bar{r}_t^2)}\,|\nabla U_{\mathfrak{p}} - \nabla U_{\mathfrak{q}}|(Y_t)\mathrm{d}t \\ &\leq -\, 2^{-1}\,\frac{\bar{r}_t^2}{\rho_\delta(\bar{r}_t^2)}\,(\alpha_{\mathfrak{p}} - G_{\mathfrak{p}}^{\kappa})\,\mathrm{d}t + 2^{-1}\,\frac{\bar{r}_t}{\rho_\delta(\bar{r}_t^2)}\,|\nabla U_{\mathfrak{p}} - \nabla U_{\mathfrak{q}}|(Y_t)\mathrm{d}t\;, \end{aligned}$$

where in the last step we have relied on the sublinearity of $g^\kappa_\mathfrak{p}\in\mathcal{G}_\kappa$ (namely that $g^\kappa_\mathfrak{p}(r)\leq G^\kappa_\mathfrak{p}\,r$ for some positive constant $G^\kappa_\mathfrak{p}>0$). Therefore it holds

$$\mathrm{d}\rho_\delta(\bar r_t^2)\leq \frac{(\alpha_\mathfrak{p}-G^\kappa_\mathfrak{p})^+}{2}\,\rho_\delta(\bar r_t^2)\mathrm{d}t\ +2^{-1}|\nabla U_\mathfrak{p}-\nabla U_\mathfrak{q}|(Y_t)\mathrm{d}t\,.$$

By taking expectation and integrating over time the above bound gives

$$\begin{aligned}\mathbb{E}[\rho_\delta(\bar r_t^2)]&\leq \mathbb{E}[\rho_\delta(\bar r_0^2)]+\frac{(\alpha_\mathfrak{p}-G^\kappa_\mathfrak{p})^+}{2}\int_0^t\mathbb{E}[\rho_\delta(\bar r_s^2)]\,\mathrm{d}s+2^{-1}\int_0^t\mathbb{E}[|\nabla U_\mathfrak{p}-\nabla U_\mathfrak{q}|(Y_t)]\,\mathrm{d}s\\&=\sqrt{\delta}+\frac{(\alpha_\mathfrak{p}-G^\kappa_\mathfrak{p})^+}{2}\int_0^t\mathbb{E}[\rho_\delta(\bar r_s^2)]\,\mathrm{d}s+\frac{t}{2}\int|\nabla U_\mathfrak{p}-\nabla U_\mathfrak{q}|\,\mathrm{d}\mathfrak{q}\,,\end{aligned}$$

where in the last step we have relied on the stationarity $Y_t\sim\mathfrak{q}$ and that $\bar r_0=0$. Therefore Gronwall Lemma yields to

$$\mathbb{E}[\bar r_t]\leq\mathbb{E}[\rho_\delta(\bar r_t^2)]\leq\frac{t}{2}\,\exp\biggl(\frac{t}{2}\,(\alpha_\mathfrak{p}-G^\kappa_\mathfrak{p})^+\biggr)\int|\nabla U_\mathfrak{p}-\nabla U_\mathfrak{q}|\,\mathrm{d}\mathfrak{q}+\sqrt{\delta}\,\exp\biggl(\frac{t}{2}\,(\alpha_\mathfrak{p}-G^\kappa_\mathfrak{p})^+\biggr)\,.$$

By letting δ to zero in the above right-hand-side, we obtain the desired upper bound

$$\begin{aligned}\mathcal{W}_{f_\mathfrak{p}}(\mathfrak{q}\,P^\mathfrak{p}_t,\,\mathfrak{q})&\overset{(51)}{\leq}\mathcal{W}_1(\mathfrak{q}\,P^\mathfrak{p}_t,\,\mathfrak{q})\leq\mathbb{E}[|X^\mathfrak{q}_t-Y_t|]=\mathbb{E}[\bar r_t]\\&\leq\frac{t}{2}\,\exp\biggl(\frac{t}{2}\,(\alpha_\mathfrak{p}-G^\kappa_\mathfrak{p})^+\biggr)\int|\nabla U_\mathfrak{p}-\nabla U_\mathfrak{q}|\,\mathrm{d}\mathfrak{q}\,.\end{aligned}$$

By putting together the last estimate with (53) and (54) we have proven

$$\mathcal{W}_{f_\mathfrak{p}}(\mathfrak{p},\,\mathfrak{q})\leq e^{-\lambda_\mathfrak{p}\,t}\,\mathcal{W}_{f_\mathfrak{p}}(\mathfrak{p},\,\mathfrak{q})+\frac{t}{2}\,\exp\biggl(\frac{t}{2}\,(\alpha_\mathfrak{p}-G^\kappa_\mathfrak{p})^+\biggr)\int|\nabla U_\mathfrak{p}-\nabla U_\mathfrak{q}|\,\mathrm{d}\mathfrak{q}\,.$$

or equivalently

$$\mathcal{W}_{f_\mathfrak{p}}(\mathfrak{p},\,\mathfrak{q})\leq\frac{t/2}{1-e^{-\lambda_\mathfrak{p}\,t}}\,\exp\biggl(\frac{t}{2}\,(\alpha_\mathfrak{p}-G^\kappa_\mathfrak{p})^+\biggr)\int|\nabla U_\mathfrak{p}-\nabla U_\mathfrak{q}|\,\mathrm{d}\mathfrak{q}\,,$$

which in the t vanishing limit reads as

$$\mathcal{W}_{f_\mathfrak{p}}(\mathfrak{p},\,\mathfrak{q})\leq(2\,\lambda_\mathfrak{p})^{-1}\int|\nabla U_\mathfrak{p}-\nabla U_\mathfrak{q}|\,\mathrm{d}\mathfrak{q}\,.$$

Combining the above bound with the equivalence (51) concludes the proof. □

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 24. *Assume the validity of* **A** *1 and* **A** *2. Then, for any $x\in\mathbb{R}^d$, it holds*

$$\mathcal{W}_1(\pi_T^{x,\psi^n},\pi_T^{x,\psi^\star})\leq\gamma_n^\nu\int|\nabla\psi^n-\nabla\psi^\star|\,\mathrm{d}\pi_T^{x,\psi^\star},\tag{55}$$

and similarly

$$\mathcal{W}_1(\pi_T^{x,\varphi^{n+1}},\pi_T^{x,\varphi^\star})\leq\gamma_n^\mu\int\bigl|\nabla\varphi^{n+1}-\nabla\varphi^\star\bigr|\,\mathrm{d}\pi_T^{x,\varphi^\star},\tag{56}$$

where γ_n^μ and γ_n^ν are given in Remark 35.

Proof. Inequality (55) 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 (49) 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 14 guarantees that there exists $\alpha_{\nu,n}$, $\alpha_{\psi^\star} > -T^{-1}$ such that

$$\kappa_{U_\mathfrak{p}}(r) \geq T^{-1} + \alpha_{\nu,n} - r^{-1} g_\nu^\kappa(r) \quad \text{and} \quad \kappa_{U_\mathfrak{q}}(r) \geq T^{-1} + \alpha_{\psi^\star} - r^{-1} g_\nu^\kappa(r) .$$

Therefore $\mathfrak{p}$ and $\mathfrak{q}$ satisfy Assumption **AO**1 and Theorem 23 gives (55).

We omit the details for the proof of (56) since it can be obtained in the same way, this time considering $\mathfrak{p} = \pi_T^{x,\varphi^{n+1}}$ and $\mathfrak{q} = \pi_T^{x,\varphi^\star}$. □

5. Proof of the main results

Owing to the results of the previous section and to the construction given in Remark 35 in the appendix, let us simply remark here that the rate sequences $(\gamma_n^\mu)_{n\in\mathbb{N}}$ and $(\gamma_n^\nu)_{n\in\mathbb{N}}$ appearing in the following results are monotone decreasing, *i.e.*, they provide faster convergence as the index n increases. It will be clear from our exposition that we will deduce from the convergence of the gradients, both the convergence of the iterates on primal and dual formulation as well as the convergence of the Hessians.

5.1. Exponential convergence of the gradients along Sinkhorn algorithm.

The convergence of the gradients will follow by iterating the following contracting result.

Proposition 25. *Assume the validity of* ***A****1 and* ***A****2. Then for any $n \geq 0$ it holds*

$$\int |\nabla\varphi^{n+1} - \nabla\varphi^\star| \mathrm{d}\mu \leq T^{-1}\gamma_n^\nu \int |\nabla\psi^n - \nabla\psi^\star| \mathrm{d}\nu , \tag{57}$$

and similarly

$$\int |\nabla\psi^{n+1} - \nabla\psi^\star| \mathrm{d}\nu \leq T^{-1}\gamma_n^\mu \int |\nabla\varphi^{n+1} - \nabla\varphi^\star| \mathrm{d}\mu , \tag{58}$$

where γ_n^μ and γ_n^ν are given in Remark 35.

Proof. Let us start by showing (57). From (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 3)

$$\nabla \log P_T \exp(-h)(x) = \frac{1}{T}\int (y-x)\, \pi_T^{x,h}(\mathrm{d}y) ,$$

we finally deduce

$$|\nabla\varphi^{n+1} - \nabla\varphi^\star|(x) = T^{-1}\left| \int y\, \pi_T^{x,\psi^n}(\mathrm{d}y) - \int y\, \pi_T^{x,\psi^\star}(\mathrm{d}y) \right| \leq T^{-1}\mathcal{W}_1(\pi_T^{x,\psi^n}, \pi_T^{x,\psi^\star}) . \tag{59}$$

Combining the above observation with Corollary 24 we end up with

$$|\nabla\varphi^{n+1} - \nabla\varphi^\star|(x) \leq \; T^{-1}\, \gamma_n^\nu \int |\nabla\psi^n - \nabla\psi^\star| \,\mathrm{d}\pi_T^{x,\psi^\star} \;, \tag{60}$$

with γ_n^ν as introduced in Remark 35. In order to conclude it is enough noticing that the $\pi_T^{x,\psi^\star}$ is the conditional probability of the Schrödinger bridge $\pi_T^\star$ given as first variable x, which means that by integrating over $\mu(\mathrm{d}x)$ we have[2]

$$\int \pi_T^{x,\psi^\star}(\mathrm{d}y)\,\mu(\mathrm{d}x) = \nu(\mathrm{d}y)\ ,$$

which combined with (60) gives (57). The contractive estimate (58) can be proven in the same fashion by relying on (56) by noticing that

$$|\nabla\psi^{n+1} - \nabla\psi^\star|(x) \le T^{-1}\mathcal{W}_1(\pi_T^{x,\varphi^{n+1}}, \pi_T^{x,\varphi^\star}) \ \text{ and } \int \pi_T^{x,\varphi^\star}(\mathrm{d}y)\,\nu(\mathrm{d}x) = \mu(\mathrm{d}y)\ .$$

□

We can now move to the pointwise convergence get the pointwise convergence of the gradients stated in Theorem 7, by relying on contractive techniques similar to the previous ones.

Lemma 26. *Assume the validity of* $\boldsymbol{A}$*1 and* $\boldsymbol{A}$*2-(i). If there are positive constants $A, B > 0$ such that*

$$|\nabla\psi^n - \nabla\psi^\star|(x) \le (A|x| + B) \quad \forall\, x \in \mathbb{R}^d\ ,$$

then it holds

$$|\nabla\varphi^{n+1} - \nabla\varphi^\star|(x) \le \hat\gamma_n^\nu\,(A|x| + B) \quad \forall\, x \in \mathbb{R}^d$$

with

$$\hat\gamma_n^\nu \coloneqq T^{-1}\,\gamma_n^\nu\ \max\Big\{(T\,\alpha_{\psi^\star} + 1)^{-1}\ , \Big(1 + \frac{A}{B}\frac{1 + \|g_\nu^\kappa\|_\infty + |\nabla\psi^\star(0)|}{\alpha_{\psi^\star} + T^{-1}}\Big)\Big\}\ ,$$

with γ_n^ν as defined in Remark 35.

Similarly if we assume $\boldsymbol{A}$*1 and* $\boldsymbol{A}$*2-(ii) and that*

$$|\nabla\varphi^n - \nabla\varphi^\star|(x) \le (A|x| + B) \quad \forall\, x \in \mathbb{R}^d\ ,$$

then it holds

$$|\nabla\psi^{n+1} - \nabla\psi^\star|(x) \le \hat\gamma_n^\mu\,(A|x| + B) \quad \forall\, x \in \mathbb{R}^d$$

with

$$\hat\gamma_n^\mu \coloneqq T^{-1}\,\gamma_n^\mu\ \max\Big\{(T\,\alpha_{\varphi^\star} + 1)^{-1}\ , \Big(1 + \frac{A}{B}\frac{1 + \|g_\mu^\kappa\|_\infty + |\nabla\varphi^\star(0)|}{\alpha_{\varphi^\star} + T^{-1}}\Big)\Big\}\ ,$$

with γ_n^μ as defined in Remark 35.

Proof. We will prove only the first contracting bound since the proof of the second one can be achieved by following the same argument. Once again let us consider λ_n^ν, C_n^ν and f_n^ν as in Remark 35, associated to $\kappa_n^\nu(r) = \alpha_{\nu,n} + T^{-1} - r^{-1}\,g_\nu^\kappa(r)$. Owing to the computations performed in Proposition 25 and Corollary 24, let us consider (60) as starting point of our proof here. Therefore we have

$$|\nabla\varphi^{n+1} - \nabla\varphi^\star|(x) \le\ T^{-1}\,\gamma_n^\nu \int |\nabla\psi^n - \nabla\psi^\star|\,\mathrm{d}\pi_T^{x,\psi^\star}\ ,$$

which combined with our assumption yields to

$$|\nabla\varphi^{n+1} - \nabla\varphi^\star|(x) \le T^{-1}\,\gamma_n^\nu\Big(A\,\mathbb{E}_{\pi_T^{x,\psi^\star}}[|Y|] + B\Big)\ . \tag{61}$$

[2]This equality can also be proven by writing explicitly $\pi_T^{x,\psi^\star}(\mathrm{d}y) = (P_T e^{-\psi^\star}(x))^{-1}\,\exp(-\frac{|y-x|^2}{2T} - \psi^\star(y))\mathrm{d}y$ and $\mu(\mathrm{d}x) = \exp(-\varphi^\star(x) + \log P_T e^{-\psi^\star}(x))\mathrm{d}x$ as follows from (4).

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

$$\begin{cases} \mathrm{d}Y^\star_t = -\Big(\frac{Y^\star_t - x}{2T} + \frac{1}{2}\nabla\psi^\star(Y^\star_t)\Big)\,\mathrm{d}t + \mathrm{d}B_t \\ Y^\star_0 \sim \pi_T^{x,\psi^\star}\,. \end{cases}$$

Then Ito formula, Corollary 16 and the boundedness of $g^\kappa_\nu \in \mathcal{G}_\kappa$ imply that

$$\begin{aligned} \mathrm{d}|Y^\star_t|^2 &= -T^{-1}|Y^\star_t|^2\,\mathrm{d}t + T^{-1}\langle Y^\star_t,\, x\rangle\,\mathrm{d}t - \langle Y^\star_t,\, \nabla\psi^\star(Y^\star_t)\rangle\,\mathrm{d}t + 1\,\mathrm{d}t + 2\,Y^\star_t\,\cdot \mathrm{d}B_t \\ &\le -(\alpha_{\psi^\star} + T^{-1})|Y^\star_t|^2\,\mathrm{d}t + (1 + T^{-1}|x| + \|g^\kappa_\nu\|_\infty + |\nabla\psi^\star(0)|)|Y^\star_t|\,\mathrm{d}t + 2\,Y^\star_t\,\cdot \mathrm{d}B_t\,, \end{aligned}$$

and therefore for any $\varepsilon \in (0, \alpha_{\psi^\star} + T^{-1})$ we have

$$\mathrm{d}|Y^\star_t|^2 \le -(\alpha_{\psi^\star} + T^{-1} - \varepsilon)|Y^\star_t|^2\,\mathrm{d}t + (4\varepsilon)^{-1}(1 + T^{-1}|x| + \|g^\kappa_\nu\|_\infty + |\nabla\psi^\star(0)|)^2\mathrm{d}t + 2\,Y^\star_t\cdot\mathrm{d}B_t\,.$$

If we consider the stopping time $\tau_M \coloneqq \inf\{t \ge 0 \,:\, |Y^\star_t| > M\}$, where we set $\inf(\emptyset) \coloneqq +\infty$, by integrating over time on $[0, t\wedge\tau_M]$, taking expectation and owing to the Optional Stopping Theorem we deduce that

$$\begin{aligned} \mathbb{E}[|Y^\star_{t\wedge\tau_M}|^2] \le \mathbb{E}[|Y^\star_0|^2] - (\alpha_{\psi^\star} + T^{-1} - \varepsilon)\int_0^t \mathbb{E}[\mathbf{1}_{s\le\tau_M}\,|Y^\star_s|^2]\,\mathrm{d}s \\ + \frac{t}{4\varepsilon}(1 + T^{-1}|x| + \|g^\kappa_\nu\|_\infty + |\nabla\psi^\star(0)|)^2\,. \end{aligned}$$

Since $\tau_M \uparrow +\infty$ almost surely as $M \uparrow +\infty$ (as a consequence of Corollary 16), in the asymptotic regime Fatou Lemma implies for any $t \ge 0$ that

$$\mathbb{E}[|Y^\star_t|^2] + (\alpha_{\psi^\star} + T^{-1} - \varepsilon)\int_0^t \mathbb{E}[|Y^\star_s|^2]\,\mathrm{d}s \le \mathbb{E}[|Y^\star_0|^2] + \frac{t}{4\varepsilon}(1 + T^{-1}|x| + \|g^\kappa_\nu\|_\infty + |\nabla\psi^\star(0)|)^2\,,$$

which combined with the stationarity of the process $Y^\star_s \sim \pi_T^{x,\psi^\star}$, gives

$$\mathbb{E}_{\pi_T^{x,\psi^\star}}[|Y|] \le \mathbb{E}_{\pi_T^{x,\psi^\star}}[|Y|^2]^{1/2} \le \frac{1 + T^{-1}|x| + \|g^\kappa_\nu\|_\infty + |\nabla\psi^\star(0)|}{\sqrt{4\varepsilon(\alpha_{\psi^\star} + T^{-1} - \varepsilon)}}\,.$$

By minimizing over $\varepsilon \in (0, \alpha_{\psi^\star} + T^{-1})$ we finally get the desired upper-bound

$$\mathbb{E}_{\pi_T^{x,\psi^\star}}[|Y|] \le \frac{1 + T^{-1}|x| + \|g^\kappa_\nu\|_\infty + |\nabla\psi^\star(0)|}{\alpha_{\psi^\star} + T^{-1}} = \frac{|x|}{T\,\alpha_{\psi^\star} + 1} + \frac{1 + \|g^\kappa_\nu\|_\infty + |\nabla\psi^\star(0)|}{\alpha_{\psi^\star} + T^{-1}}\,. \tag{62}$$

By combining the above bound with (61) we finally conclude that

$$|\nabla\varphi^{n+1} - \nabla\varphi^\star|(x) \le T^{-1}\,\gamma^\nu_n\Big(\frac{A}{T\,\alpha_{\psi^\star} + 1}\,|x| + A\frac{1 + \|g^\kappa_\nu\|_\infty + |\nabla\psi^\star(0)|}{\alpha_{\psi^\star} + T^{-1}} + B\Big) \le \hat{\gamma}^\nu_n(A\,|x| + B)$$

where the last step holds true for the chosen rate $\hat{\gamma}^\nu_n$. 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

$$\mathbb{E}_{\pi_T^{x,\varphi^\star}}[|Y|] < \frac{|x|}{T\,\alpha_{\varphi^\star} + 1} + \frac{1 + \|g^\kappa_\mu\|_\infty + |\nabla\varphi^\star(0)|}{\alpha_{\varphi^\star} + T^{-1}}\,. \tag{63}$$

∎

5.2. **Exponential convergence of dual and primal Sinkhorn algorithm.** Let us firstly show how the pointwise convergence of the gradients implies the (pointwise and integrated) convergence of the Sinkhorn iterates, *i.e.*, Theorem 11.

Proof of Theorem 11. Owing to the normalizations (26), (27) and to Theorem 7, we immediately deduce that

$$|\varphi^{\diamond n}-\varphi^\star|(x)=\left|\varphi^{\diamond n}(x)-\int\varphi^{\diamond n}\mathrm{d}\mu-\varphi^\star(x)+\int\varphi^\star\mathrm{d}\mu\right|=\left|\int(\varphi^n-\varphi^\star)(x)-(\varphi^n-\varphi^\star)(y)\,\mathrm{d}\mu(y)\right|$$
$$\le\int\left|(\varphi^n-\varphi^\star)(x)-(\varphi^n-\varphi^\star)(y)\right|\mathrm{d}\mu(y)\le\int\int_0^1|\nabla(\varphi^n-\varphi^\star)|(y+t(x-y))\,|x-y|\,\mathrm{d}t\,\mathrm{d}\mu(y)$$
$$\overset{(20)}{\le}\frac{1}{\hat\gamma^\mu_{n-1}}\prod_{k=0}^{n-1}\hat\gamma^\mu_k\,\hat\gamma^\nu_k\int\int_0^1(A|y+t(x-y)|+B)|x-y|\,\mathrm{d}t\,\mathrm{d}\mu(y)$$
$$\le\frac{1}{\hat\gamma^\mu_{n-1}}\prod_{k=0}^{n-1}\hat\gamma^\mu_k\,\hat\gamma^\nu_k\left[A\,|x|^2+(A\,M_1(\mu)+B)\,|x|+B\,M_1(\mu)+2\,A\,M_2(\mu)\right].$$

The second pointwise bound can be proven in the same fashion. □

As a consequence of Theorem 11 we may also deduce the convergence of the L^1-norms along the adjusted marginals and along the real marginals.

Corollary 27. *Assume the validity of* ***A****1 and* ***A****2 and* (19) *for some positive constants* A, $B>0$. *Then for any* $n\ge1$ *it holds*

$$\begin{aligned}&\|\varphi^{\diamond n}-\varphi^\star\|_{\mathrm{L}^1(\mu)}\le\frac{1}{\hat\gamma^\mu_{n-1}}\prod_{k=0}^{n-1}\hat\gamma^\mu_k\,\hat\gamma^\nu_k\left[3\,A\,M_2(\mu)+(A\,M_1(\mu)+B)\,M_1(\mu)+B\,M_1(\mu)\right],\\&\|\psi^{\diamond n}-\psi^\star\|_{\mathrm{L}^1(\nu)}\le\prod_{k=0}^{n-1}\hat\gamma^\mu_k\,\hat\gamma^\nu_k\left[3\,A\,M_2(\nu)+(A\,M_1(\nu)+B)\,M_1(\nu)+B\,M_1(\nu)\right],\end{aligned}\tag{64}$$

and

$$\begin{aligned}&\|\varphi^{\diamond n}-\varphi^\star\|_{\mathrm{L}^1(\mu^n)}\le\frac{C(A,B,\mu)}{\hat\gamma^\mu_{n-1}}\prod_{k=0}^{n-1}\hat\gamma^\mu_k\,\hat\gamma^\nu_k,\\&\|\psi^{\diamond n}-\psi^\star\|_{\mathrm{L}^1(\nu^n)}\le C(A,B,\nu)\prod_{k=0}^{n-1}\hat\gamma^\mu_k\,\hat\gamma^\nu_k,\end{aligned}\tag{65}$$

where

$$\begin{aligned}C(A,B,\mu)\coloneqq&\left[3\,A\,M_2(\mu)+(A\,M_1(\mu)+B)\,M_1(\mu)+B\,M_1(\mu)\right]\\&+A\,C_2(\mu)\left(\sqrt{\mathcal{H}(\mu^1|\mu)}+\frac{\mathcal{H}(\mu^1|\mu)}{2}\right)+(A\,M_1(\mu)+B)\,C_1(\mu)\,\sqrt{\mathcal{H}(\mu^1|\mu)}\end{aligned}\tag{66}$$

and

$$\begin{aligned}C(A,B,\nu)\coloneqq&\left[3\,A\,M_2(\nu)+(A\,M_1(\nu)+B)\,M_1(\nu)+B\,M_1(\nu)\right]\\&+A\,C_2(\nu)\left(\sqrt{\mathcal{H}(\nu^0|\nu)}+\frac{\mathcal{H}(\nu^0|\nu)}{2}\right)+(A\,M_1(\nu)+B)\,C_1(\nu)\,\sqrt{\mathcal{H}(\nu^0|\nu)}\,.\end{aligned}$$

Proof. The proof of the integrated bounds along the marginals (64) is a straightforward consequence of the pointwise convergence, whereas the bounds along the adjusted marginals is a consequence of the weighted Csiszár-Kullback-Pinsker inequalities [10, Theorem 2.1] which imply

$$M_1(\mu^n) \le M_1(\mu) + C_1(\mu)\sqrt{\mathcal{H}(\mu^n|\mu)} \quad \text{and} \quad M_2(\mu^n) \le M_2(\mu) + C_2(\mu)\left(\sqrt{\mathcal{H}(\mu^n|\mu)} + \frac{\mathcal{H}(\mu^n|\mu)}{2}\right)$$

$$M_1(\nu^n) \le M_1(\nu) + C_1(\nu)\sqrt{\mathcal{H}(\nu^n|\nu)} \quad \text{and} \quad M_2(\nu^n) \le M_2(\nu) + C_2(\nu)\left(\sqrt{\mathcal{H}(\nu^n|\nu)} + \frac{\mathcal{H}(\nu^n|\nu)}{2}\right),$$

where $C_1(\mu)$, $C_1(\nu)$, $C_2(\mu)$, $C_2(\nu)$ are positive constants (independent from $n \in \mathbb{N}$). For sake of clarity we postponed the proof of the above moment bounds to Lemma 28, below. The proof of (65) then follows from the fact that 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 [59, 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 28. *Assume **A**2-(i). Then for any probability measure $\mathfrak{p} \in \mathcal{P}(\mathbb{R}^d)$ such that $\mathcal{H}(\mathfrak{p}|\nu) < +\infty$ it holds*

$$M_1(\mathfrak{p}) \le M_1(\nu) + C_1(\nu)\sqrt{\mathcal{H}(\mathfrak{p}|\nu)} \quad \textit{and} \quad M_2(\mathfrak{p}) \le M_2(\nu) + C_2(\nu)\left(\sqrt{\mathcal{H}(\mathfrak{p}|\nu)} + \frac{\mathcal{H}(\mathfrak{p}|\nu)}{2}\right)$$

$$\textit{with}\ \ C_1(\nu) \coloneqq \inf_{\sigma_\nu \in (0,\alpha_\nu/2)} \left(\frac{2}{\sigma_\nu} + \frac{2}{\sigma_\nu}\log\int e^{\sigma_\nu|x|^2}\,\mathrm{d}\nu\right)^{1/2}$$

$$\textit{and}\ \ C_2(\nu) \coloneqq \inf_{\sigma_\nu \in (0,\alpha_\nu/2)} \left(\frac{3}{\sigma_\nu} + \frac{2}{\sigma_\nu}\int e^{\sigma_\nu|x|^2}\mathrm{d}\nu\right).$$

*Similarly if **A**2-(ii) holds true, then for any probability measure $\mathfrak{p} \in \mathcal{P}(\mathbb{R}^d)$ such that $\mathcal{H}(\mathfrak{p}|\mu) < +\infty$ it holds*

$$M_1(\mathfrak{p}) \le M_1(\mu) + C_1(\mu)\sqrt{\mathcal{H}(\mathfrak{p}|\mu)} \quad \textit{and} \quad M_2(\mathfrak{p}) \le M_2(\nu) + C_2(\mu)\left(\sqrt{\mathcal{H}(\mathfrak{p}|\mu)} + \frac{\mathcal{H}(\mathfrak{p}|\mu)}{2}\right)$$

$$\textit{with}\ \ C_1(\mu) \coloneqq \inf_{\sigma_\mu \in (0,\alpha_\mu/2)} \left(\frac{2}{\sigma_\mu} + \frac{2}{\sigma_\mu}\log\int e^{\sigma_\mu|x|^2}\,\mathrm{d}\mu\right)^{1/2}$$

$$\textit{and}\ \ C_2(\mu) \coloneqq \inf_{\sigma_\mu \in (0,\alpha_\mu/2)} \left(\frac{3}{\sigma_\mu} + \frac{2}{\sigma_\mu}\int e^{\sigma_\mu|x|^2}\mathrm{d}\mu\right).$$

Proof. For any $\sigma_\nu \in (0, \alpha_\nu/2)$, from the weighted Csiszár-Kullback-Pinsker inequalities [10, 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

$$M_1(\mathfrak{p}) = M_1(\nu) + \sigma_\nu^{-1/2}\int \sigma_\nu^{1/2}\,|x|\,\mathrm{d}(\mathfrak{p}-\nu) \le M_1(\nu) + \sqrt{\mathcal{H}(\mathfrak{p}|\nu)}\left(\frac{2}{\sigma_\nu} + \frac{2}{\sigma_\nu}\log\int e^{\sigma_\nu|x|^2}\,\mathrm{d}\nu\right)^{1/2}$$

$$M_2(\mathfrak{p}) = M_2(\nu) + \frac{2}{\sigma_\nu}\int\frac{\sigma_\nu}{2}\,|x|^2\,\mathrm{d}(\mathfrak{p}-\nu) \le M_2(\nu) + \left(\sqrt{\mathcal{H}(\mathfrak{p}|\nu)} + \frac{\mathcal{H}(\mathfrak{p}|\nu)}{2}\right)\left(\frac{3}{\sigma_\nu} + \frac{2}{\sigma_\nu}\int e^{\sigma_\nu|x|^2}\mathrm{d}\nu\right)$$

which are finite because of Lemma 37. Minimizing 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. □

Finally, let us conclude the section with the proof of the exponential convergence of Sinkhorn algorithm on the primal side.

Proof of Theorem 12. Let us preliminary point out that as a first consequence of Corollary 27, 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, (1) and (6) imply that

$$\log \frac{\mathrm{d}\pi_T^\star}{\mathrm{d}\pi^{n,n}}(x,y) = \varphi^n(x) - \varphi^\star(x) + \psi^n(y) - \psi^\star(y)\,,$$

and hence the symmetric relative entropies can be rewritten as

$$\begin{aligned}\mathscr{H}(\pi^{n,n}|\pi_T^\star) + \mathscr{H}(\pi_T^\star|\pi^{n,n}) &= \int (\varphi^n - \varphi^\star) \oplus (\psi^n - \psi^\star)\, \mathrm{d}\pi_T^\star - \int (\varphi^n - \varphi^\star) \oplus (\psi^n - \psi^\star)\, \mathrm{d}\pi^{n,n} \\ &= \int (\varphi^{\diamond n} - \varphi^\star) \oplus (\psi^{\diamond n} - \psi^\star)\, \mathrm{d}\pi_T^\star - \int (\varphi^{\diamond n} - \varphi^\star) \oplus (\psi^{\diamond n} - \psi^\star)\, \mathrm{d}\pi^{n,n} \\ &= \int (\varphi^{\diamond n} - \varphi^\star)\, \mathrm{d}\mu^n - \int (\varphi^{\diamond n} - \varphi^\star)\, \mathrm{d}\mu\,.\end{aligned}$$

By combining the above with Corollary 27, we then deduce that

$$\mathscr{H}(\pi^{n,n}|\pi_T^\star) + \mathscr{H}(\pi_T^\star|\pi^{n,n}) \leq \frac{D(A,B,\mu)}{\hat{\gamma}_{n-1}^\mu} \prod_{k=0}^{n-1} \hat{\gamma}_k^\mu\, \hat{\gamma}_k^\nu\,,$$

with

$$\begin{aligned}D(A,B,\mu) &\coloneqq C(A,B,\mu) + 3\,A\,M_2(\mu) + (A\,M_1(\mu) + B)\,M_1(\mu) + B\,M_1(\mu) \\ &= 2\Big[3\,A\,M_2(\mu) + (A\,M_1(\mu) + B)\,M_1(\mu) + B\,M_1(\mu)\Big] \\ &\quad + A\,C_2(\mu)\left(\sqrt{\mathcal{H}(\mu^1|\mu)} + \frac{\mathcal{H}(\mu^1|\mu)}{2}\right) + (A\,M_1(\mu) + B)\,C_1(\mu)\sqrt{\mathcal{H}(\mu^1|\mu)}\,,\end{aligned} \tag{67}$$

with $C_1(\mu)$, $C_2(\mu)$ being the constants introduced in Lemma 28.

Similarly (1) and (5) imply

$$\log \frac{\mathrm{d}\pi_T^\star}{\mathrm{d}\pi^{n+1,n}}(x,y) = \varphi^{n+1}(x) - \varphi^\star(x) + \psi^n(y) - \psi^\star(y)\,,$$

hence

$$\mathscr{H}(\pi^{n+1,n}|\pi_T^\star) + \mathscr{H}(\pi_T^\star|\pi^{n+1,n}) = \int (\psi^{\diamond n} - \psi^\star)\, \mathrm{d}\nu^n - \int (\psi^{\diamond n} - \psi^\star)\, \mathrm{d}\nu\,.$$

The latter combined with Corollary 27 yields to

$$\mathscr{H}(\pi^{n+1,n}|\pi_T^\star) + \mathscr{H}(\pi_T^\star|\pi^{n+1,n}) \leq D(A,B,\nu) \prod_{k=0}^{n-1} \hat{\gamma}_k^\mu\, \hat{\gamma}_k^\nu\,,$$

with

$$
\begin{aligned}
D(A,B,\nu) \coloneqq\, & C(A,B,\nu) + 3\,A\,M_2(\nu) + (A\,M_1(\nu)+B)\,M_1(\nu) + B\,M_1(\nu)\\
=\, & 2\Big[3\,A\,M_2(\nu) + (A\,M_1(\nu)+B)\,M_1(\nu) + B\,M_1(\nu)\Big]\\
& + A\,C_2(\nu)\bigg(\sqrt{\mathcal{H}(\nu^0|\nu)} + \frac{\mathcal{H}(\nu^0|\nu)}{2}\bigg) + (A\,M_1(\nu)+B)\,C_1(\nu)\,\sqrt{\mathcal{H}(\nu^0|\nu)}\,.
\end{aligned} \tag{68}
$$

Finally let us notice that from (5) and (6) we may also deduce that

$$
\begin{aligned}
\mathscr{H}(\pi^{n,n}|\pi^{n,n-1}) = \mathscr{H}(\mu^n|\mu) \quad &\text{and} \quad \mathscr{H}(\pi^{n,n-1}|\pi^{n,n}) = \mathscr{H}(\mu|\mu^n)\,,\\
\mathscr{H}(\pi^{n+1,n}|\pi^{n,n}) = \mathscr{H}(\nu^n|\nu) \quad &\text{and} \quad \mathscr{H}(\pi^{n,n}|\pi^{n+1,n}) = \mathscr{H}(\nu|\nu^n)\,.
\end{aligned}
$$

and hence the bound for the adjusted marginals is a consequence of the previous ones, and the trivial inequalities

$$
\begin{aligned}
&\mathscr{H}(\pi^{n,n}|\pi^{n,n-1}) = \mathscr{H}(\mu^n|\mu) \le \mathscr{H}(\pi^{n,n}|\pi^\star_T)\,, \quad \mathscr{H}(\pi^{n,n-1}|\pi^{n,n}) = \mathscr{H}(\mu|\mu^n) \le \mathscr{H}(\pi^\star_T|\pi^{n,n})\,,\\
&\mathscr{H}(\pi^{n+1,n}|\pi^{n,n}) = \mathscr{H}(\nu^n|\nu) \le \mathscr{H}(\pi^{n+1,n}|\pi^\star_T)\,, \quad \mathscr{H}(\pi^{n,n}|\pi^{n+1,n}) = \mathscr{H}(\nu|\nu^n) \le \mathscr{H}(\pi^\star_T|\pi^{n+1,n})\,.
\end{aligned}
$$

□

Remark 29. *Let us remark here that from the previous result we can consider a sharper multiplicative constant $C_S(A,B,\mu)$ in Corollary 27. Indeed, instead of relying on the monotonicity of relative entropies along Sinkhorn algorithm, in (66) we could define $C_S(A,B,\mu)$ as*

$$
C_S(A,B,\mu) \coloneqq \Big[3\,A\,M_2(\mu) + (A\,M_1(\mu)+B)\,M_1(\mu) + B\,M_1(\mu)\Big] + \varepsilon_\mu(n)\,,
$$

where

$$
\varepsilon_\mu(n) \coloneqq A\,C_2(\mu)\bigg(\sqrt{\mathcal{H}(\mu^n|\mu)} + \frac{\mathcal{H}(\mu^n|\mu)}{2}\bigg) + (A\,M_1(\mu)+B)\,C_1(\mu)\,\sqrt{\mathcal{H}(\mu^n|\mu)}
$$

is a positive constant exponentially small as $n\uparrow+\infty$ thanks to Theorem 12. Then, in Theorem 12 instead of (67), we can consider the sharper multiplicative constant

$$
D_S(A,B,\mu) \coloneqq 2\Big[3\,A\,M_2(\mu) + (A\,M_1(\mu)+B)\,M_1(\mu) + B\,M_1(\mu)\Big] + \varepsilon_\mu(n)\,.
$$

The same reasoning applies for $C_S(A,B,\nu)$ and $D_S(A,B,\nu)$.

5.3. **Exponential convergence of the Hessians along Sinkhorn algorithm.** For notations sake for any couple of vectors $v,\,w\in\mathbb{R}^d$ we will denote by $v\otimes w = v\,w^\intercal$ the matrix given by their tensor product. Notice that for any $v,\,w\in\mathbb{R}^d$ the Frobenius norm of their product reads as

$$
\|v\otimes w\|_{\mathrm{F}} = |v|\,|w|\,. \tag{69}
$$

Let us also recall here that for $h\in\{\varphi^n,\,\psi^n,\,\varphi^\star,\,\psi^\star\}$ the gradient and the Hessian along the semigroup has the explicit formulation (cf. (10) and (28) respectively)

$$
\begin{aligned}
\nabla\log P_T\exp(-h)(x) &= \frac{1}{T}\int (y-x)\,\pi_T^{x,h}(\mathrm{d}y)\\
\nabla^2\log P_T\exp(-h)(x) &= -\,T^{-1}\,\mathrm{Id} + T^{-2}\,\mathrm{Cov}(\pi_T^{x,h})\,.
\end{aligned}
$$

where we recall the conditional distribution $\pi_T^{x,h}$ being defined at (11)

$$\pi_T^{x,h}(\mathrm{d}y) \propto \exp\Big(-\frac{|y-x|^2}{2T} - h(y)\Big)\mathrm{d}y\ .$$

Throught this section we will always assume the validity of the hypothesis of Theorem 7, *i.e.* **A** 1, **A**2 and that (19) is met.

Lemma 30. *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*

$$\|\nabla^2\varphi^{n+1} - \nabla^2\varphi^\star\|_{\mathrm{F}}(x) \le T^{-2}\,\mathbb{E}_{(Y,Z)\sim\pi}\Big[|Y-Z|\,(|Y|+|Z|+C_1\,T^{-1}\,|x|+C_2)\Big]\ . \tag{70}$$

Proof. As a byproduct of (4) and (28) we immediately have for any $x \in \mathbb{R}^d$ and for any coupling $\pi \in \Pi(\pi_T^{x,\psi^\star}, \pi_T^{x,\psi^n})$

$$\begin{aligned}
T^2\,(\nabla^2\varphi^{n+1} - \nabla^2\varphi^\star)(x) =& \mathrm{Cov}(\pi_T^{x,\psi^\star}) - \mathrm{Cov}(\pi_T^{x,\psi^n})\\
=& \mathbb{E}_{(Y,Z)\sim\pi}\Big[Y^{\otimes 2} - Z^{\otimes 2}\Big] + \mathbb{E}_{Z\sim\pi_T^{x,\psi^n}}[Z]^{\otimes 2} - \mathbb{E}_{Y\sim\pi_T^{x,\psi^\star}}[Y]^{\otimes 2}\\
=& \mathbb{E}_{(Y,Z)\sim\pi}\Big[(Y-Z)\otimes Y\Big] + \mathbb{E}_{(Y,Z)\sim\pi}\Big[Z\otimes(Y-Z)\Big]\\
&- \mathbb{E}_{(Y,Z)\sim\pi}\Big[Y-Z\Big]\,\mathbb{E}_{Z\sim\pi_T^{x,\psi^n}}[Z] - \mathbb{E}_{Y\sim\pi_T^{x,\psi^\star}}[Y]\,\mathbb{E}_{(Y,Z)\sim\pi}\Big[Y-Z\Big]
\end{aligned}$$

By applying the Frobenius norm, and recalling (69), we then deduce that

$$T^2\,\|\nabla^2\varphi^{n+1} - \nabla^2\varphi^\star\|_{\mathrm{F}}(x) \le \mathbb{E}_{(Y,Z)\sim\pi}\Big[|Y-Z|\,\Big(|Y|+|Z|+\mathbb{E}_{Z\sim\pi_T^{x,\psi^n}}[|Z|] + \mathbb{E}_{Y\sim\pi_T^{x,\psi^\star}}[|Y|]\Big)\Big]$$

In order to bound the last two expected values in the right hand side we will proceed as in the proof of Lemma 26. Particularly, (62) already proves that

$$\mathbb{E}_{Y\sim\pi_T^{x,\psi^\star}}[|Y|] < \frac{|x|}{T\,\alpha_{\psi^\star}+1} + \frac{1+\|g_\nu^\kappa\|_\infty + |\nabla\psi^\star(0)|}{\alpha_{\psi^\star}+T^{-1}}\ .$$

By reasoning in the same way we can prove that

$$\mathbb{E}_{Z\sim\pi_T^{x,\psi^n}}[|Z|] < \frac{|x|}{T\,\alpha_{\nu,n}+1} + \frac{1+\|g_\nu^\kappa\|_\infty + |\nabla\psi^n(0)|}{\alpha_{\psi^n}+T^{-1}}\ ,$$

Particularly, the pointwise convergence of the gradients of Theorem 7 and the convergences $\alpha_{\mu,n}\uparrow \alpha_{\varphi^\star}$, $\alpha_{\nu,n}\uparrow\alpha_{\psi^\star}$ stated in Theorem 14 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. □

Before moving to the proof of the Hessians' convergence, let us recall here (cf. Corollary 16) that $V(y) = 1+|y|^2$ is a Lyapunov function for (12) 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

$$\mathcal{L}_{\psi^\star}V(y)\ \vee\ \mathcal{L}_{\psi^n}V(y)\ \le B_\nu - A_\nu\,V(y)\ \text{ and }\ \mathcal{L}_{\varphi^\star}V(y)\ \vee\ \mathcal{L}_{\varphi^n}V(y)\ \le B_\mu - A_\mu\,V(y)\ , \tag{71}$$

where $\mathcal{L}_h := \Delta/2 - \frac{1}{2}\langle T^{-1}(y-x) + \nabla h(y),\, \nabla\rangle$ is the generator associated to (12). 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 7 and the convergences $\alpha_{\mu,n} \uparrow \alpha_{\varphi^\star}$, $\alpha_{\nu,n} \uparrow \alpha_{\psi^\star}$ stated in Theorem 14.

Proof of Theorem 13. For sake of notations let us introduce the constant $C_x := \max\{1,\, C_1\, T^{-1}\,|x| + C_2\}$. We will proceed as in Corollary 24, this time considering a distorted Wasserstein semi-distance

$$\mathcal{W}_{f_x^\nu}(\cdot,\cdot) := \inf_{\pi\in\Pi(\cdot,\cdot)} \mathbb{E}_{(Y,Z)\sim\pi}\Big[f_x^\nu(|Y-Z|)\,(1+\varepsilon V(Y)+\varepsilon V(Z))\Big]\,. \tag{72}$$

In the above definition $V(y) := 1+|y|^2$ and we consider the bounded concave function f_x^ν and the parameter $\varepsilon\in(0,1)$, small enough such that (84) holds. For sake of clarity we have postponed their definition as well as the proof of their properties to the end of this section. Let us just state here that there is no dependence from the index $n\in\mathbb{N}$, $\mathcal{W}_{f_x^\nu}$ contracts along the semigroup $(P_t^n)_{t\geq 0}$ associated to the SDE

$$\mathrm{d}Y_t^n = -\Big(\frac{Y_t^n - x}{2T} + \frac{1}{2}\nabla\psi^n(Y_t^n)\Big)\mathrm{d}t + \mathrm{d}B_t\,,$$

and there exists a rate $\lambda>0$ (independent from $n\in\mathbb{N}$, explicitly given at (92)) such that

$$\mathcal{W}_{f_x^\nu}(\mu_1\,P_t^n,\,\mu_2\,P_t^n) \leq e^{-\lambda\,t}\,\mathcal{W}_{f_x^\nu}(\mu_1,\,\mu_2) \quad \forall\,\mu_1,\mu_2\in\mathcal{P}(\mathbb{R}^d)\,. \tag{73}$$

Moreover, $\mathcal{W}_{f_x^\nu}$ satisfies a weak triangle inequality, *i.e.*

$$\mathcal{W}_{f_x^\nu}(\mu_1,\mu_2) \leq C_\Delta(\mathcal{W}_{f_x^\nu}(\mu_1,\mu_3) + \mathcal{W}_{f_x^\nu}(\mu_3,\mu_2)) \quad \forall\,\mu_1,\mu_2,\mu_3\in\mathcal{P}(\mathbb{R}^d)\,. \tag{74}$$

with $C_\Delta = \max\{3,\, 2+2\,R_2^2\}$, with the radius $R_2>0$ such that

$$\begin{aligned} &\frac{C_I}{2}\,r \leq f_x^\nu(r) \leq r \quad \text{and} \quad \frac{1}{2} \leq (f_x^\nu)'(r) \leq 1 \quad \forall\, r<R_2\,,\\ &f_x^\nu(r) = f_x^\nu(R_2) \quad \text{and hence} \quad (f_x^\nu)'(r) = 0 \quad \forall\, r>R_2\,, \end{aligned} \tag{75}$$

for a positive constant $C_I\in(0,1)$. This particularly implies that for any coupling $\pi\in\Pi(\pi_T^{x,\psi^\star},\,\pi_T^{x,\psi^n})$ we have

$$\begin{aligned} T^{-2}\,\mathbb{E}_{(Y,Z)\sim\pi}\Big[|Y-Z|\,(|Y|+|Z|+C_x)\Big] &\leq \frac{C_x}{T^2\,\varepsilon^{1/2}}\,\mathbb{E}_{(Y,Z)\sim\pi}\Big[|Y-Z|\,(1+\varepsilon^{1/2}\,|Y|+\varepsilon^{1/2}\,|Z|)\Big]\\ &\leq \begin{cases} \frac{2}{C_I}\,\frac{C_x}{T^2\,\varepsilon^{1/2}}\,\mathbb{E}_{(Y,Z)\sim\pi}\Big[f_x^\nu(|Y-Z|)\,(1+\varepsilon^{1/2}\,|Y|+\varepsilon^{1/2}\,|Z|)\Big] & \text{if } |Y-Z|\leq R_2\,,\\ \frac{C_x}{T^2\,\varepsilon}\,f_x^\nu(R_2)^{-1}\,\mathbb{E}_{(Y,Z)\sim\pi}\Big[f_x^\nu(R_2)\,(1+\varepsilon^{1/2}\,|Y|+\varepsilon^{1/2}\,|Z|)^2\Big] & \text{if } |Y-Z|>R_2 \end{cases}\\ &\leq \frac{10}{3}\,\frac{C_x}{T^2\,\varepsilon^{1/2}}\left(\frac{2}{C_I}\vee\frac{f_x^\nu(R_2)^{-1}}{\varepsilon^{1/2}}\right)\mathbb{E}_{(Y,Z)\sim\pi}\Big[f_x^\nu(|Y-Z|)\,(1+\varepsilon\,V(Y)+\varepsilon\,V(Z))\Big]\,. \end{aligned}$$

Owing to (70) and by minimizing the above over $\pi\in\Pi(\pi_T^{x,\psi^\star},\,\pi_T^{x,\psi^n})$, we deduce then that

$$\|\nabla^2\varphi^{n+1} - \nabla^2\varphi^\star\|_{\mathrm{F}}(x) \leq \frac{10}{3}\,\frac{C_x}{T^2\,\varepsilon^{1/2}}\left(\frac{2}{C_I}\vee\frac{f_x^\nu(R_2)^{-1}}{\varepsilon^{1/2}}\right)\mathcal{W}_{f_x^\nu}(\pi_T^{x,\psi^\star},\,\pi_T^{x,\psi^n})\,. \tag{76}$$

Now, recall that $\pi_T^{x,\psi^\star}$ corresponds to the invariant probability of the SDE

$$\mathrm{d}Y_t^\star = -\Big(\frac{Y_t^\star - x}{2T} + \frac{1}{2}\nabla\psi^\star(Y_t^\star)\Big)\mathrm{d}t + \mathrm{d}B_t$$

and let $(P^\star_t)_{t\geq 0}$ be its corresponding semigroup. Particularly we have $\pi_T^{x,\psi^\star} P^\star_t = \pi_T^{x,\psi^\star}$ for any $t\geq 0$. Similarly, for any $t\geq 0$ it holds $\pi_T^{x,\psi^n} P^n_t = \pi_T^{x,\psi^n}$, with $(P^n_t)_{t\geq 0}$ being the semigroup introduced above.

Given the above premises, from the weak triangle inequality (74) we deduce that

$$\begin{aligned}\mathcal{W}_{f^\nu_x}(\pi_T^{x,\psi^\star},\,\pi_T^{x,\psi^n}) &= \mathcal{W}_{f^\nu_x}(\pi_T^{x,\psi^\star} P^\star_t,\,\pi_T^{x,\psi^n} P^n_t)\\ &\leq C_\Delta\left(\mathcal{W}_{f^\nu_x}(\pi_T^{x,\psi^\star} P^\star_t,\,\pi_T^{x,\psi^\star} P^n_t)+\mathcal{W}_{f^\nu_x}(\pi_T^{x,\psi^\star} P^n_t,\,\pi_T^{x,\psi^n} P^n_t)\right)\\ &\overset{(73)}{\leq} C_\Delta\,\mathcal{W}_{f^\nu_x}(\pi_T^{x,\psi^\star} P^\star_t,\,\pi_T^{x,\psi^\star} P^n_t)+C_\Delta\,e^{-\lambda\,t}\,\mathcal{W}_{f^\nu_x}(\pi_T^{x,\psi^\star},\,\pi_T^{x,\psi^n})\,.\end{aligned}$$

Therefore for any $t>\lambda^{-1}\log C_\Delta$ it holds

$$\mathcal{W}_{f^\nu_x}(\pi_T^{x,\psi^\star},\,\pi_T^{x,\psi^n}) \leq \frac{C_\Delta}{1-C_\Delta\,e^{-\lambda\,t}}\,\mathcal{W}_{f^\nu_x}(\pi_T^{x,\psi^\star} P^\star_t,\,\pi_T^{x,\psi^\star} P^n_t)\,. \tag{77}$$

We will show that $\mathcal{W}_{f^\nu_x}(\pi_T^{x,\psi^\star} P^\star_t,\,\pi_T^{x,\psi^\star} P^n_t)$ is exponentially small as n diverges, by relying on the synchronous coupling technique. Hence consider the diffusion processes

$$\begin{cases}\mathrm{d}Y^\star_t = -\left(\frac{Y^\star_t - x}{2T}+\frac12\nabla\psi^\star(Y^\star_t)\right)\mathrm{d}t+\mathrm{d}B_t\,,\\ \mathrm{d}Z_t = -\left(\frac{Z_t - x}{2T}+\frac12\nabla\psi^n(Z_t)\right)\mathrm{d}t+\mathrm{d}B_t\,,\\ Y^\star_0 = Z_0\sim\pi_T^{x,\psi^\star}\,,\end{cases} \tag{78}$$

where $(B_t)_{t\geq 0}$ is the same d-dimensional Brownian motion. Notice that $Y^\star_t\sim\pi_T^{x,\psi^\star}P^\star_t$ whereas $Z_t\sim\pi_T^{x,\psi^\star}P^n_t$, which yields to

$$\mathcal{W}_{f^\nu_x}(\pi_T^{x,\psi^\star} P^\star_t,\,\pi_T^{x,\psi^\star} P^n_t) \leq \mathbb{E}_{Y^\star_t,Z_t}\left[f^\nu_x(|Y^\star_t - Z_t|)\,(1+\varepsilon V(Y^\star_t)+\varepsilon V(Z^\star_t))\right]\,. \tag{79}$$

By construction we immediately get

$$\mathrm{d}|Y^\star_t - Z_t| = -\frac{1}{2T}|Y^\star_t - Z_t|\,\mathrm{d}t-\frac12\left\langle\nabla\psi^\star(Y^\star_t)-\nabla\psi^n(Z_t),\,\frac{Y^\star_t - Z_t}{|Y^\star_t - Z_t|}\right\rangle\mathrm{d}t$$

and hence if we introduce the stopping time $\tau\coloneqq\inf\{t>0\colon |Y^\star_t - Z_t|>R_2\}$, we deduce that for any $t<\tau$ it holds

$$\begin{aligned}\mathrm{d}f^\nu_x(|Y^\star_t - Z_t|) &= -\frac{(f^\nu_x)'(|Y^\star_t - Z_t|)}{2}\left(T^{-1}|Y^\star_t - Z_t|+\left\langle\nabla\psi^\star(Y^\star_t)-\nabla\psi^n(Z_t),\,\frac{Y^\star_t - Z_t}{|Y^\star_t - Z_t|}\right\rangle\right)\mathrm{d}t\\ \leq -(T^{-1}+\kappa_{\psi^n}(|Y^\star_t - Z_t|))&\,\frac{(f^\nu_x)'(|Y^\star_t - Z_t|)}{2}\,|Y^\star_t - Z_t|\,\mathrm{d}t+\frac{(f^\nu_x)'(|Y^\star_t - Z_t|)}{2}\,|\nabla\psi^\star-\nabla\psi^n|(Y^\star_t)\mathrm{d}t\\ &\overset{(31)}{\leq} -(T^{-1}+\alpha_{\nu,n}-|Y^\star_t - Z_t|^{-1}g^\kappa_\nu(|Y^\star_t - Z_t|))\,\frac{(f^\nu_x)'(|Y^\star_t - Z_t|)}{2}\,|Y^\star_t - Z_t|\,\mathrm{d}t\\ &\quad+\frac{(f^\nu_x)'(|Y^\star_t - Z_t|)}{2}\,|\nabla\psi^\star-\nabla\psi^n|(Y^\star_t)\mathrm{d}t\,.\end{aligned}$$

Since $\alpha_{\nu,n} > \alpha_\nu - T^{-1}$ (cf. Theorem 14), the sublinearity $g^\kappa_\nu(r) \leq G^\kappa_\nu\, r$ and the upper bound (75) for $(f^\nu_x)'$, imply

$$\begin{aligned}\mathrm{d}f^\nu_x(|Y^\star_t - Z_t|) &\leq \frac{(G^\kappa_\nu - \alpha_\nu)^+}{2}\,|Y^\star_t - Z_t|\,\mathrm{d}t + \frac{1}{2}\,|\nabla\psi^\star - \nabla\psi^n|(Y^\star_t)\mathrm{d}t\\ &\overset{(75)}{\leq} C_I^{-1}\,(G^\kappa_\nu - \alpha_\nu)^+\, f^\nu_x(|Y^\star_t - Z_t|)\,\mathrm{d}t + \frac{1}{2}\,|\nabla\psi^\star - \nabla\psi^n|(Y^\star_t)\mathrm{d}t\,,\end{aligned}$$

where the last step is possible since for any $t < \tau$ it holds $|Y^\star_t - Z_t| \leq R_2$. Since $f^\nu_x(|Y^\star_t - Z_t|) = f^\nu_x(R_2)$ for any $t \geq \tau$, the above differential inequality holds for any $t \geq 0$ and Gronwall Lemma finally yields for any $t \geq 0$ to

$$\begin{aligned}f^\nu_x(|Y^\star_t - Z_t|) &\leq \exp(C_I^{-1}\,(G^\kappa_\nu - \alpha_\nu)^+\,t)\,\frac{1}{2}\int_0^t |\nabla\psi^\star - \nabla\psi^n|(Y^\star_s)\,\mathrm{d}s\\ &\overset{(20)}{\leq} \exp(C_I^{-1}\,(G^\kappa_\nu - \alpha_\nu)^+\,t)\,\frac{1}{2}\int_0^t (A|Y^\star_s| + B)\mathrm{d}\,s\prod_{k=0}^{n-1}\hat\gamma^\mu_k\,\hat\gamma^\nu_k\,.\end{aligned}$$

where the last step follows from Theorem 7. By recalling (79), so far we have proven that

(80)
$$\mathcal{W}_{f^\nu_x}(\pi^{x,\psi^\star}_T P^\star_t, \pi^{x,\psi^\star}_T P^n_t) \leq \frac{1}{2}\,e^{C_I^{-1}(G^\kappa_\nu - \alpha_\nu)^+ t}\int_0^t \mathbb{E}\Big[(A|Y^\star_s| + B)(1 + \varepsilon V(Y^\star_t) + \varepsilon V(Z_t))\Big]\mathrm{d}s\prod_{k=0}^{n-1}\hat\gamma^\mu_k\,\hat\gamma^\nu_k\,.$$

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 $Y^\star_s \sim \pi^{x,\psi^\star}_T$ we have

(81)
$$\begin{aligned}&\int_0^t \mathbb{E}\Big[(A|Y^\star_s| + B)(1 + \varepsilon V(Y^\star_t) + \varepsilon V(Z_t))\Big]\,\mathrm{d}s\\ &\leq Bt\Big(1 + \varepsilon\mathbb{E}[V(Y^\star_0)] + \varepsilon\mathbb{E}[V(Z_t)]\Big) + A\,t\,\mathbb{E}[|Y^\star_0|] + \frac{A\varepsilon t}{2}(2\,\mathbb{E}[|Y^\star_0|^2] + \mathbb{E}[V(Y^\star_0)^2] + \mathbb{E}[V(Z_t)^2])\,.\end{aligned}$$

At this point it is enough noticing that the geometric drift condition (38) obtained in the proof Corollary 16 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 39 in the Appendix, where we show in (104) 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 (76), (77) and (80), and by minimizinge over $t > \lambda^{-1}\log C_\Delta$ we have proven that

$$\|\nabla^2\varphi^{n+1} - \nabla^2\varphi^\star\|_{\mathrm{F}}(x) \leq C(x,T,\nu,A,B)\prod_{k=0}^{n-1}\hat\gamma^\mu_k\,\hat\gamma^\nu_k\,,$$

where the above constant is equal to

(82)
$$C(x,T,\nu,A,B) \coloneqq \inf_{t>\lambda^{-1}\log C_\Delta}\frac{10}{3}\,\frac{C_x}{T^2\varepsilon^{1/2}}\left(\frac{2}{C_I}\vee\frac{f^\nu_x(R_2)^{-1}}{\varepsilon^{1/2}}\right)\frac{C_\Delta}{1 - C_\Delta e^{-\lambda t}}\,\frac{1}{2}e^{C_I^{-1}(G^\kappa_\nu - \alpha_\nu)^+ t}U(t,\nu,x,A,B,T)\,.$$

By following the same line of reasoning, it is possible proving that

$$\|\nabla^2\psi^{n+1} - \nabla^2\psi^\star\|_{\mathrm{F}}(x) \leq C(x,T,\mu,A,B)\,\frac{1}{\hat\gamma^\mu_n}\prod_{k=0}^{n}\hat\gamma^\mu_k\,\hat\gamma^\nu_k\,,$$

and the constant $C(x,T,\mu,A,B)$ can be built in analogy to (82). □

Explicit construction and contractive properties of $\mathcal{W}_{f_x^\nu}$. Let us finally conclude the section with the explicit construction of f_x^ν and with the proof of (73). The following result follows form [36, Theorem 2.2].

Firstly, notice the geometric drift condition (71) implies

$$\begin{aligned}&\mathcal{L}_{\psi^n}V(z)+\mathcal{L}_{\psi^n}V(y)<0 \quad \forall (z,y)\notin B(2B_\nu/A_\nu)\,,\\ &\varepsilon\mathcal{L}_{\psi^n}V(z)+\varepsilon\mathcal{L}_{\psi^n}V(y)<-\frac{A_\nu}{2}(1\wedge 4B_\nu\varepsilon)(1+\varepsilon V(z)+\varepsilon V(y)) \quad \forall (z,y)\notin B(4B_\nu(1+A_\nu^{-1}))\,,\end{aligned} \tag{83}$$

where $B(r^2)\subseteq\mathbb{R}^{2d}$ denotes the centered ball of radius r. For later convenience let us also define the radii

$$\begin{aligned}R_1&:=\sup\{|x-y|\,:\,(x,y)\in B(2B_\nu/A_\nu)\}\,,\\ R_2&:=\sup\{|x-y|\,:\,(x,y)\in B(4B_\nu(1+A_\nu^{-1}))\}\,.\end{aligned}$$

Now, take $\varepsilon\in(0,1)$ satisfying the condition

$$(4B_\nu\varepsilon)^{-1}\ge\int_0^{R_1}\int_0^s\exp\biggl(\frac{G_\nu^\kappa-\alpha_\nu)^+}{4}(s^2-r^2)+2\,\varepsilon^{1/2}(s-r)\biggr)\mathrm{d}r\,\mathrm{d}s\,, \tag{84}$$

which is always possible since the left hand side diverges as ε vanishes, whereas the right hand side is bounded.

Finally, define

$$\begin{aligned}&f_x^\nu(r):=\int_0^{r\wedge R_2}\phi(s)g(s)\,\mathrm{d}s\,,\quad\text{with}\ \ \phi(r):=\exp\biggl(-\frac{(G_\nu^\kappa-\alpha_\nu)^+}{8}r^2-2\,\varepsilon^{1/2}\,r\biggr)\,,\\ &\Phi(r):=\int_0^r\phi(s)\,\mathrm{d}s\quad\text{and}\quad g(r):=1-\frac{\int_0^{r\wedge R_1}\frac{\Phi(s)}{\phi(s)}\mathrm{d}s}{4\int_0^{R_1}\frac{\Phi(s)}{\phi(s)}\mathrm{d}s}-\frac{\int_0^{r\wedge R_2}\frac{\Phi(s)}{\phi(s)}\mathrm{d}s}{4\int_0^{R_2}\frac{\Phi(s)}{\phi(s)}\mathrm{d}s}\,.\end{aligned}$$

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 (84) equivalently reads as $4B_\nu\varepsilon\le\xi$. 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 (75) with $C_I:=\phi(R_2)$, 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

$$\begin{aligned}(f_x^\nu)''(r)&=-\biggl(\frac{(G_\nu^\kappa-\alpha_\nu)^+}{4}r+2\,\varepsilon^{1/2}\biggr)\phi(r)g(r)-\frac{\mathbf{1}_{\{r<R_1\}}\xi^{-1}+\beta^{-1}}{4}\Phi(r)\\ &\le-\biggl(\frac{(G_\nu^\kappa-\alpha_\nu)^+}{4}r+2\,\varepsilon^{1/2}\biggr)(f_x^\nu)'(r)-\frac{\mathbf{1}_{\{r<R_1\}}\xi+\beta}{4}f_x^\nu(r)\,.\end{aligned} \tag{85}$$

We conclude with the proof of the triangle inequality and the contractive property for the distorted Wasserstein semi-distance introduced at (72)

$$\mathcal{W}_{f_x^\nu}(\cdot,\cdot) := \inf_{\pi\in\Pi(\cdot,\cdot)} \mathbb{E}_{(Y,Z)\sim\pi}\left[f_x^\nu(|Y-Z|)\,(1+\varepsilon V(Y)+\varepsilon V(Z))\right] .$$

Proof of the weak triangle inequality (74). The following proof is an adaptation of [48, 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

(86)
$$f_x^\nu(|y-z|)(1+V(y)+\varepsilon V(z)) \le C_\Delta\,\left[f_x^\nu(|y-p|)(1+\varepsilon V(y)+\varepsilon V(p))+f_x^\nu(|p-z|)(1+\varepsilon V(p)+\varepsilon V(z))\right] .$$

Firstly, notice that it holds

$$|y-z|\le R_2 \quad \Rightarrow V(y)\le \max\{2,\,1+2\,R_2^2\}\,V(z) \quad \forall\, y,z\in\mathbb{R}^d\ . \tag{87}$$

Without loss of generalities assume that $|z|\le|y|$ (and hence $V(z)\le V(y)$). Since $f_x^\nu(r)\le f_x^\nu(R_2)$, if $|y-p|\ge R_2$ then

$$\begin{aligned} f_x^\nu(|y-z|)(1+\varepsilon V(y)+\varepsilon V(z)) &\le f_x^\nu(R_2)(1+2\,\varepsilon V(y)) = f_x^\nu(|y-p|)(1+2\varepsilon V(y))\\ &\le 2\,f_x^\nu(|y-p|)(1+\varepsilon V(y)+\varepsilon V(p))\ . \end{aligned} \tag{88}$$

On the other hand, if $|y-p|\le R_2$ then (87) and the subadditivity of f_x^ν (which is guaranteed by its concavity) imply

$$\begin{aligned} &f_x^\nu(|y-z|)(1+\varepsilon V(y)+\varepsilon V(z)) \le (f_x^\nu(|y-p|)+f_x^\nu(|p-z|))(1+\varepsilon V(y)+\varepsilon V(z))\\ &\quad\le f_x^\nu(|y-p|)(1+2\varepsilon V(y))+f_x^\nu(|p-z|)(1+\max\{2,\,1+2\,R_2^2\}\,\varepsilon V(z)+\varepsilon V(z))\ . \end{aligned} \tag{89}$$

As a byproduct of (88) and (89) we get the validity of (86) (and hence of (74)) with

$$C_\Delta := \max\{3,\,2+2\,R_2^2\}\ .$$

□

Proof of the contraction (73). 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

$$\begin{cases} \mathrm{d}Z_t = -\left(\frac{Z_t-x}{2T}+\frac12\nabla\psi^n(Z_t)\right)\mathrm{d}t+\mathrm{d}B_t\ ,\\ \mathrm{d}Y_t = -\left(\frac{Y_t-x}{2T}+\frac12\nabla\psi^n(Y_t)\right)\mathrm{d}t+\mathrm{d}\hat{B}_t \quad \forall\, t\in[0,\tau) \ \text{ and } Y_t=Z_t \quad \forall\, t\ge\tau\ ,\\ (Z_0,Y_0)\sim\pi\ , \end{cases}$$

where $\tau:=\inf\{s\ge0 : Z_s=Y_s\}$, and $(\hat{B}_t)_{t\ge0}$ is defined as

$$\mathrm{d}\hat{B}_t := (\mathrm{Id}-2\,e_t\,e_t^\intercal\,\mathbf{1}_{\{t<\tau\}})\,\mathrm{d}B_t \qquad \text{where} \quad e_t:=\begin{cases}\frac{Z_t-Y_t}{|Z_t-Y_t|} & \text{when } |Z_t-Y_t|>0\,,\\ u & \text{when } |Z_t-Y_t|=0\,.\end{cases}$$

with $u\in\mathbb{R}^d$ being a fixed (arbitrary) unit-vector. By Lévy's characterization, $(\hat{B}_t)_{t\ge0}$ 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 $\mathrm{d}W_t:=e_t^\intercal\mathrm{d}B_t$ is a one-dimensional Brownian motion. By setting $r_t=|Z_t-Y_t|$ and by applying Ito formula as in the proof of Theorem 23, the trivial bound $\alpha_{\nu,n} > \alpha_\nu - T^{-1}$ (cf. Theorem 14) and the sublinearity $g^\kappa_\nu(r) \le G^\kappa_\nu\, r$ imply that

$$\begin{aligned}
\mathrm{d}f^\nu_x(r_t) &\le \Big(2\,(f^\nu_x)''(r_t) - \frac{r_t\,(f^\nu_x)'(r_t)}{2}(\alpha_{\nu,n} + T^{-1} - r_t^{-1} g^\kappa_\nu(r_t))\Big)\,\mathrm{d}t + 2\,(f^\nu_x)'(r_t)\,\mathrm{d}W_t \\
&\le \Big(2\,(f^\nu_x)''(r_t) + \frac{r_t\,(f^\nu_x)'(r_t)}{2}(G^\kappa_\nu - \alpha_\nu)^+\Big)\,\mathrm{d}t + 2\,(f^\nu_x)'(r_t)\,\mathrm{d}W_t \\
&\overset{(85)}{\le} -4\,\varepsilon^{1/2}\,(f^\nu_x)'(r_t)\mathrm{d}t - \mathbf{1}_{\{r_t<R_1\}}\frac{\xi}{2} f^\nu_x(r_t)\,\mathrm{d}t - \mathbf{1}_{\{r_t<R_2\}}\frac{\beta}{2} f^\nu_x(r_t)\,\mathrm{d}t + 2\,(f^\nu_x)'(r_t)\,\mathrm{d}W_t\,.
\end{aligned} \tag{90}$$

Next, we notice that from the geometric drift condition (71) and the definition of R_1, R_2 it follows that

(91)
$$\begin{aligned}
\mathrm{d}(1+\varepsilon V(Z_t)+\varepsilon V(Y_t)) &= \varepsilon(\mathcal{L}_{\psi^n} V(Z_t) + \mathcal{L}_{\psi^n} V(Y_t))\,\mathrm{d}t + 2\varepsilon\,\langle Z_t+Y_t,\,\mathrm{d}B_t\rangle - 4\varepsilon\,\langle Y_t,\,e_t\rangle\,\mathrm{d}W_t \\
&\le 2\varepsilon\,B_\nu\,\mathrm{d}t - \varepsilon\,A_\nu(V(Z_t)+V(Y_t))\,\mathrm{d}t + 2\varepsilon\,\langle Z_t+Y_t,\,\mathrm{d}B_t\rangle - 4\varepsilon\,\langle Y_t,\,e_t\rangle\,\mathrm{d}W_t \\
&\overset{(84)}{\le} \mathbf{1}_{\{r_t<R_1\}}\Big(\xi/2 - \varepsilon\,A_\nu(V(Z_t)+V(Y_t))\Big)\mathrm{d}t + \varepsilon\,\mathbf{1}_{\{r_t\in[R_1,R_2)\}}(\mathcal{L}_{\psi^n} V(Z_t) + \mathcal{L}_{\psi^n} V(Y_t))\,\mathrm{d}t \\
&\quad + \varepsilon\,\mathbf{1}_{\{r_t\ge R_2\}}(\mathcal{L}_{\psi^n} V(Z_t) + \mathcal{L}_{\psi^n} V(Y_t))\,\mathrm{d}t + 2\varepsilon\,\langle Z_t+Y_t,\,\mathrm{d}B_t\rangle - 4\varepsilon\,\langle Y_t,\,e_t\rangle\,\mathrm{d}W_t \\
&\overset{(83)}{\le} \Big(\mathbf{1}_{\{r_t<R_1\}}\,\xi/2 - \mathbf{1}_{\{r_t\ge R_2\}}\,\lambda\,(1+\varepsilon V(Z_t)+\varepsilon V(Y_t))\Big)\,\mathrm{d}t + 2\varepsilon\,\langle Z_t+Y_t,\,\mathrm{d}B_t\rangle - 4\varepsilon\,\langle Y_t,\,e_t\rangle\,\mathrm{d}W_t
\end{aligned}$$

where in the last step we have taken

$$\lambda := \min\{\beta,\, A_\nu,\, 4A_\nu B_\nu \varepsilon\}/2\,. \tag{92}$$

Finally, notice that the choice of coupling by reflection gives to the covariation between $f^\nu_x(r_t)$ and $1+\varepsilon V(Z_t)+\varepsilon V(Y_t)$ the following elegant expression

$$\mathrm{d}[f^\nu_x(r_\cdot),\,1+\varepsilon V(Z_\cdot)+\varepsilon V(Y_\cdot)]_t = 4\varepsilon\, r_t\,(f^\nu_x)'(r_t)\,\mathrm{d}t < 4\varepsilon^{1/2}\,(1+\varepsilon V(y)+\varepsilon V(z))\,(f^\nu_x)'(r_t)\,\mathrm{d}t\,. \tag{93}$$

where the last step follows from the trivial series of inequalities

$$\begin{aligned}
4\varepsilon\,|y-z| &\le 4\varepsilon(1+\varepsilon V(y)+\varepsilon V(z))\Big(\frac{|y|}{1+\varepsilon V(y)} + \frac{|z|}{1+\varepsilon V(z)}\Big) \\
&\le 4\varepsilon(1+\varepsilon V(y)+\varepsilon V(z))\,\sup_{y\in\mathbb{R}^d}\frac{2|y|}{1+\varepsilon V(y)} = 4(1+\varepsilon V(y)+\varepsilon V(z))\sqrt{\frac{\varepsilon}{1+\varepsilon}}\,.
\end{aligned}$$

If for sake of notation we set $F_V(z,y) := f^\nu_x(|y-z|)(1+\varepsilon V(z)+\varepsilon V(y))$ the inequalities (90), (91) and (93) gives

$$\begin{aligned}
\mathrm{d}F_V(Z_t,\,Y_t) &= (1+\varepsilon V(Z_t)+\varepsilon V(Y_t))\,\mathrm{d}f^\nu_x(r_t) + f^\nu_x(r_t)\,\mathrm{d}(1+\varepsilon V(Z_t)+\varepsilon V(Y_t)) \\
&\quad + \mathrm{d}[f^\nu_x(r_\cdot),\,1+\varepsilon V(Z_\cdot)+\varepsilon V(Y_\cdot)]_t \\
&\le -\mathbf{1}_{\{r_t<R_1\}}\frac{\xi}{2}\,(F_V(Z_t,Y_t) - f^\nu_x(r_t))\,\mathrm{d}t - \Big(\mathbf{1}_{\{r_t<R_2\}}\frac{\beta}{2} + \mathbf{1}_{\{r_t\ge R_2\}}\lambda\Big)\,F_V(Z_t,Y_t)\,\mathrm{d}t + \mathrm{d}M_t \\
&\le -\lambda\,F_V(Z_t,Y_t)\,\mathrm{d}t + \mathrm{d}M_t\,,
\end{aligned}$$

where

$$\mathrm{d}M_t := 2\,(f^\nu_x)'(r_t)(1+\varepsilon V(Z_t)+\varepsilon V(Y_t))\,\mathrm{d}W_t + 2\varepsilon\,f^\nu_x(r_t)\,\langle Z_t+Y_t,\,\mathrm{d}B_t\rangle - 4\varepsilon\,f^\nu_x(r_t)\,\langle Y_t,\,e_t\rangle\,\mathrm{d}W_t$$

is a local martingale. Hence $e^{\lambda t}F_V(Z_t,Y_t)$ is a local-supermartingale.

Now, for any $M \in \mathbb{N}$ consider now the stopping time

$$\tau_M := \inf\{t \geq 0 \colon |Z_t - Y_t| \leq M^{-1} \ \text{ or } \ |Y_t| \vee |Z_t| \geq M\} \,,$$

and notice that $T_M \uparrow +\infty$ as M grows (cf. Corollary 16). Then the previous discussion, Gronwall Lemma and Fatou Lemma give

$$e^{\lambda t}\,\mathbb{E}[F_V(Z_t, Y_t)] \leq \liminf_{M\to+\infty} \mathbb{E}[\mathbf{1}_{\{t<\tau_M\}}\, e^{\lambda t} F_V(Z_t, Y_t)] = \mathbb{E}[F_V(Z_0, Y_0)]$$

In conclusion, since $Z_t \sim \mu_1 P^n_t$ and $Y_t \sim \mu_2 P^n_t$ we have

$$\mathcal{W}_{f^\nu_x}(\mu_1\, P^n_t, \mu_2\, P^n_t) \leq \mathbb{E}[F_V(Z_t, Y_t)] \leq e^{-\lambda\, t}\, \mathbb{E}_\pi[F_V(Z_0, Y_0)] \,.$$

By minimizing the above bound over $\pi \in \Pi(\mu_1, \mu_2)$ concludes the proof of (73). □

Appendix A. Dynamical Schrödinger problem and Sinkhorn bridges

The original motivation of the Schrödinger problem comes from statistical mechanics, in particular the evolution in time of a particle system conditionally to observations. Indeed, it is usefol to record here that it can be equivalently cast as a minimization problem on $\mathcal{P}(\Omega)$, *i.e.*, the space of probability measures on the path space $\Omega := \mathcal{C}([0,T], \mathbb{R}^d)$, endowed with the uniform convergence topology and equipped with its Borel σ-field $\mathcal{B}(\Omega)$. More precisely, Sanov Theorem [33, Theorem 2.2.1] tells that SP consists in finding the solution to

$$\text{minimize } \mathscr{H}(\mathrm{P}|\mathrm{R}) \text{ under the constraint } \mathrm{P} \in \mathcal{P}(\Omega)\,, \mathrm{P}_0 = \mu \text{ and } \mathrm{P}_T = \nu\,, \tag{94}$$

where $\mathrm{R}(\mathrm{d}\omega) := \int \mathbb{P}_x(\mathrm{d}\omega)\,\mathrm{d}x$ denotes the stationary Wiener measure, with $\mathbb{P}_x$ being the canonical distribution of the standard d-dimensional Brownian motion over the time window $[0,T]$ started from $x \in \mathbb{R}^d$, while P_t denotes the marginal of P at time t, *i.e.*, $\mathrm{P}_t := (X_t)_\#\mathrm{P}$ (with $X_t(\omega) := \omega_t$ being the canonical projection map). The above minimization problem on the path space is referred to as the Schrödinger bridge problem. Since $\mathrm{P} \mapsto \mathscr{H}(\mathrm{P}|\mathrm{R})$ is strictly convex on $\mathcal{P}(\Omega)$, its optimizer (when it exists) is unique and is called the "Schrödinger bridge" and we denote it by $\mathrm{P}^\star$. It is well-known that this problem and SP are equivalent and that $\mathrm{P}^\star$ can be obtained by lifting the optimal coupling $\pi^\star_T$ [52, Sec. 2]. More precisely, owing to the additive property of relative entropy [52, Formula (72)], for any $\mathrm{P} \in \mathcal{P}(\Omega)$ it holds

$$\mathscr{H}(\mathrm{P}|\mathrm{R}) = \mathscr{H}((X_0, X_T)_\#\mathrm{P}|\mathrm{R}_{0,T}) + \int \mathscr{H}(\mathrm{P}^{xy}|\mathrm{R}^{xy})\,\mathrm{d}(X_0, X_T)_\#\mathrm{P}(x,y)\,,$$

where P^{xy} is the xy-bridge of P and the map $((x,y), A) \mapsto \mathrm{P}^{xy}(A)$ is a regular conditional distribution of P w.r.t. $(\omega_t)_{t\in[0,T]} \mapsto (\omega_0, \omega_T)$. Therefore the dynamic minimization problem (94) splits in two independent minimization problems, namely a constrained minimization problem on the coupling space, which corresponds to (SP), and an unconstrained minimization problem for the xy-bridges. Clearly the minimum value of the latter problem is equal to zero and it is attained when choosing $\mathrm{P}^{xy} = \mathrm{R}^{xy}$. In conclusion, the optimal Schrödinger bridge $\mathrm{P}^\star$ is obtained by lifting the Schrödinger plan $\pi^\star_T$ via Brownian bridges, *i.e.*,

$$\mathrm{P}^\star(\cdot) = \int_{\mathbb{R}^d\times\mathbb{R}^d} \mathrm{R}^{xy}(\cdot)\,\mathrm{d}\pi^\star_T(x,y)\,. \tag{95}$$

Particularly, the above formula allows to translate to the path space results regarding the couplings and the (static) SP.

Similarly, the Sinkhorn algorithm (7) can be lifted to the path space from Sinkhorn plans $\pi_T^{n,n}$ and $\pi_T^{n+1,n}$. Indeed using the same lifting (95), *i.e.*, we could consider the path measures

$$\mathrm{P}^{n,n}(\cdot) \coloneqq \int_{\mathbb{R}^d\times\mathbb{R}^d} \mathrm{R}^{xy}(\cdot)\,\mathrm{d}\pi^{n,n}(x,y) \quad \text{and} \quad \mathrm{P}^{n+1,n}(\cdot) \coloneqq \int_{\mathbb{R}^d\times\mathbb{R}^d} \mathrm{R}^{xy}(\cdot)\,\mathrm{d}\pi^{n+1,n}(x,y)\ ,$$

to whom we refer to as the Sinkhorn bridges. Notice that we immediately have

$$\mathrm{P}^{n+1,n} = \arg\min_{\{\mathrm{P}\colon \mathrm{P}_0=\mu\}} \mathscr{H}(\cdot|\mathrm{P}^{n,n})\ , \qquad \mathrm{P}^{n+1,n+1} = \arg\min_{\{\mathrm{P}\colon \mathrm{P}_T=\nu\}} \mathscr{H}(\cdot|\mathrm{P}^{n+1,n})\ ,$$

which is equivalent to (7) owing again to the additive property of the relative entropy since it implies

$$\mathscr{H}(\mathrm{P}^{n+1,n}|\mathrm{P}^{n,n}) = \mathscr{H}(\pi^{n+1,n}|\pi^{n,n}) \quad \text{and} \quad \mathscr{H}(\mathrm{P}^{n+1,n+1}|\mathrm{P}^{n+1,n}) = \mathscr{H}(\pi^{n+1,n+1}|\pi^{n+1,n})\ .$$

Then we may straightforwardly lift Theorem 12 to the path space and obtain the following equivalent result

Theorem 31 (Exponential convergence of Sinkhorn bridges)**.** *Assume the validity of* ***A*** *1 and* ***A*** *2 and* (19) *for some positive constants* A, $B>0$*. Then, for any* $n\geq 1$ *it holds*

$$\mathscr{H}(\mathrm{P}^{n,n}|\mathrm{P}^\star) + \mathscr{H}(\mathrm{P}^\star|\mathrm{P}^{n,n}) \leq \frac{D(A,B,\mu)}{\hat{\gamma}^\mu_{n-1}} \prod_{k=0}^{n-1} \hat{\gamma}^\mu_k\,\hat{\gamma}^\nu_k\ ,$$

$$\mathscr{H}(\mathrm{P}^{n+1,n}|\mathrm{P}^\star) + \mathscr{H}(\mathrm{P}^\star|\mathrm{P}^{n+1,n}) \leq D(A,B,\nu) \prod_{k=0}^{n-1} \hat{\gamma}^\mu_k\,\hat{\gamma}^\nu_k\ ,$$

and

$$\mathscr{H}(\mathrm{P}^{n,n}|\pi^{n,n-1}) + \mathscr{H}(\pi^{n,n-1}|\mathrm{P}^{n,n}) = \mathscr{H}(\mu^n|\mu) + \mathscr{H}(\mu|\mu^n) \leq \frac{D(A,B,\mu)}{\hat{\gamma}^\mu_{n-1}} \prod_{k=0}^{n-1} \hat{\gamma}^\mu_k\,\hat{\gamma}^\nu_k\ ,$$

$$\mathscr{H}(\mathrm{P}^{n+1,n}|\mathrm{P}^{n,n}) + \mathscr{H}(\mathrm{P}^{n,n}|\mathrm{P}^{n+1,n}) = \mathscr{H}(\nu^n|\nu) + \mathscr{H}(\nu|\nu^n) \leq D(A,B,\nu) \prod_{k=0}^{n-1} \hat{\gamma}^\mu_k\,\hat{\gamma}^\nu_k\ ,$$

where the multiplicative constants are explicitly given at (67) *and* (68)*.*

As a consequence, for T *large enough (e.g.* (21)*), for any* $\hat{\gamma}^\mu_\infty\,\hat{\gamma}^\nu_\infty<\lambda<1$*, there exists* $C\geq 0$ *such that for any* $n\in\mathbb{N}^*$ *it holds*

$$\mathscr{H}(\mathrm{P}^{n,n}|\mathrm{P}^\star) + \mathscr{H}(\mathrm{P}^\star|\mathrm{P}^{n,n}) + \mathscr{H}(\mathrm{P}^{n+1,n}|\mathrm{P}^\star) + \mathscr{H}(\mathrm{P}^\star|\mathrm{P}^{n+1,n}) \leq C\,\lambda^n\ .$$

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* (23)*.*

Appendix B. Convergence along the iterates marginals

Sinkhorn iterates may be considered as potentials of appropriate Schrödinger problems. Indeed, the decomposition given in (5) and (6) implies that

- the couple (φ^{n+1},ψ^n) corresponds to a couple of Schrödinger potentials (as defined in (1)) associated to the Schrödinger problem with reference measure $\mathrm{R}_{0,T}$ and with marginals μ and $\nu^n \coloneqq (\mathrm{proj}_y)_\#\pi^{n+1,n}$;
- the couple $(\varphi^{n+1},\psi^{n+1})$ corresponds to a couple of Schrödinger potentials (as defined in (1)) associated to the Schrödinger problem with reference measure $\mathrm{R}_{0,T}$ and with marginals $\mu^{n+1} \coloneqq (\mathrm{proj}_x)_\#\pi^{n+1,n+1}$ and ν.

This simple observation guarantees us to consider the following convergence result

Theorem 32. *Assume the validity of* **A** *1 and* **A** *2. Then for any $n \geq 1$ it holds*

$$\int |\nabla\varphi^n - \nabla\varphi^\star| \mathrm{d}\mu^n \leq \frac{\gamma_\infty^\nu}{T} \prod_{k=0}^{n-2} \frac{\gamma_k^\mu\, \gamma_k^\nu}{T^2} \int |\nabla\psi^0 - \nabla\psi^\star| \mathrm{d}\nu\,,$$

$$\int |\nabla\psi^n - \nabla\psi^\star| \mathrm{d}\nu^n \leq \frac{\gamma_\infty^\mu}{\gamma_{n-1}^\mu} \prod_{k=0}^{n-1} \frac{\gamma_k^\mu\, \gamma_k^\nu}{T^2} \int |\nabla\psi^0 - \nabla\psi^\star| \mathrm{d}\nu\,,$$

where $(\gamma_k^\mu)_{k\in\mathbb{N}}$ and $(\gamma_k^\nu)_{k\in\mathbb{N}}$ are given in Remark 35 and the above productorial $\prod_{k=0}^{n-2}$ should equal 1 if the upper index $n-2$ is negative. Particularly, as soon as $T^2 > \gamma_\infty^\mu\, \gamma_\infty^\nu$ we get the exponential convergence to zero of the above L^1-norms along the adjusted marginal.

Moreover, if we set the initial Sinkhorn iterate equal to $\psi^0 = U_\nu$ (i.e., $\varphi^0 = 0$), then

$$\|\nabla\varphi^n - \nabla\varphi^\star\|_{\mathrm{L}^1(\mu^n)} \leq \frac{\gamma_\infty^\nu}{T} \prod_{k=0}^{n-2} \frac{\gamma_k^\mu\, \gamma_k^\nu}{T^2}\, T^{-1} \left(M_1(\mu) + M_1(\nu)\right)\,,$$

$$\|\nabla\psi^n - \nabla\psi^\star\|_{\mathrm{L}^1(\nu^n)} \leq \frac{\gamma_\infty^\mu}{\gamma_{n-1}^\mu} \prod_{k=0}^{n-1} \frac{\gamma_k^\mu\, \gamma_k^\nu}{T^2}\, T^{-1} \left(M_1(\mu) + M_1(\nu)\right)\,.$$

Proof. The proof is an adaptation of the proof of Theorem 4 and boils down to showing that for all $n \geq 0$

$$\int |\nabla\varphi^{n+1} - \nabla\varphi^\star| \mathrm{d}\mu^n \leq T^{-1}\, \gamma_\infty^\nu \int |\nabla\psi^n - \nabla\psi^\star| \mathrm{d}\nu\,,$$

$$\int |\nabla\psi^{n+1} - \nabla\psi^\star| \mathrm{d}\nu^n \leq T^{-1}\, \gamma_\infty^\mu \int |\nabla\varphi^{n+1} - \nabla\varphi^\star| \mathrm{d}\mu\,,$$

and then simply rely on Theorem 4. We will show only the first bound since the proof of the second one can be obtained by meaning of the same argument.

Owing to (59) and on the conditional property of π_T^{x,ψ^n}[3], *i.e.*, $\int \pi_T^{x,\psi^n}(\mathrm{d}y)\, \mu^n(\mathrm{d}x) = \nu(\mathrm{d}y)$, it is enough showing that

$$\mathcal{W}_1(\pi_T^{x,\psi^n}, \pi_T^{x,\psi^\star}) \leq \gamma_\infty^\nu \int |\nabla\psi^n - \nabla\psi^\star|\, \mathrm{d}\pi_T^{x,\psi^n},$$

which trivially follows from Theorem 23. More precisely it is enough arguing as in Corollary 24, this time with $\mathfrak{p} = \pi_T^{x,\psi^\star}$ and $\mathfrak{q} = \pi_T^{x,\psi^n}$ (*i.e.*, exchanging the role between the marginals). Finally, when $\psi^0 = U_\nu$ we can obtain the above bounds as in Corollary 9. □

Appendix C. Explicit rates of convergence

The following result is a standard construction when dealing with coupling by reflection. The first two items come from [34, Theorem 1] whereas the monotonicity property stated in the last item has been observed in [22, Proposition 2.1].

Proposition 33. *For any $\kappa \in \mathcal{C}((0,+\infty),\mathbb{R})$ such that*

$$\liminf_{r\to+\infty} \kappa(r) > 0 \quad \textit{and} \quad \int_0^1 s\, \kappa(s)^- \,\mathrm{d}s < +\infty\,,$$

[3]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 $\pi_T^{x,\psi^n}(\mathrm{d}y)\mu^n(\mathrm{d}x) = \pi^{n,n}(\mathrm{d}x\,\mathrm{d}y)$.

there exist a strictly increasing concave function $f_\kappa \colon [0,+\infty) \to [0,+\infty)$*, a positive rate* λ_κ *and a positive constant* C_κ *such that*

(1) it holds

$$C_\kappa\, r \le f_\kappa(r) \le r \quad \text{and} \quad C_\kappa \le {f_\kappa}'(r) \le 1 \quad \forall\, r \in [0,+\infty)\ ; \tag{96}$$

(2) for any $r>0$ *the following differential inequality holds*

$$2\, {f_\kappa}''(r) - \frac{r\, {f_\kappa}'(r)}{2}\, \kappa(r) \le -\lambda_\kappa\, f(r)\ ;$$

(3) the maps $\kappa \mapsto \lambda_\kappa$ *and* $\kappa \mapsto C_\kappa$ *are monotone, i.e., for any* κ*,* $\bar\kappa$

$$\bar\kappa(r) \ge \kappa(r) \quad \forall\, r>0 \quad \Rightarrow \quad \lambda_{\bar\kappa} \ge \lambda_\kappa \text{ and } C_{\bar\kappa} \ge C_\kappa\ .$$

Proof. This is a standard result, whose proof can be found in [35]. For readers' convenience we mimic here the construction given there. Therefore let us define

$$\begin{aligned} R_0 &:= \inf\{R \ge 0 :\ \kappa(r) \ge 0 \quad \forall r \ge R\} \\ R_1 &:= \inf\{R \ge R_0 :\ \kappa(r) R (R-R_0) \ge 8 \quad \forall r \ge R\} \end{aligned} \tag{97}$$

and consider the function

$$f_\kappa(r) := \int_0^r \phi(s) g(s)\, \mathrm{d}s\ , \quad \text{with}\ \ \phi(r) := \exp\!\left(-\frac{1}{4}\int_0^r s\, \kappa(s)^-\, \mathrm{d}s\right),$$

$$\Phi(r) := \int_0^r \phi(s)\, \mathrm{d}s \quad \text{and} \quad g(r) := 1 - \frac{\int_0^{r\wedge R_1} \frac{\Phi(s)}{\phi(s)}\mathrm{d}s}{2\int_0^{R_1} \frac{\Phi(s)}{\phi(s)}\mathrm{d}s}\ ,$$

where the negative part is defined as $a^- := \max\{0, -a\}$. Finally, let us define the positive constant C_κ and the rate λ_κ as the quantities

$$C_\kappa := \frac{\phi(R_0)}{2} \quad \text{and} \quad \lambda_\kappa := \left(\int_0^{R_1} \frac{\Phi(s)}{\phi(s)}\mathrm{d}s\right)^{-1}.$$

Let us start by simply noticing that

$$\phi(R_0) \le \phi(s) \le 1\ , \quad \phi(R_0)\, r \le \Phi(r) \le r \quad \text{and} \quad \frac{1}{2} \le g(r) \le 1\ ,$$

which immediately proves the bounds for ${f_\kappa}'(r) = \phi(r) g(r)$ with $C = \phi(R_0)/2$. From the previous bound on g we immediately deduce also

$$\Phi(r)/2 \le f_\kappa(r) \le \Phi(r)\ , \tag{98}$$

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 ${f_\kappa}'(r) = \phi(r) g(r)$ and

$$\begin{aligned} {f_\kappa}''(r) &= \phi'(r) g(r) + \phi(r) g'(r) = -\frac{r}{4}\, \kappa(r)^-\, \phi(r) g(r) + \phi(r) g'(r) \\ &= -\frac{\kappa(r)^-}{4}\, r\, f'(r) + \phi(r) g'(r) \le \frac{\kappa(r)}{4}\, r\, {f_\kappa}'(r) + \phi(r) g'(r)\,. \end{aligned}$$

Indeed as a byproduct we get

$${f_\kappa}''(r) - \frac{\kappa(r)}{4}\, r\, {f_\kappa}'(r) \le \phi(r) g'(r)\,,$$

and since for any $r < R_1$ it holds $g'(r) = -\frac{\lambda_\kappa}{2}\,\Phi(r)/\phi(r)$, we deduce

$$f_\kappa{}''(r) - \frac{\kappa(r)}{4}\, r\, f_\kappa{}'(r) \le -\frac{\lambda_\kappa}{2}\,\Phi(r) \overset{(98)}{\le} -\frac{\lambda_\kappa}{2}\, f_\kappa(r) \quad \forall\, r < R_1\,.$$

At the same time, for any $r \ge R_0$ we have $\phi(r) = \phi(R_0)$ which implies

$$\Phi(r) = \Phi(R_0) + (r - R_0)\phi(R_0) \quad \forall r \ge R_0\,. \tag{99}$$

Now if we introduce $\tilde{\Phi}(r) \coloneqq \Phi(r)/r$, 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 which we deduce that for any $r \ge R_1$ the ratio function $\tilde{\Phi}$ is decreasing and therefore that

$$\frac{\Phi(r)}{r} \le \frac{\Phi(R_1)}{R_1} \quad \forall\, r \ge R_1\,. \tag{100}$$

Given this premises, for any $r \ge R_1$ it holds

$$\phi(r) = \phi(R_0)\,, \quad g(r) = \frac{1}{2} \quad \text{and} \quad \kappa(r)\, R_1(R_1 - R_0) \ge 8$$

(cf. Definition (97)), which implies that $f_\kappa{}'(r)$ is constantly equal to $\frac{\phi(R_0)}{2}$ for any $r \ge R_1$. Therefore, for any $r \ge R_1$ we have

$$\begin{aligned} f_\kappa{}''(r) - \frac{\kappa(r)}{4}\, r\, f_\kappa{}'(r) &= -\frac{\kappa(r)}{4}\, r\, \frac{\phi(R_0)}{2} = -\frac{\kappa(r)}{8}\, r\, \phi(R_0) \\ \overset{(97)}{\le} -\frac{r\,\phi(R_0)}{R_1(R_1 - R_0)} &\overset{(100)}{\le} -\frac{\Phi(r)}{\Phi(R_1)}\frac{\phi(R_0)}{R_1 - R_0} \overset{(\dagger)}{\le} -\frac{\lambda_\kappa}{2}\,\Phi(r) \overset{(98)}{\le} -\frac{\lambda_\kappa}{2}\, f(r) \end{aligned}$$

where inequality (†) follows from the observation that $\phi(s) = \phi(R_0)$ for any $r \ge R_0$ and that

$$\begin{aligned} \lambda_\kappa^{-1} &= \int_0^{R_1} \Phi(s)/\phi(s)\,\mathrm{d}s \ge \int_{R_0}^{R_1} \Phi(s)/\phi(s)\,\mathrm{d}s \overset{(99)}{=} \int_{R_0}^{R_1} \frac{\Phi(R_0) + (s - R_0)\phi(R_0)}{\phi(R_0)}\,\mathrm{d}s \\ &= \frac{\Phi(R_0}{\phi(R_0)}(R_1 - R_0) + \frac{(R_1 - R_0)^2}{2} = \frac{(R_1 - R_0)}{2\,\phi(R_0)}\,(2\Phi(R_0) + (R_1 - R_0)\phi(R_0)) \\ &\overset{(99)}{=} \frac{(R_1 - R_0)}{2\,\phi(R_0)}\,(\Phi(R_0) + \Phi(R_1)) \ge \frac{(R_1 - R_0)}{2\,\phi(R_0)}\,\Phi(R_1)\,. \end{aligned}$$

The proof of the monotonicity in item *(iii)* follows from the definition of C and from the observation that we may rewrite the definition of λ_κ as

$$\lambda_\kappa = \left(\int_0^{R_1}\int_0^s \exp\left(\frac{1}{4}\int_t^s u\,\kappa(u)^-\,\mathrm{d}u\right)\mathrm{d}t\;\mathrm{d}s\right)^{-1}\,,$$

which clearly implies the desired monotonicity. □

Corollary 34. *If f, λ, C are defined as in Proposition 33 and if we consider the Wasserstein distance*

$$\mathcal{W}_f(\cdot,\cdot) \coloneqq \inf_{\pi\in\Pi(\cdot,\cdot)} \mathbb{E}_{(X,Y)\sim\pi}\big[f(|X - Y|)\big]\,.$$

induced by the concave function f, then it is equivalent to the standard 1-Wasserstein, i.e.

$$C\,\mathcal{W}_1(\cdot,\cdot) \le \mathcal{W}_f(\cdot,\cdot) \le \mathcal{W}_1(\cdot,\cdot)\,,$$

Proof. The fact that $\mathcal{W}_f$ is a distance follows from the fact that $f(0)=0$, f is strictly increasing, concave and hence also subadditive (which implies the triangular inequality). The equivalence with the usual 1-Wasserstein distance follows from (96). □

Given the above explicit construction, we are finally ready to give explicit expressions for the convergence rates presented in the Introduction.

Remark 35 (On the exponential convergence rates in Theorem 4 and Theorem 32)**.** *The explicit formulas for the convergence rates of Theorem 4 and Theorem 32 is*

$$\gamma_n^\mu \coloneqq (2\, C_n^\mu\, \lambda_n^\mu)^{-1} \quad \text{and} \quad \gamma_n^\nu \coloneqq (2\, C_n^\nu\, \lambda_n^\nu)^{-1} \ ,$$

where λ_n^ν, C_n^ν are the constants associated to the concave function f_n^ν built in the Proposition 33 when considering $\kappa_n^\nu(r) = \alpha_{\nu,n} + T^{-1} - r^{-1}\, g_\nu^\kappa(r)$, which comes from the lower bound (31), given in Theorem 14, for the integrated convexity profile κ_{ψ^n}.

Similarly λ_n^μ, C_n^μ are the constants associated to f_n^μ built in the Proposition 33 when considering $\kappa_n^\mu(r) = \alpha_{n+1}^\mu + T^{-1} - r^{-1}\, g_\mu^\kappa(r)$, which comes from the lower bound (34), given in Theorem 14, for the integrated convexity profile $\kappa_{\varphi^{n+1}}$.

The asymptotic rates

$$\gamma_\infty^\mu \coloneqq (2\, C_\infty^\mu\, \lambda_\infty^\mu)^{-1} \quad \text{and} \quad \gamma_\infty^\nu \coloneqq (2\, C_\infty^\nu\, \lambda_\infty^\nu)^{-1} \ ,$$

are instead defined when considering the concave functions f_∞^μ and f_∞^ν associated respectively to the choices of moduli $\kappa_\infty^\mu(r) = \alpha_{\varphi^\star} + T^{-1} - r^{-1}\, g_\mu^\kappa(r)$ and $\kappa_\infty^\nu(r) = \alpha_{\psi^\star} + T^{-1} - r^{-1}\, g_\mu^\kappa(r)$.

We conclude this remark by discussing the monotonicity property of the rates sequences $(\gamma_n^\mu)_{n\in\mathbb{N}}$ and $(\gamma_n^\mu)_{n\in\mathbb{N}}$. This follows from the monotonicity of the sequences $(\alpha_{\mu,n})_{n\in\mathbb{N}}$ and $(\alpha_{\nu,n})_{n\in\mathbb{N}}$ introduced in Theorem 14. Indeed we have proven there that $(\alpha_{\mu,n})_{n\in\mathbb{N}}$ is an increasing sequence converging to $\alpha_{\varphi^\star}$, which implies that the same holds for the sequences $(\kappa_n^\mu(r))_{n\in\mathbb{N}}$, which converges to $\kappa_\infty^\mu(r)$, for any fixed $r>0$. Hence the third item of the previous theorem implies that $(\lambda_n^\mu)_{n\in\mathbb{N}}$ and $(C_n^\mu)_{n\in\mathbb{N}}$ are monotone increasing sequences and therefore $(\gamma_n^\mu)_{n\in\mathbb{N}}$ is monotone decreasing. Moreover, from the definition of λ_κ it immediately follows also that $C_n^\mu \uparrow C_\infty^\mu$, $\lambda_n^\mu \uparrow \lambda_\infty^\mu$ and hence $\gamma_n^\mu \downarrow \gamma_\infty^\mu$.

The same reasoning shows that $\kappa_n^\nu \uparrow \kappa_\infty^\nu$ and hence $C_n^\nu \uparrow C_\infty^\nu$, $\lambda_n^\nu \uparrow \lambda_\infty^\nu$, which implies $\gamma_n^\nu \downarrow \gamma_\infty^\nu$.

C.1. **Explicit rates for marginals strictly log-concave.** In this section we prove the statement of Examples 6, 8 and 10. Therefore assume the validity of **A**1 and **A**2 with $g_\mu^\kappa = g_\nu^\kappa = g_\mu^\ell = g_\nu^\ell \equiv 0$. Let us firslty observe that Theorem 14 in this particular setting simply reads as

Theorem 36. *Assume the validity of **A** 1 and **A** 2 with $g_\mu^\kappa = g_\nu^\kappa = g_\mu^\ell = g_\nu^\ell \equiv 0$. 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 $n \geq 1$ for any $n\in\mathbb{N}$ it holds $r>0$ it holds*

$$\kappa_{\varphi^n}(r) \geq \alpha_{\mu,n} \quad \text{and} \quad \kappa_{\psi^n}(r) \geq \alpha_{\nu,n} \ .$$

These two sequences are defined as

$$\begin{cases} \alpha_{\mu,0} \coloneqq \alpha_\mu - T^{-1} \ , \\ \alpha_{\mu,n+1} \coloneqq \alpha_\mu - T^{-1} + \Big(T^2\, \beta_\nu + (\alpha_{\mu,n} + T^{-1})^{-1} \Big)^{-1} \ , \quad n\in\mathbb{N} \ , \end{cases}$$

and

$$\begin{cases}\alpha_{\nu,0} \coloneqq \alpha_\nu - T^{-1}\,, \\ \alpha_{\nu,n+1} \coloneqq \alpha_\nu - T^{-1} + \left(T^2\,\beta_\mu + (\alpha_{\nu,n} + T^{-1})^{-1}\right)^{-1}\,, \quad n\in\mathbb{N}\,.\end{cases}$$

Moreover, both sequences converge respectively to

$$\alpha_{\varphi^\star} \coloneqq \frac{1}{2}\left(\alpha_\mu + \sqrt{\alpha_\mu^2 + 4\alpha_\mu/(T^2\beta_\nu)}\right) - T^{-1} \quad \text{and} \quad \alpha_{\psi^\star} \coloneqq \frac{1}{2}\left(\alpha_\nu + \sqrt{\alpha_\nu^2 + 4\alpha_\nu/(T^2\beta_\mu)}\right) - T^{-1}\,, \tag{101}$$

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 (1).

Proof. This is a particular instance of Theorem 14 when $g^\kappa_\mu = g^\kappa_\nu = g^\ell_\mu = g^\ell_\nu \equiv 0$. 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 (101). We will only prove the first one since the second identity can be proven in the same way. From Theorem 14 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 $\theta^\mu_n \coloneqq \alpha_{\mu,n} + T^{-1}$. Clearly $\theta^\mu_n > 0$, $\theta^\mu_n \uparrow \theta^\mu_\infty \coloneqq \alpha_{\varphi^\star} + T^{-1}$ and the latter limit value can be seen as a fixed point for the iteration

$$\theta^\mu_{n+1} = \alpha_\nu + (T^2\,\beta_\nu + (\theta^\mu_n)^{-1})^{-1}\,.$$

A straightforward computation shows that the there are just two possible fixed point solutions, namely

$$\frac{1}{2}\left(\alpha_\mu - \sqrt{\alpha_\mu^2 + 4\alpha_\mu/(T^2\beta_\nu)}\right) \text{ and } \frac{1}{2}\left(\alpha_\mu + \sqrt{\alpha_\mu^2 + 4\alpha_\mu/(T^2\beta_\nu)}\right)\,.$$

Since one solution is negative, whereas $(\theta^\mu_n)_{n\in\mathbb{N}}$ is a positive increasing sequence, we immediately deduce that θ^μ_∞ equals the largest (and positive) fixed point. This proves (101). □

From the previous result we immediately deduce the identities appearing in Examples 6, 8 and 10. Indeed Theorem 36 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 obtain the validity of Corollary 24 this time with the rates appearing in Theorem 36, *i.e.*,

$$\mathcal{W}_1(\pi_T^{x,\psi^n}, \pi_T^{x,\psi^\star}) \leq \gamma^\nu_n \int |\nabla\psi^n - \nabla\psi^\star|\,\mathrm{d}\pi_T^{x,\psi^\star} \quad \text{with } \gamma^\nu_n = (\theta^\nu_n)^{-1} = (\alpha_{\nu,n} + T^{-1})^{-1}\,,$$

and

$$\mathcal{W}_1(\pi_T^{x,\varphi^{n+1}}, \pi_T^{x,\varphi^\star}) \leq \gamma^\mu_n \int \left|\nabla\varphi^{n+1} - \nabla\varphi^\star\right|\mathrm{d}\pi_T^{x,\varphi^\star} \quad \text{with } \gamma^\mu_n = (\theta^\mu_n)^{-1} = (\alpha_{\mu,n} + T^{-1})^{-1}\,.$$

Finally, the lower-bounds in provided (18) can be computed by solving $T > \theta\,\gamma^\mu_\infty$ and $T > \theta^{-1}\,\gamma^\nu_\infty$, whereas the sufficient condition (25) for the exponential convergence is derived from (24) by imposing $\hat\gamma^\mu_\infty < 1$ and $\hat\gamma^\nu_\infty < 1$.

Appendix D. Technical results and moment bounds

Lemma 37 (Exponential integrability of the marginals)**.** *Assume* **A** 2*-(i). Then for any* $\sigma_\nu \in (0,\alpha_\nu/2)$ *it holds* $\exp(\sigma_\nu|x|^2)\in \mathrm{L}^1(\nu)$*. Similarly if* **A** 2*-(ii) holds true, then for any* $\sigma_\mu\in(0,\alpha_\mu/2)$ *it holds* $\exp(\sigma_\mu|x|^2)\in\mathrm{L}^1(\mu)$.

Proof. We will prove only the first claim since the second one can be proven in the same way. It is enough noticing that for any $x\in\mathbb{R}^d$ it holds

$$\begin{aligned}\langle\nabla U_\nu(x),\, x\rangle &= \langle \nabla U_\nu(x)-\nabla U_\nu(0),\, x\rangle + \langle\nabla U_\nu(0),\, x\rangle \geq \kappa_{U_\nu}(|x|)\,|x|^2 - |\nabla U_\nu(0)|\,|x|\\ &\geq \alpha_\nu\,|x|^2 - (g^\kappa_\nu(|x|)+|\nabla U_\nu(0)|)|x| \geq \alpha_\nu\,|x|^2 - \bar G_\nu\,|x|\end{aligned}$$

where above we have set $\bar G_\nu = \|g^\kappa_\nu\|_\infty + |\nabla U_\nu(0)|$. Therefore for any $x\in\mathbb{R}^d$ it holds

$$U_\nu(x) = U_\nu(0) + \int_0^1 \langle\nabla U_\nu(tx), x\rangle\,\mathrm{d}t \geq U_\nu(0) + \int_0^1 (\alpha_\nu\, t\,|x|^2 - \bar G_\nu\,|x|)\,\mathrm{d}t = U_\nu(0) + \frac{\alpha_\nu}{2}\,|x|^2 - \bar G_\nu\,|x|\ .$$

Finally we may deduce for any $\sigma_\nu\in(0,\alpha_\nu/2)$

$$\begin{aligned}\|\exp(\sigma_\nu|x|^2)\|_{\mathrm{L}^1(\nu)} &= \int_{\{|x|\leq R\}} \exp(\sigma_\nu|x|^2)\mathrm{d}\nu(x) + \int_{\{|x|>R\}} \exp(\sigma_\nu|x|^2)\mathrm{d}\nu(x)\\ &\leq e^{\sigma_\nu|R|^2}\,\nu(\{|x|\leq R\}) + e^{-U_\nu(0)}\int_{\{|x|>R\}} \exp(-(\alpha_\nu/2-\sigma_\nu)|x|^2 + \bar G_\nu|x|)\,\mathrm{d}x\\ &\leq e^{\sigma_\nu|R|^2}\,\nu(\{|x|\leq R\}) + e^{-U_\nu(0)}\int_{\{|x|>R\}} \exp(-(\alpha_\nu/4-\sigma_\nu/2)|x|^2)\,\mathrm{d}x < +\infty\ ,\end{aligned}$$

where above we have set $R \coloneqq 2\bar G_\nu\,(\alpha_\nu/2-\sigma_\nu)^{-1}$. □

Proof of Proposition 3. 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 [23, Proposition 5.2], once we have noticed that **A** 2-(i) and Lemma 37 guarantees the validity of $\exp(\sigma_\nu|x|^2)\in\mathrm{L}^1(\nu)$ for some positive $\sigma_\nu>0$. Therefore from (4) 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

$$\int_{\mathbb{R}^d}\exp\left(\bar\sigma|y|^2 - \psi^n(y) - \frac{|\bar x-y|^2}{2T}\right)\mathrm{d}y < +\infty \quad \forall\, x\in\mathbb{R}^d\ . \tag{102}$$

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 (10) for $h=\psi^n$, *i.e.*

$$\nabla\log P_T\exp(-\psi^n)(x)=T^{-1}\int(y-x)\,\pi_T^{x,\psi^n}(\mathrm{d}y)=-\frac{x}{T}+\frac{1}{T}\frac{\int y\exp(-\psi^n(y)-\frac{|x-y|^2}{2T})\mathrm{d}y}{\int\exp(-\psi^n(y)-\frac{|x-y|^2}{2T})\mathrm{d}y}\,.$$

The bound (102) 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 13, under the geometric-drift condition (38) obtained in the proof of Corollary 16. More precisely if for any even $p\geq 2$ we set $V_p(y)=1+|y|^p$, similarly to what happened in (71), Corollary 16 and (38), the pointwise convergence of the gradients of Theorem 7 and the convergences $\alpha_{\mu,n}\uparrow\alpha_{\varphi^\star}$, $\alpha_{\nu,n}\uparrow\alpha_{\psi^\star}$ (stated in Theorem 14) 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

$$\mathcal{L}_{\psi^\star}V_p(y)\ \vee\ \mathcal{L}_{\psi^n}V_p(y)\ \leq B_{\nu,p}-A_{\nu,p}\,V_p(y)\ \text{ and }\ \mathcal{L}_{\varphi^\star}V_p(y)\ \vee\ \mathcal{L}_{\varphi^n}V_p(y)\ \leq B_{\mu,p}-A_{\mu,p}\,V_p(y)\ ,$$

where $\mathcal{L}_h\coloneqq\Delta/2-\frac{1}{2}\langle T^{-1}(y-x)+\nabla h(y),\,\nabla\rangle$ is the generator associated to (12). We will bound the moments appearing in the proof of Theorem 13 in terms of the above constants $A_{\mu,p},\,A_{\nu,p}>0$ and $B_{\mu,p},\,B_{\nu,p}$.

Lemma 38. *Take $p\geq 2$ and set $V_p(y)=1+|y|^p$. Let $(Y_t^\star)_{t\geq0}$ and $(Z_t)_{t\geq 0}$ be defined as in (78) in the proof of Theorem 13. 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

$$\mathrm{d}V_p(Y_t^\star)=\mathcal{L}_{\psi^\star}V_p(Y_t^\star)\,\mathrm{d}t+4|Y_t^\star|^2\langle Y_t^\star,\,\mathrm{d}B_t\rangle\leq-A_{\nu,p}V_p(Y_t^\star)\,\mathrm{d}t+B_{\nu,p}\,\mathrm{d}t+p\,|Y_t^\star|^{p-2}\langle Y_t^\star,\,\mathrm{d}B_t\rangle\,.$$

Therefore, up to considering a stopping time as already detailed in the proof of the contraction (73), by taking expectation and integrating over time, from the stationarity of the process $Y_t^\star\sim\pi_T^{x,\psi^\star}$ we deduce

$$\mathbb{E}[V_p(Y_t^\star)]\leq B_{\nu,p}/A_{\nu,p}\quad\forall\,t\geq 0\,. \tag{103}$$

Similarly, when considering $h=\psi^n$ we get

$$\mathrm{d}V_p(Z_t)=\mathcal{L}_{\psi^\star}V_p(Z_t)\,\mathrm{d}t+4|Z_t|^2\langle Z_t,\,\mathrm{d}B_t\rangle\leq-A_{\nu,p}V_p(Z_t)\,\mathrm{d}t+B_{\nu,p}\,\mathrm{d}t+p\,|Z_t|^{p-2}\langle Z_t,\,\mathrm{d}B_t\rangle\,.$$

Up to considering again a stopping time, taking expectation and integrating over time yields 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 (103), 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 (81), (62) and Lemma 38 (for $p=2,\,4$) we finally conclude that

Corollary 39. *Let $(Y_t^\star)_{t\geq 0}$ and $(Z_t)_{t\geq 0}$ be defined as in (78) in the proof of Theorem 13. Then (81) can be bounded as*

$$\int_0^t \mathbb{E}\Big[(A|Y_s^\star| + B)(1+\varepsilon V_2(Y_t^\star) + \varepsilon V_2(Z_t))\Big]\,\mathrm{d}s \leq U(t,\nu,x,A,B,T)\ ,$$

with

(104)

$$U(t,\nu,x,A,B,T) \coloneqq Bt\bigg(1+\varepsilon\frac{B_{\nu,2}}{A_{\nu,2}}(1+e^{t\,A_{\nu,2}})+\varepsilon B_{\nu,2}\,t\,e^{t\,A_{\nu,2}}\bigg)$$
$$+\frac{A\,t}{\alpha_{\psi^\star}+T^{-1}}\,(T^{-1}|x|+1+\|g_\nu^\kappa\|_\infty+|\nabla\psi^\star(0)|)+A\varepsilon\,t\bigg(\frac{B_{\nu,2}}{A_{\nu,2}}+\frac{B_{\nu,4}}{A_{\nu,4}}(1+e^{t\,A_{\nu,4}})+B_{\nu,4}\,t\,e^{t\,A_{\nu,4}}\bigg)\,.$$

References

[1] 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.

[2] Jonas Adler, Axel Ringh, Ozan Öktem, and Johan Karlsson. Learning to solve inverse problems using Wasserstein loss. *arXiv preprint arXiv:1710.10898*, 2017.

[3] Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein generative adversarial networks. In *International conference on machine learning*, pages 214–223. PMLR, 2017.

[4] 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.

[5] Jean-David Benamou. Optimal transportation, modelling and numerical simulation. *Acta Numerica*, 30:249–325, 2021.

[6] 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.

[7] Jean-David Benamou, Guillaume Carlier, Simone Di Marino, and Luca Nenna. An entropy minimization approach to second-order variational mean-field games. *Mathematical Models and Methods in Applied Sciences*, 29(08):1553–1583, 2019.

[8] Espen Bernton, Promit Ghosal, and Marcel Nutz. Entropic Optimal Transport: Geometry and Large Deviations. *Duke Mathematical Journal*, 171(16):3363–3400, 2022.

[9] Garrett Birkhoff. Extensions of Jentzsch's theorem. *Transactions of the American Mathematical Society*, 85(1):219–227, 1957.

[10] 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.

[11] 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.

[12] Guillaume Carlier. On the Linear Convergence of the Multimarginal Sinkhorn Algorithm. *SIAM Journal on Optimization*, 32(2):786–794, 2022.

[13] Guillaume Carlier and Maxime Laborde. A Differential Approach to the Multi-Marginal Schrödinger System. *SIAM Journal on Mathematical Analysis*, 52(1):709–717, 2020.

[14] Guillaume Carlier, Paul Pegon, and Luca Tamanini. Convergence rate of general entropic optimal transport costs. *Preprint, arXiv:2206.03347*, 2022.

[15] 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.

[16] 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.

[17] Yongxin Chen, Tryphon T Georgiou, and Michele Pavon. Stochastic Control Liaisons: Richard Sinkhorn Meets Gaspard Monge on a Schrodinger Bridge. *SIAM Review*, 63(2):249–313, 2021.
[18] Sinho Chewi and Aram-Alexandre Pooladian. An entropic generalization of Caffarelli's contraction theorem via covariance inequalities. *arXiv preprint arXiv:2203.04954*, 2022.
[19] Alberto Chiarini, Giovanni Conforti, Giacomo Greco, and Luca Tamanini. Gradient estimates for the Schrödinger potentials: convergence to the Brenier map and quantitative stability. *arXiv preprint arXiv:2207.14262*, 2022.
[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, 2020.
[21] 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(1-2):1–47, 2019.
[22] Giovanni Conforti. Coupling by reflection for controlled diffusion processes: turnpike property and large time behavior of Hamilton Jacobi Bellman equations. *arXiv preprint arXiv:2206.04009*, 2022.
[23] Giovanni Conforti. Quadratic Schrodigner problem: second order estimates for potentials and logarithmic Sobolev inequality for the optimal couplings. *arXiv preprint arXiv:2301.00083*, 2022.
[24] Giovanni Conforti and Luigia Ripani. Around the entropic Talagrand inequality. *Bernoulli*, 26(2):1431–1452, 2020.
[25] 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.
[26] Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in Neural Information Processing Systems*, pages 2292–2300, 2013.
[27] Valentin De Bortoli, Emile Mathieu, Michael Hutchinson, James Thornton, Yee Whye Teh, and Arnaud Doucet. Riemannian score-based generative modeling. *arXiv preprint arXiv:2202.02763*, 2022.
[28] 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.
[29] Eustasio del Barrio, Alberto Gonzalez-Sanz, Jean-Michel Loubes, and Jonathan Niles-Weed. An improved central limit theorem and fast convergence rates for entropic transportation costs. *arXiv preprint arXiv:2204.09105*, 2022.
[30] George Deligiannidis, Valentin De Bortoli, and Arnaud Doucet. Quantitative uniform stability of the iterative proportional fitting procedure. *arXiv preprint arXiv:2108.08129*, 2021.
[31] 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.
[32] Ayelet Dominitz and Allen Tannenbaum. Texture mapping via optimal mass transport. *IEEE transactions on visualization and computer graphics*, 16(3):419–433, 2009.
[33] P. Dupuis and R.S. Ellis. *A Weak Convergence Approach to the Theory of Large Deviations*. Wiley series in Probability and Statistics. John Wiley & sons, 1997.
[34] Andreas Eberle. Reflection couplings and contraction rates for diffusions. *Probability Theory and Related Fields*, 166(3-4):851–886, 2016.
[35] Andreas Eberle, Arnaud Guillin, and Raphael Zimmer. Couplings and quantitative contraction rates for Langevin dynamics. *The Annals of Probability*, 47(4):1982–2010, 2019.
[36] 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.
[37] 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.
[38] Matthias Erbar, Jan Maas, and Michiel Renger. From large deviations to Wasserstein gradient flows in multiple dimensions. *Electronic Communications in Probability*, 20:12 pp., 2015.
[39] Max Fathi, Nathael Gozlan, and Maxime Prodhomme. A proof of the Caffarelli contraction theorem via entropic regularization. *Calculus of Variations and Partial Differential Equations*, 59(96), 2020.
[40] Joel Franklin and Jens Lorenz. On the scaling of multidimensional matrices. *Linear Algebra and its Applications*, 114-115:717–735, 1989.
[41] Aude Genevay, Gabriel Peyré, and Marco Cuturi. Learning generative models with sinkhorn divergences. In *International Conference on Artificial Intelligence and Statistics*, pages 1608–1617. PMLR, 2018.
[42] Ivan Gentil, Christian Léonard, and Luigia Ripani. Dynamical aspects of the generalized Schrödinger problem via Otto calculus–A heuristic point of view. *Revista Matemática Iberoamericana*, 36(4):1071–1112, 2020.

[43] 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.
[44] Promit Ghosal and Marcel Nutz. On the Convergence Rate of Sinkhorn's Algorithm. *arxiv preprint arXiv:2212.06000*, 2022.
[45] 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.
[46] 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.
[47] Giacomo Greco, Maxence Noble, Giovanni Conforti, and Alain Durmus. Non-asymptotic convergence bounds for Sinkhorn iterates and their gradients: a coupling approach. *In preparation*.
[48] M. Hairer, J. C. Mattingly, and M. 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.
[49] Soheil Kolouri, Se Rim Park, Matthew Thorpe, Dejan Slepcev, and Gustavo K Rohde. Optimal mass transport: Signal processing and machine-learning applications. *IEEE signal processing magazine*, 34(4):43–59, 2017.
[50] Flavien Léger. A gradient descent perspective on Sinkhorn. *Applied Mathematics & Optimization*, 84(2):1843–1855, 2021.
[51] C. Léonard. From the Schrödinger problem to the Monge–Kantorovich problem. *Journal of Functional Analysis*, 262(4):1879–1920, 2012.
[52] 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.
[53] J.C. Mattingly, A.M. Stuart, and D.J. Higham. Ergodicity for SDEs and approximations: locally Lipschitz vector fields and degenerate noise. *Stochastic Processes and their Applications*, 102(2):185–232, 2002.
[54] Gonzalo Mena and Jonathan Niles-Weed. Statistical bounds for entropic optimal transport: sample complexity and the central limit theorem. *Advances in Neural Information Processing Systems*, 32, 2019.
[55] 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.
[56] Toshio Mikami. Monge's problem with a quadratic cost by the zero-noise limit of h-path processes. *Probability Theory and Related Fields*, 129(2):245–260, 2004.
[57] Toshio Mikami and Michèle Thieullen. Duality theorem for the stochastic optimal control problem. *Stochastic processes and their applications*, 116(12):1815–1835, 2006.
[58] Léonard Monsaingeon, Luca Tamanini, and Dmitry Vorotnikov. The dynamical Schrödinger problem in abstract metric spaces. *Preprint, arXiv:2012.12005*, 2020.
[59] Marcel Nutz. Introduction to entropic optimal transport. *Lecture notes, Columbia University*, 2021.
[60] Marcel Nutz and Johannes Wiesel. Entropic optimal transport: convergence of potentials. *Probability Theory and Related Fields*, 184(1):401–424, 2022.
[61] Marcel Nutz and Johannes Wiesel. Stability of Schrödinger Potentials and Convergence of Sinkhorn's Algorithm. *arxiv preprint arxiv:2201.10059*, 2022.
[62] Soumik Pal. On the difference between entropic cost and the optimal transport cost. *arXiv preprint arXiv:1905.12206*, 2019.
[63] Soumik Pal and Ting-Kam Leonard Wong. Multiplicative Schrödinger problem and the Dirichlet transport. *Probability Theory and Related Fields*, 178(1-2):613–654, 2020.
[64] Gabriel Peyré. Entropic Approximation of Wasserstein Gradient Flows. *SIAM Journal on Imaging Sciences*, 8(4):2323–2351, 2015.
[65] Gabriel Peyré and Marco Cuturi. Computational Optimal Transport. *Foundations and Trends in Machine Learning*, 11(5-6):355–607, 2019.
[66] Aram-Alexandre Pooladian and Jonathan Niles-Weed. Entropic estimation of optimal transport maps. *Preprint, arXiv:2109.12004*, 2021.
[67] Aram-Alexandre Pooladian and Jonathan Niles-Weed. Entropic estimation of optimal transport maps. *arXiv preprint arXiv:2109.12004*, 2021.
[68] G. O. Roberts and R. L. Tweedie. Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli*, 4(2):341–363, 1996.
[69] Ludger Ruschendorf. Convergence of the iterative proportional fitting procedure. *The Annals of Statistics*, pages 1160–1174, 1995.
[70] Erwin Schrödinger. La théorie relativiste de l'électron et l' interprétation de la mécanique quantique. *Ann. Inst Henri Poincaré*, (2):269 – 310, 1932.

[71] Richard Sinkhorn. A Relationship Between Arbitrary Positive Matrices and Doubly Stochastic Matrices. *The Annals of Mathematical Statistics*, 35(2):876–879, 1964.
[72] Richard Sinkhorn and Paul Knopp. Concerning nonnegative matrices and doubly stochastic matrices. *Pacific Journal of Mathematics*, 21(2):343–348, 1967.
[73] Justin Solomon, Fernando De Goes, Gabriel Peyré, Marco Cuturi, Adrian Butscher, Andy Nguyen, Tao Du, and Leonidas Guibas. Convolutional wasserstein distances: Efficient optimal transportation on geometric domains. *ACM Transactions on Graphics (ToG)*, 34(4):1–11, 2015.
[74] G. Udny Yule. On the Methods of Measuring Association Between Two Attributes. *Journal of the Royal Statistical Society*, 75(6):579–652, 1912.

École Polytechnique
Current address: Département de Mathématiques Appliqués, Palaiseau, France.
Email address: giovanni.conforti@polytechnique.edu

École Polytechnique
Current address: Département de Mathématiques Appliqués, Palaiseau, France.
Email address: alain.durmus@polytechnique.edu

Eindhoven University of Technology
Current address: Department of Mathematics and Computer Science, 5600MB Eindhoven, Netherlands.
Email address: g.greco@tue.nl