
BERLIN SUMMER SCHOOL ON APPLIED STOCHASTIC ANALYSIS

Berlin, 28–31 July 2026

Lecture Notes

An Introduction to Schrödinger Bridges and Sinkhorn Semigroups

Pierre Del Moral

Inria Bordeaux – Sud-Ouest



Scientific Organising Committee

Peter Bank, Jannis Dause, Jakob Kellermann, Martin Nicolaus,
Gemma Sedrakjan, Wilhelm Stannat, Roman Wüthrich

Lecture notes reproducing an article with Ajay Jasra to appear in the Journal of Theoretical Probability. Preprint available at <https://arxiv.org/abs/2601.12633> (January 2026)

Introduction

These lecture notes were prepared for the Berlin Summer School on Applied Stochastic Analysis, held in Berlin on 28–31 July 2026. They reproduce, with only minor editorial additions, an article to appear in the *Journal of Theoretical Probability*; the corresponding preprint is available at <https://arxiv.org/abs/2601.12633>.

Entropic optimal transport problems play an increasingly important role in machine learning and generative modelling. In contrast with optimal transport maps which often have limited applicability in high dimensions, Schrödinger bridges can be solved using the celebrated Sinkhorn’s algorithm, a.k.a. the iterative proportional fitting procedure. The stability properties of Sinkhorn bridges when the number of iterations tends to infinity is a very active research area in applied probability and machine learning. Traditional proofs of convergence are mainly based on nonlinear versions of Perron-Frobenius theory and related Hilbert projective metric techniques, gradient descent, Bregman divergence techniques and Hamilton-Jacobi-Bellman equations, including propagation of convexity profiles based on coupling diffusions by reflection methods. The objective of this review article is to present, in a self-contained manner, recently developed Sinkhorn/Gibbs-type semigroup analysis based upon contraction coefficients and Lyapunov-type operator-theoretic techniques. These powerful, off-the-shelf semigroup methods are based upon transportation cost inequalities (e.g. log-Sobolev, Talagrand quadratic inequality, curvature estimates), ϕ -divergences, Kantorovich-type criteria and Dobrushin contraction-type coefficients on weighted Banach spaces as well as Wasserstein distances. This novel semigroup analysis allows one to unify and simplify many arguments in the stability of Sinkhorn algorithm. It also yields new contraction estimates w.r.t. generalized ϕ -entropies, as well as weighted total variation norms, Kantorovich criteria and Wasserstein distances

These notes are addressed primarily to doctoral students and early-career researchers in probability, statistics, and machine learning who wish to enter this area with a working, proof-oriented understanding of why and how Sinkhorn’s algorithm converges, rather than a purely algorithmic acquaintance with it. We hope they will also be of use to more experienced researchers looking for a unified reference to the semigroup approach.

We are grateful to the organisers of the Berlin Summer School on Applied Stochastic Analysis for the invitation to lecture on this material, and to the participants for their questions and comments, which have helped clarify several points in the exposition. Remaining errors and imprecisions are, of course, our own.

Contents

1	Entropic optimal transport	5
1.1	Schrödinger and Sinkhorn bridges	5
1.2	Sinkhorn semigroups	7
1.3	Literature review	9
1.4	Article focus and structure	11
1.5	Some basic notation	11
2	Some background on bridges	14
2.1	Overview	14
2.2	Conjugate and dual transitions	15
2.3	State space models	16
2.4	Dynamic and static bridges	17
2.5	Continuous vs discrete time models	19
2.6	Some half-bridge entropy formulae	20
3	Stability of Markov operators	20
3.1	Overview	20
3.2	Entropy and divergence criteria	20
3.3	Kantorovich semi-distances	22
3.4	Contraction coefficients	23
3.4.1	Dobrushin coefficient	23
3.4.2	Kantorovich Lipschitz norm	23
3.4.3	Lyapunov methodologies	24
3.5	Transportation cost inequality	26
3.5.1	Log-Sobolev constants	26
3.5.2	Some illustrations	27
3.5.3	A stability theorem	28
4	A brief review on Sinkhorn bridges	30
4.1	Overview	30
4.2	Entropy-regularized transport	30
4.3	Some monotone properties	31
4.4	Sinkhorn potentials	32
4.5	Entropic projections	34
4.6	Commutation formulae	36
4.7	Gradient and Hessian formulae	36
5	Stability analysis	37
5.1	Overview	37
5.2	Linear decay rates	38
5.3	Bounded cost functions	38
5.3.1	Universal contraction rates	38
5.3.2	Schrödinger bridges	40
5.4	Gaussian bridges	41
5.4.1	Linear Gaussian models	41

5.4.2	Sinkhorn bridges	41
5.4.3	Riccati equations	43
5.4.4	Schrödinger bridges	44
5.4.5	A convergence theorem	45
5.5	Log-concave-at-infinity models	46
5.5.1	Curvature estimates	47
5.5.2	Contraction inequalities	48
5.5.3	Some quantitative estimates	49
5.5.4	Strongly convex models	50
5.6	Lyapunov techniques	51
5.6.1	Regularity conditions	51
5.6.2	A technical lemma	53
5.6.3	Some uniform estimates	54
5.6.4	A contraction theorem	56
5.6.5	Entropy estimates	58

1 Entropic optimal transport

1.1 Schrödinger and Sinkhorn bridges

Let μ and η be two probability measures on measurable spaces $\mathbb{X}$ and $\mathbb{Y}$ and let $\Pi(\mu, \eta)$ be the set of couplings of μ and η on the product space $(\mathbb{X} \times \mathbb{Y})$, with prescribed first and second coordinate marginals (μ, η) . The (static) Schrödinger bridge $P_{\mu, \eta}$ from μ to η with respect to a reference probability measure P on $(\mathbb{X} \times \mathbb{Y})$ is defined by the bridge map

$$P \mapsto P_{\mu, \eta} := \arg \min_{Q \in \Pi(\mu, \eta)} \mathcal{H}(Q|P) \quad \text{with} \quad \mathcal{H}(Q|P) := \int \log \left(\frac{dQ}{dP} \right) dQ \quad (1.1)$$

where dQ/dP the Radon-Nikodym derivative of Q with respect to P .

The bridge map $P \mapsto P_{\mu, \eta}$ is well defined as soon as $\mathcal{H}(Q|P) < \infty$ for some $Q \in \Pi(\mu, \eta)$, see for instance, the seminal article by Csiszár [39], [97, Section 6] and the survey article by Léonard [87] including the references therein. In contrast with conventional optimal transport maps, with limited applicability in high dimensions, Schrödinger bridges (1.1) can be solved using the celebrated Sinkhorn’s algorithm, a.k.a. iterative proportional fitting procedure or bi-proportional fitting in the statistics literature [40, 71, 72, 110, 111, 112], as well as the relative state attraction and matrix scaling [65, 80, 108, 109] approaches in computer sciences. These ideas provide a concrete roadmap to solve optimal transport problems and, as mentioned, detail numerous applications in computer science and statistics.

Different communities adopt distinct perspectives on the Sinkhorn algorithm. One common approach formulates Sinkhorn bridges as a sequence of transport plans on the product space $(\mathbb{X} \times \mathbb{Y})$, parameterized by dual potential functions on $\mathbb{X}$ and $\mathbb{Y}$. These potentials satisfy a nonlinear system of log-integral equations equivalent to Bregman’s iterative projection scheme for relative entropy minimization. The resulting alternating projection procedure admits an interpretation in terms of successive entropic half-bridge projections. Other formulations rely instead on continuous-time generative forward-backward diffusion models. To facilitate communication between fields, a concise overview of the principal formulations arising in applied probability, optimal control, and machine learning is provided in Sections 2 and 4.

To minimize the technical exposition in the introduction, we adopt a more direct and elementary approach based on conjugate Markov transitions (i.e., backward transitions and Bayes’ rule). This more conventional probabilistic framework streamlines the description and highlights the connection between the convergence analysis of the Sinkhorn algorithm and classical stability techniques for time-inhomogeneous Markov processes [60, 103, 104, 105, 106]; see also [9, 48] for related developments in the setting of time-varying positive semigroups. The present article further provides a self-contained treatment of new stability results derived from Markov semigroup contraction coefficients, log-Sobolev inequalities, Lyapunov-type operator-theoretic methods, and Riccati matrix differential equations recently introduced in [3, 4, 49].

We now briefly outline the basic principles of the algorithm in terms of conjugate transitions; Section 2.3 presents a more detailed summary. Given a Markov transition $K(x, dy)$ from $\mathbb{X}$ into $\mathbb{Y}$ we set

$$(\mu K)(dy) := \int_{\mathbb{X}} \mu(dx) K(x, dy) \quad \text{and} \quad (\mu \times K)(d(x, y)) := \mu(dx) K(x, dy). \quad (1.2)$$

Given a probability measure $R(d(y, x))$ on $(\mathbb{Y} \times \mathbb{X})$, we denote by $R^\flat$ the probability measure on $(\mathbb{X} \times \mathbb{Y})$ defined by

$$R^\flat(d(x, y)) := R(d(y, x)).$$

Observe that

$$P := \mu \times K \implies P = P_{\mu, \mu K} \in \Pi(\mu, \mu K) \quad \text{and} \quad (P_{\mu, \eta})^\flat = P_{\eta, \mu}^\flat := (P^\flat)_{\eta, \mu}. \quad (1.3)$$

The Sinkhorn iterations are defined sequentially for any $n \geq 0$ by a collection of probability distributions $\mathcal{P}_n$ on the product space $(\mathbb{X} \times \mathbb{Y})$ denoted by

$$\mathcal{P}_{2n} = \mu \times \mathcal{K}_{2n} \quad \text{and} \quad \mathcal{P}_{2n+1} = (\eta \times \mathcal{K}_{2n+1})^\flat \quad (1.4)$$

starting from a reference measure of the form $\mathcal{P}_0 = P := (\mu \times K)$ at rank $n = 0$. For $n \geq 1$, the Markov transitions $\mathcal{K}_n$ in (1.4) are defined sequentially by the conditioning Bayes-type formulae

$$\begin{cases} (\pi_{2n} \times \mathcal{K}_{2n+1})^\flat = \mu \times \mathcal{K}_{2n} & \text{and} \quad \pi_{2n+1} \times \mathcal{K}_{2(n+1)} = (\eta \times \mathcal{K}_{2n+1})^\flat \\ \text{with the distributions} \quad \pi_{2n} := \mu \mathcal{K}_{2n} & \text{and} \quad \pi_{2n+1} := \eta \mathcal{K}_{2n+1}. \end{cases} \quad (1.5)$$

Equivalently, for absolutely continuous models, with some abusive notation we have

$$\mathcal{P}_{2n}(d(x, y)) := \mu(dx) \mathcal{K}_{2n}(x, dy) = \pi_{2n}(dy) \mathcal{K}_{2n+1}(y, dx) \quad (1.6)$$

with the distribution

$$\pi_{2n}(dy) = (\mu \mathcal{K}_{2n})(dy) \quad \text{and} \quad \mathcal{K}_{2n+1}(y, dx) := \frac{\mu(dx) \mathcal{K}_{2n}(x, dy)}{(\mu \mathcal{K}_{2n})(dy)}.$$

Similarly, replacing in the r.h.s. of (1.6) the distribution π_{2n} by η we set

$$\mathcal{P}_{2n+1}(d(x, y)) := \eta(dy) \mathcal{K}_{2n+1}(y, dx) = \pi_{2n+1}(dx) \mathcal{K}_{2(n+1)}(x, dy) \quad (1.7)$$

with the distribution

$$\pi_{2n+1}(dx) = (\eta \mathcal{K}_{2n+1})(dx) \quad \text{and} \quad \mathcal{K}_{2(n+1)}(x, dy) := \frac{\eta(dy) \mathcal{K}_{2n+1}(y, dx)}{(\eta \mathcal{K}_{2n+1})(dx)}.$$

When $n \rightarrow \infty$, the marginal measures (π_{2n+1}, π_{2n}) converge towards (μ, η) and Sinkhorn bridges $\mathcal{P}_n$ converge towards the Schrödinger bridge $P_{\mu, \eta}$. In a mode of convergence that we define later, one has

$$(\pi_{2n+1}, \pi_{2n}) \xrightarrow{n \rightarrow \infty} (\mu, \eta) \quad \text{and} \quad \mathcal{P}_n \xrightarrow{n \rightarrow \infty} P_{\mu, \eta}.$$

The limiting properties as n grows are what we call, loosely, stability and this term also includes rates of convergence.

1.2 Sinkhorn semigroups

For finite state spaces, the Sinkhorn algorithm is a well known simple procedure that repeatedly applies column and row normalization on matrix spaces [56, 68, 85, 111, 112]. In problems of low-to-moderate dimension, whilst the recursions can be computed with ease, it is challenging to obtain explicit rates of convergence. For Gaussian models, Sinkhorn equations can also be solved analytically using sequential Bayes conjugacy principles [3, 35, 36, 37, 83]. As expected, this recursive formulation is closely related to the Kalman filter and related Riccati matrix difference equations, and it yields algorithms that can be implemented in practical settings without further approximations. A self contained analysis of Schrödinger bridges and the stability properties of Sinkhorn bridges for a general class of linear-Gaussian models is provided in Section 5.4. It is noted that in this Gaussian setting, the stability of the recursions is related to the stability of the Riccati equation; this concept is made concrete later in the article. In more general situations, while Sinkhorn iterations as presented in (1.5) may look appealing and easy to implement, one should note that they are based on nonlinear conditional/conjugate transformations with generally no finite-dimensional recursive solutions and, therefore, they do not lead to a practical algorithm. In this context, stability analysis of Sinkhorn bridges is crucial in the convergence analysis of any numerical approximations, including time discretization of diffusion bridges, neural networks and related score matching approximation techniques [23, 24, 41, 42, 113, 114, 115].

As underlined in [3], the key observation is that Sinkhorn semigroups belong to the class of time varying Markov processes. Indeed, integrating (1.5) we check the collection of fixed point equations

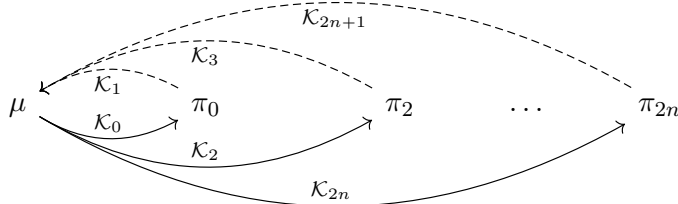
$$\mu = \pi_{2n}\mathcal{K}_{2n+1} = \mu\mathcal{S}_{2n+1} \quad \text{and} \quad \eta = \pi_{2n+1}\mathcal{K}_{2(n+1)} = \eta\mathcal{S}_{2(n+1)} \quad (1.8)$$

with the forward-backward Markov transitions

$$\begin{aligned} \mathcal{S}_{2n+1} &:= \mathcal{K}_{2n}\mathcal{K}_{2n+1} \quad \text{and} \quad \mathcal{S}_{2(n+1)} := \mathcal{K}_{2n+1}\mathcal{K}_{2(n+1)} \\ \implies \pi_{2n-1}\mathcal{S}_{2n+1} &= \pi_{2n+1} \quad \text{and} \quad \pi_{2n}\mathcal{S}_{2(n+1)} = \pi_{2(n+1)} \end{aligned} \quad (1.9)$$

where the product of transitions $\mathcal{K}_n\mathcal{K}_{n+1}$ is a shorthand notation for the composition $\mathcal{K}_n \circ \mathcal{K}_{n+1}$ of the right or left-action operators defined respectively in (1.11) (operation on functions) and in (1.2). In view of the fixed point equations (1.8) along with (1.9), when $n \rightarrow \infty$, we expect Sinkhorn distributions π_{2n} to converge towards η while π_{2n+1} converge towards μ .

A schematic picture of the time varying forward-backward semigroup of the distribution flow π_{2n} is given below:



Note that the product $\mathcal{S}_{2n+1} = \mathcal{K}_{2n}\mathcal{K}_{2n+1}$ can be interpreted as the 2-block Gibbs transition with target measure $\mathcal{P}_{2n}$ on $(\mathbb{X} \times \mathbb{Y})$. The transition $\mathcal{K}_{2n}(x, dy)$ is the $\mathcal{P}_{2n}$ -distribution of the coordinate y given x ; while $\mathcal{K}_{2n+1}(y, dx)$ is the $\mathcal{P}_{2n}$ -distribution of the coordinate x given y .

In addition, whenever the sequence of entropies of Sinkhorn distributions (1.9) w.r.t. the fixed point measures (1.8) are summable we have

$$\mathcal{H}(P_{\mu,\eta}|\mathcal{P}_{2n}) = \sum_{l \geq n} [\mathcal{H}(\eta|\pi_{2l}) + \mathcal{H}(\mu|\pi_{2l+1})] \iff \lim_{n \rightarrow \infty} \mathcal{H}(P_{\mu,\eta}|\mathcal{P}_{2n}) = 0. \quad (1.10)$$

Informally, under some conditions, in Theorem 5.2 we see the exponential decays of the marginals. Formally, a detailed proof of the above assertion is provided in Section 4.5, see also Proposition 4.1 for equivalent conditions. This crucial observation shows that the convergence of Sinkhorn bridges is exactly characterized by the convergence of Sinkhorn semigroups toward their limiting distributions.

As shown in (1.8) and (1.9) Sinkhorn transition semigroups $\mathcal{S}_n$ belong to the class of time varying Markov processes sharing a common invariant target measure at each individual step. These classes of time varying Markov semigroups arise in a variety of areas in applied mathematics including nonlinear filtering, physics and molecular chemistry; see for instance [9, 45, 52, 48, 103, 104, 105, 106] and references therein. The stability analysis of time inhomogeneous Markov processes is a broad and notoriously difficult subject, far more challenging than their time homogeneous counterparts; this is because the operators may drastically change during the semigroup evolution. The articles [60, 103, 104, 105] provide several useful probabilistic and analytical tools, including coupling methods, spectral analysis as well as functional inequalities such as Nash or log-Sobolev inequalities. Nevertheless, Sinkhorn semigroups are defined sequentially by nonlinear conjugate transformations and hence the application of these refined analytical techniques in this context is rather challenging.

The nonlinear Sinkhorn/Gibb-type semigroup approach taken in this article differs significantly from the conventional approaches developed in the optimal transport literature; this is discussed in some detail in Section 1.3. This novel framework allows one to unify and simplify many arguments in the stability of Sinkhorn bridges, as well as to improve upon some existing results. We shall survey, in a self contained manner, some powerful off-the-shelf semigroup techniques recently developed in the series of articles [3, 4, 49], themselves based on the contraction operator-theoretic analysis developed in [9, 48, 52, 53, 51, 54]. The approaches that we focus upon are:

- Transportation cost inequalities including local log-Sobolev and Talagrand quadratic cost inequalities as well as Fisher-Lipschitz-type inequalities and a novel entropic stability theorem.
- An operator-theoretic framework based on contraction coefficients, including Lyapunov methodologies on Weighted Banach spaces, Kantorovich criteria and Wasserstein distances. This approach allows one to transfer what is known about Markov chains to the context of Sinkhorn semigroups.

In the context of Sinkhorn and Schrödinger bridges, the article [49] presents a semigroup approach combining transportation cost inequalities including log-Sobolev and Talagrand

quadratic inequalities with the theory of Riccati matrix difference equations arising in filtering and optimal control theory. In the present article, we illustrate these inequalities in the context of log-concave-at-infinity models. Other types of models, such as sub-Gaussian or log-concave, are discussed in [49]. The analysis of marginal models with lower/upper bounded curvature is closely related to Gaussian models. Using Brascamp-Lieb and Cramer-Rao inequalities, the analysis of these strongly convex models also relies on discrete time algebraic Riccati equations (see for instance Appendix D in [49] and Section 5.5.4 in the present article). When applied to the stability of Sinkhorn semigroups, they also yield a series of novel contraction estimates in terms of fixed point of Riccati equations. As mentioned above, a detailed analysis of Sinkhorn bridges for linear-Gaussian models is presented in [3], including finite-dimensional recursive formulation of the Sinkhorn algorithm and sharp convergence rates in terms of Riccati matrix difference equations.

The article [4] introduces an operator-theoretic framework yielding new exponential decays of Sinkhorn iterates towards Schrödinger bridges with respect to general classes of ϕ -divergences and Kantorovich-type criteria, including the relative entropy, squared Hellinger integrals, α -divergences, $\mathbb{L}_p$ -norms, Havrda-Charvat entropies, Jensen-Shannon divergence, as well as Jeffreys and Rényi divergences. [4] also develops a novel Lyapunov contraction approach under minimal regularity conditions. This methodology provides quantitative exponential stability estimates for a large class of Sinkhorn semigroups on weighted Banach spaces. Contraction estimates are expressed in terms of Kantorovich semi-distances, Wasserstein distances as well as Dobrushin-type contraction coefficients and weighted total variation norms. This novel Lyapunov methodology is applied to a variety of situations, ranging from polynomial growth potentials and heavy tailed marginals on general normed spaces, to more sophisticated boundary state space models, including semi-circle transitions, Beta, Weibull, exponential marginals as well as semi-compact models. This approach also allows one to consider statistical finite mixture of the above models, including kernel-type density estimators of complex data distributions arising in generative modeling.

As a result of the methods we consider, in this article, we can often obtain quantitative and refined estimates. A precise description of the convergence rates and estimation constants are provided in the series of articles [3, 4, 9, 49]. We remark that the articles [3, 4, 49] are partly based on well-established techniques in Riccati difference equations and in the stability theory of Markov semigroups. The exponential stability of Sinkhorn bridges in [3, 49] rely on the stability analysis of Riccati matrix difference equations presented in [47]. The contraction properties of Sinkhorn semigroups presented in [4] combine the contraction analysis of Markov semigroups developed in [52, 53] with the Lyapunov operator theoretic framework on weighted Banach spaces presented in [48, 54], and further developed in the articles [9, 51]. In the context of time-homogeneous models, we also refer to the pioneering articles by Hasminskii [77] and Meyn and Tweedie [91], see also the book [92] and the more recent seminal article by Hairer and J. C. Mattingly [76].

1.3 Literature review

The stability properties of Schrödinger and Sinkhorn bridges when the number of iterations tends to ∞ is a very active research area in applied probability and machine learning. Most of the literature on the exponential convergence of Sinkhorn equations is concerned with finite state spaces [17, 66, 111, 112, 116] as well as compact state spaces or bounded cost functions

using nonlinear versions of Perron-Frobenius theory [18, 80] and related Hilbert projective metrics techniques [26, 43, 69, 57], based on the seminal article by Birkhoff [15]. See also [75] for a refined analysis of Sinkhorn iterates and their gradient on the d -dimensional torus as well as the more recent article [30] on compact spaces and bounded cost functions.

Sublinear rates have been developed in the articles [5, 22, 61]. Linear rates on non-necessarily compact spaces were first obtained by Léger in [86] using elegant gradient descent and Bregman divergence techniques (see also the recent articles [42, 84]). Refined convergence rates at least one order faster for a general class of models including sub-Gaussian target marginals has also been developed in [73] under some exponential moments estimates on Schrödinger potentials.

The first exponential stability analysis of unbounded models and not necessarily compactly supported marginals is presented in the article [34]. In this study, the authors investigate quantitative contraction rates for target marginal distributions with an asymptotically positive log-concavity profile and quadratic symmetric cost functions associated with a sufficiently large regularization parameter. The approach taken in [34] relies on well known links between gradient and Hessian of Sinkhorn and Schrödinger potentials along Hamilton-Jacobi-Bellman equations with conditional expectations and covariances [28, 100], see also [50]. The main conditions in [34] are expressed in terms of the integrated convexity profiles of marginal potentials. Extending ideas from [33] the proofs are based on the propagation of these convexity profiles based on coupling diffusion by reflection techniques introduced in [62]. The exponential entropic decays presented in [34] were further developed and recently refined in [29] to apply to all values of the regularization parameter and some classes of perturbation of log-concave models. However this stability approach requires one to estimate the semi-concavity properties of generally unknown dual Schrödinger bridge transitions and it only applies to entropically regularized transport problems with symmetric quadratic-type costs. Thus, this framework excludes the important cases when the reference transition is associated with linear Gaussian transitions arising in Ornstein-Uhlenbeck-type diffusion generative models and denoising diffusions.

The article [63] is also based on an extension of Hilbert's projective metric for spaces of integrable functions of bounded growth. Extending results in [26, 43] to unbounded cost functions, the recent article [63] presents exponential decays of these new classes of Hilbert's projective metrics with respect to some cone of functions in settings where the marginal distributions have sufficiently light tails compared to the growth of the cost function. These results also imply the exponential convergence of the total variation distance. On normed finite dimensional spaces, the main condition in [63] is that the log of the reference transition density grows faster than a polynomial of order p and the tails of both marginals decay faster than $e^{-r^{p+q}}$ for some $q > 0$. This condition does not cover Gaussian models. It is out of the scope of this article to present more details of these different approaches. Despite the numerous references given above, a complete literature review is not possible. We refer to reader to the books [97, 102] as well as the survey articles [27, 70, 79, 80, 107, 121, 123], recent article articles [3, 4, 29, 34, 49, 74] and the numerous references therein. We underline that all the methodologies discussed above differ from the recently developed Sinkhorn semigroup contraction analysis and the Lyapunov operator-theoretic framework discussed in Section 1.2.

1.4 Article focus and structure

The objective of this survey article is to detail, in a pedagogical manner, the subject of Sinkhorn semigroup techniques, along with quantitative stability estimates. Although, as mentioned above, we do discuss some existing results, we do not survey the entire literature with all of the results that have been derived. One of the main motivations of this paper is to equip the reader with many of the tools that are needed to analyze Sinkhorn semigroups, without having to go beyond the technical methods that have been employed in the analysis of non-homogenous Markov chains (e.g. [48, 60]). This in turn leads to short and elegant proofs of existing results, along with, in some cases, a substantial improvement over those currently published. We again remark that, other than the finite state-space or Gaussian case, Sinkhorn algorithm recursions are numerically intractable. An important ingredient in the mathematical analysis of numerical approximations is the time stability of the previously mentioned recursions. Indeed an understanding of time stability could lead to a motivation for new numerical methods. This idea is analogous to the stability of the non-linear filter which has been studied in some depth for more than 30 years (e.g. [52] or [11, 45] and the references therein). Once one understands whether a non-linear filter is stable, one may hope to derive numerical algorithms which approximate the filter and can be stable themselves (we refer the reader to [11, 45] to understand what stability is in non-linear filtering).

This article is structured as follows. In the following subsection 1.5 we provide the main notations that are adopted in the paper. Section 2 presents some background analysis on Schrödinger and Sinkhorn bridges. In section 3 we give a summary on the stability of Markov operators which is the approach that we adopt. In section 4 a brief review of Sinkhorn bridges is given. Finally in section 5 we detail our stability analysis.

1.5 Some basic notation

Integral operators

Given a Markov transition $K(x, dy)$ from $\mathbb{X}$ into $\mathbb{Y}$ and a function f on $\mathbb{Y}$ we set

$$K(f)(x) := \int_{\mathbb{Y}} K(x, dy) f(y). \quad (1.11)$$

We are using K to denote both the left-action operator on measure (1.2) and the right-action operator on functions (1.11). Denote by $\Pi(\mu, \cdot)$ be the set of probability measures on $(\mathbb{X} \times \mathbb{Y})$ with $\mathbb{X}$ -marginal μ ; and $\Pi(\cdot, \eta)$ is the set of probability measures on $(\mathbb{X} \times \mathbb{Y})$ with $\mathbb{Y}$ -marginal η . Note that

$$(\mu \otimes \eta)(d(x, y)) := \mu(dx) \eta(dy) \implies (\mu \otimes \eta) \in \Pi(\mu, \eta).$$

The Dirac delta measure located at x is denote by δ_x . We also denote by $|\nu| := \nu_+ + \nu_-$ the total variation measure associated with a Hahn-Jordan decomposition $\nu = \nu_+ - \nu_-$ of a bounded signed measure ν . For indicator functions $f = 1_A$ of a measurable subset $A \subset \mathbb{X}$ and a given measure μ on $\mathbb{X}$ sometimes we write $\mu(A)$ instead of $\mu(1_A)$.

We let $\mathcal{B}(\mathbb{X})$ denote the set of measurable functions on $\mathbb{X}$ and $\|f\| := \sup_{x \in \mathbb{X}} |f(x)|$ denotes the uniform norm of a function $f \in \mathcal{B}(\mathbb{X})$. Denote by $\mathbf{Prob}(\mathbb{X})$ the set of probability measures on $\mathbb{X}$. Next, we let $g \in \mathcal{B}(\mathbb{X})$ be a lower semi-continuous (abbreviated l.s.c.) function, lower bounded away from zero, and we denote by $\mathcal{B}_g(\mathbb{X})$ the subspace of functions

$f \in \mathcal{B}(\mathbb{X})$ such that $\|f/g\| < \infty$. Denote by $\text{Prob}_g(\mathbb{X})$ be the set of probability measures μ on $\mathbb{X}$ such that $\mu(g) < \infty$. $\mathcal{B}_\infty(\mathbb{X})$ is the collection of locally upper bounded and uniformly positive functions that grow at infinity (in the sense that their sub-level sets are compact). $\mathcal{B}_0(\mathbb{X}) = \{\theta : 1/\theta \in \mathcal{B}_\infty(\mathbb{X})\}$ is the sub-algebra of (bounded) positive functions $\theta(x)$ which are locally lower bounded and vanish at infinity in the sense that, for any $0 < \epsilon < \|\theta\|$, the set $\{\theta \geq \epsilon\}$ is a non-empty compact set.

Differential operators

We let $\nabla f(x) = [\partial_{x_i} f(x)]_{1 \leq i \leq d}$ be the gradient column vector associated with some smooth function $f(x)$ from $\mathbb{R}^d$ into $\mathbb{R}$. Given some smooth function $F(x)$ from $\mathbb{R}^d$ into $\mathbb{R}^p$ we denote by $\nabla h = [\nabla F^1, \dots, \nabla F^p] \in \mathbb{R}^{n \times p}$ the gradient matrix associated with the column vector function $F = (F^i)_{1 \leq i \leq p}$, for some $p \geq 1$. For instance, for a given matrix $A \in \mathbb{R}^{q \times p}$ we have $Ah(x) \in \mathbb{R}^q$ and therefore

$$\nabla(AF(x)) = (\nabla F(x)) A' \quad (1.12)$$

where A' is the transpose of the matrix A . Given a smooth function $f(x, y)$ from $\mathbb{R}^d \times \mathbb{R}^d$ into $\mathbb{R}$ and $i = 1, 2$ we denote by ∇_i the gradient operator defined by

$$\nabla_i f = \begin{bmatrix} (\nabla_i f)^1 \\ \vdots \\ (\nabla_i f)^d \end{bmatrix} \quad \text{with} \quad (\nabla_1 f)^j := \partial_{x^j} f \quad \text{and} \quad (\nabla_2 f)^j := \partial_{y^j} f.$$

For smooth p -column valued functions $F(x, y) = (F^j(x, y))_{1 \leq j \leq p}$ we also set

$$\nabla_i F = [\nabla_i F^1, \dots, \nabla_i F^p].$$

In this notation, the second order differential operators $\nabla_i \nabla_j$ on function $f(x, y)$ from $\mathbb{R}^d \times \mathbb{R}^d$ into $\mathbb{R}$ are defined by $(\nabla_i \nabla_j)(f) := \nabla_i(\nabla_j(f))$ and we set $\nabla_i^2 := \nabla_i \nabla_i$ when $i = j$.

Conditional covariances

Let $\mathcal{G}l_d$ denote the general linear group of $(d \times d)$ -invertible matrices, and $\mathcal{S}_d^+ \subset \mathcal{G}l_d$ the subset of positive definite matrices. We sometimes use the Löwner partial ordering notation $v_1 \geq v_2$ to mean that a symmetric matrix $v_1 - v_2$ is positive semi-definite (equivalently, $v_2 - v_1$ is negative semi-definite), and $v_1 > v_2$ when $v_1 - v_2$ is positive definite (equivalently, $v_2 - v_1$ is negative definite). Given $v \in \mathcal{S}_d^+$ we denote by $v^{1/2}$ the principal (unique) symmetric square root.

Consider bounded positive integral operators from $\mathbb{R}^d$ into itself of the form

$$K(x, dy) = e^{-W(x, y)} dy \quad \text{and} \quad K_V(x, dy) := \frac{K(x, dy) e^{-V(y)}}{K(e^{-V})(x)} \quad (1.13)$$

for some real valued functions $W(x, y)$ and $V(y)$. Denote by W_x and $W_{x,x}$ the gradient and the Hessian of the transition potential $W(x, y)$ with respect to the first coordinate; that is

$$W_x(y) := \nabla_1 W(x, y) \quad \text{and} \quad W_{x,x}(y) := \nabla_1^2 W(x, y). \quad (1.14)$$

The gradient and the Hessian formulae are

$$\begin{aligned} f(x) &:= \log K(e^{-V})(x) \\ \implies \nabla f(x) &= -K_V(W_x)(x) \quad \text{and} \quad \nabla^2 f(x) = -K_V(W_{x,x})(x) + \text{cov}_{K_V, W}(x) \end{aligned} \quad (1.15)$$

with the conditional covariance

$$\text{cov}_{K_V, W}(x) := \frac{1}{2} \int K_V(x, dy_1) K_V(x, dy_2) (W_x(y_1) - W_x(y_2)) (W_x(y_1) - W_x(y_2))' \quad (1.16)$$

where we recall v' is the transpose of a matrix or vector v . Denote by $\text{mean}_K(x)$ and $\text{cov}_K(x)$ the conditional mean and covariance functions

$$\begin{aligned} \text{mean}_K(x) &:= \int K(x, dy) y \\ \text{cov}_K(x) &:= \frac{1}{2} \int K(x, dy_1) K(x, dy_2) (y_1 - y_2) (y_1 - y_2)'. \end{aligned} \quad (1.17)$$

Assume the log-density function W satisfy the strong convexity condition

$$\tau^{-1} \leq \nabla^2 W(x, y) \leq \bar{\tau}^{-1} \quad \text{for some } \tau, \bar{\tau} \in \mathcal{S}_d^+. \quad (1.18)$$

In this context, the Brascamp-Lieb inequality ensures that

$$\text{cov}_K(x) \leq \int K(x, dy) (\nabla^2 W(x, y))^{-1} \leq \tau.$$

Conversely, the Cramer-Rao inequality also ensures that

$$\text{cov}_K(x)^{-1} \leq \int K(x, dy) \nabla^2 W(x, y) \leq \bar{\tau}^{-1}.$$

We summarize the above discussion with the covariance estimate

$$(1.18) \implies \bar{\tau} \leq \text{cov}_K(x) \leq \tau. \quad (1.19)$$

Riccati difference equations

The Frobenius matrix norm of a given matrix $v \in \mathbb{R}^{d_1 \times d_2}$ is defined by $\|v\|_F^2 = \text{Tr}(v'v)$, with the trace operator $\text{Tr}(\cdot)$. The spectral norm is defined by $\|v\|_2 = \sqrt{\lambda_{\max}(v'v)}$ and $\lambda_{\max}(v'v)$ is the maximum eigenvalue of $v'v$. Note that, for $x \in \mathbb{R}^{d \times 1} = \mathbb{R}^d$, the Frobenius and the spectral norm $\|x\|_F = \sqrt{x'x} = \|x\|_2$ coincides with the Euclidean norm. When there is no possible confusion, sometimes we use the notation $\|\cdot\|$ for any equivalent matrix or vector norm.

For any $u, v \in \mathcal{S}_d^+$ we recall the Ando-Hemmen inequality

$$\|u^{1/2} - v^{1/2}\| \leq \left[\lambda_{\min}^{1/2}(u) + \lambda_{\min}^{1/2}(v) \right]^{-1} \|u - v\| \quad (1.20)$$

that holds for any unitary invariant matrix norm $\|\cdot\|$, including the spectral and the Frobenius norms, see for instance Theorem 6.2 on page 135 in [81], as well as Proposition 3.2 in [8].

Let us associate with some $\varpi \in \mathcal{S}_d^+$ the increasing map Ricc_ϖ from $\mathcal{S}_d^0$ into $\mathcal{S}_d^+$ defined by

$$\begin{aligned} \text{Ricc}_\varpi : \mathcal{S}_d^0 &\mapsto \mathcal{S}_d^+ \\ v &\rightsquigarrow \text{Ricc}_\varpi(v) := (I + (\varpi + v)^{-1})^{-1}. \end{aligned} \quad (1.21)$$

A refined stability analysis of Riccati matrix difference equation

$$v_{n+1} := \text{Ricc}_\varpi(v_n) \quad (1.22)$$

and the limiting stationary matrices $r = \text{Ricc}_\varpi(r)$ associated with these maps is provided in [3]. Notice that

$$\begin{aligned} r = \text{Ricc}_\varpi(r) &\iff r^{-1} = I + (\varpi + r)^{-1} \iff \varpi r^{-1} + I = (\varpi + r)r^{-1} = (\varpi + r) + I \\ &\iff \varpi r^{-1} = \varpi + r \iff r^{-1} = I + \varpi^{-1}r \iff I = r + r\varpi^{-1}r. \end{aligned} \quad (1.23)$$

Also note the increasing property

$$v_1 \leq v_2 \implies \text{Ricc}_\varpi(v_1) \leq \text{Ricc}_\varpi(v_2)$$

For any $n \geq p \geq 1$ and $v \in \mathcal{S}_0^d$ we also have the estimates

$$\text{Ricc}_\varpi^p(0) \leq \text{Ricc}_\varpi^n(0) \leq \text{Ricc}_\varpi^n(v) \leq \text{Ricc}_\varpi^{n-1}(I) \leq \text{Ricc}_\varpi^{p-1}(I) \leq I \quad (1.24)$$

The matrix fixed point equations (1.23) belong to the class of discrete algebraic Riccati equations (DARE), and no analytical solutions are available for general models. In our simplified framework, the unique positive definite fixed point (1.23) of the Riccati difference equation (1.22) is given by

$$(I + \varpi^{-1})^{-1} \leq r := -\frac{\varpi}{2} + \left(\varpi + \left(\frac{\varpi}{2} \right)^2 \right)^{1/2} \leq I. \quad (1.25)$$

In addition, there exists some constant c_ϖ such that

$$\|v_n - r\|_2 \leq c_\varpi (1 + \lambda_{\min}(\varpi + r))^{-2n} \|v_0 - r\|_2 \quad (1.26)$$

The contraction rates in (1.26) are based on Floquet-type representation of Riccati flows and they are sharp (see Theorem 1.3 in [47]). For further discussion see [3, 47], and the references therein.

The positive and negative part of $a \in \mathbb{R}$ are defined respectively by $a_+ = \max(a, 0)$ and $a_- = \max(-a, 0)$. Given $a, b \in \mathbb{R}$ we also set $a \vee b = \max(a, b)$ and $a \wedge b = \min(a, b)$.

2 Some background on bridges

2.1 Overview

The theory of entropic optimal transport is at the carrefour of statistics and applied probability, optimal control theory, machine learning and artificial intelligence. Entropic bridges have emerged as powerful techniques in a various fields of physics, economics, engineering

and computer sciences. Every community often has a different way to interpret and formulate Schrödinger and Sinkhorn bridges and every formulation has its own advantages and limitations. To guide the reader, this section provides a comprehensive overview of some of the main formulations used in this field. We remark that if the reader is already an expert in one of the possible presentations of Schrödinger and Sinkhorn bridges then it is possible skip most of the following sections with little loss in continuity.

Section 2.2 presents a probabilistic formulation of Sinkhorn bridges in terms of conjugate and dual integral operators. Equivalent Bayes' type formulations in terms of state space models are discussed in Section 2.3. Section 2.4 underlines the links between static and dynamic Schrödinger bridges in the context of discrete generation Markov processes. Forward-backward formulations for continuous time generative-type diffusion models are discussed in 2.5. Section 2.6 provides an alternative analytical interpretation of Sinkhorn bridges in terms of entropic projections.

2.2 Conjugate and dual transitions

When $\delta_x K \ll \mu K$ for any $x \in \mathbb{X}$, we have the Bayes' formula

$$\mu \times K = ((\mu K) \times K_\mu^*)^\flat \quad \text{with} \quad K_\mu^*(y, dx) := \mu(dx) \frac{d\delta_x K}{d\mu K}(y). \quad (2.1)$$

A more general definition of the dual transition K_μ^* is provided in Section 2 in the article [53]. Note the reversible property

$$\mu(dx)K(x, dy)K_\mu^*(y, d\bar{x}) = \mu(d\bar{x})K(\bar{x}, dy)K_\mu^*(y, dx) \quad \text{and therefore} \quad \mu K K_\mu^* = \mu.$$

By symmetry arguments, when $\delta_y L \ll \eta L$ for any $y \in \mathbb{Y}$, we have the Bayes' formula

$$(\eta \times L)^\flat = (\eta L) \times L_\eta^* \quad \text{with} \quad L_\eta^*(x, dy) := \eta(dy) \frac{d\delta_y L}{d\eta L}(x). \quad (2.2)$$

Again, we have the reversible property

$$\eta(dy)L(y, dx)L_\eta^*(x, d\bar{y}) = \eta(d\bar{y})L(\bar{y}, dx)L_\eta^*(x, dy) \quad \text{and therefore} \quad \eta L L_\eta^* = \eta.$$

Consider some reference probability measure $P \in \Pi(\mu, \cdot)$. We have disintegration formula $P = (\mu \times K)$, for some Markov transition K from $\mathbb{X}$ into $\mathbb{Y}$. Sinkhorn bridges $\mathcal{P}_n$ are defined sequentially starting from a reference probability

$$\mathcal{P}_0 := P \iff (\mathcal{P}_0 := \mu \times \mathcal{K}_0 \quad \text{with} \quad \mathcal{K}_0 := K).$$

Rewritten in a slightly different way, the Bayes' rule (2.1) takes the form

$$\mathcal{P}_0 := \mu \times \mathcal{K}_0 = (\pi_0 \times \mathcal{K}_1)^\flat \in \Pi(\mu, \pi_0) \quad \text{with} \quad \pi_0 := \mu \mathcal{K}_0 \quad \text{and} \quad \mathcal{K}_1 := (\mathcal{K}_0)_\mu^*.$$

In the same vein, Bayes' rule (2.2) takes the form

$$\mathcal{P}_1 := (\eta \times \mathcal{K}_1)^\flat = \pi_1 \times \mathcal{K}_2 \in \Pi(\pi_1, \eta) \quad \text{with} \quad \pi_1 := \eta \mathcal{K}_1 \quad \text{and} \quad \mathcal{K}_2 := (\mathcal{K}_1)_\eta^*.$$

Iterating the above procedure we define for even indices a collection of bridges

$$\mathcal{P}_{2n} := \mu \times \mathcal{K}_{2n} = (\pi_{2n} \times \mathcal{K}_{2n+1})^\flat \in \Pi(\mu, \pi_{2n}) \quad (2.3)$$

with the marginal distributions

$$\pi_{2n} := \mu \mathcal{K}_{2n} \quad \text{and the transitions} \quad \mathcal{K}_{2n+1} := (\mathcal{K}_{2n})_{\mu}^*.$$

Sinkhorn bridges for odd indices are given by

$$\mathcal{P}_{2n+1} := (\eta \times \mathcal{K}_{2n+1})^{\flat} = \pi_{2n+1} \times \mathcal{K}_{2(n+1)} \in \Pi(\pi_{2n+1}, \eta) \quad (2.4)$$

with the marginal distributions

$$\pi_{2n+1} := \eta \mathcal{K}_{2n+1} \quad \text{and the transitions} \quad \mathcal{K}_{2(n+1)} := (\mathcal{K}_{2n+1})_{\eta}^*.$$

2.3 State space models

In terms of random variables (X, Y) with distribution $P = \mu \times K$ we have

$$P(d(x, y)) = \mathbb{P}((X, Y) \in d(x, y)) = \mathbb{P}(X \in dx) \mathbb{P}(Y \in dy \mid X = x) = \mu(dx) K(x, dy)$$

with the distributions

$$\begin{aligned} \mu(dx) &:= \mathbb{P}(X \in dx) \quad \text{and} \quad K(x, dy) := \mathbb{P}(Y \in dy \mid X = x) \\ \implies (\mu K)(dy) &= \mathbb{P}(Y \in dy) \quad \text{and} \quad K_{\mu}^*(y, dx) = \mathbb{P}(X \in dx \mid Y = y). \end{aligned}$$

For continuous random variables on $\mathbb{X} = \mathbb{R}^p = \mathbb{Y}$ with conditional densities with respect to the Lebesgue measure we have

$$\begin{aligned} \mu(dx) &= p_X(x) dx \quad \text{and} \quad K(x, dy) = p_{Y|X}(y|x) dy \\ \implies P(d(x, y)) &= p_{X,Y}(x, y) dx dy := p_X(x) p_{Y|X}(y|x) dx dy \end{aligned}$$

where $p_{X,Y}(x, y)$ is the joint density of (X, Y) , $p_X(x)$ the marginal density of X and $p_{Y|X}(y|x)$ is the conditional density of Y given $X = x$. In this context, Bayes' rule (2.1) takes the form

$$p_{X,Y}(x, y) = p_X(x) p_{Y|X}(y|x) = p_Y(y) p_{X|Y}(x|y) \quad \text{and} \quad K_{\mu}^*(y, dx) = p_{X|Y}(x|y) dx.$$

Sinkhorn bridges

Consider random variables (X_n, Y_n) such that

$$\mathcal{P}_n(d(x, y)) = \mathbb{P}((X_n, Y_n) \in d(x, y))$$

with the Sinkhorn bridges $\mathcal{P}_n$ defined in (1.4). For even indices we have

$$\begin{aligned} \mathcal{P}_{2n} &:= \mu \times \mathcal{K}_{2n} \\ \iff \mu(dx) &= \mathbb{P}(X_{2n} \in dx) \quad \text{and} \quad \mathcal{K}_{2n}(x, dy) = \mathbb{P}(Y_{2n} \in dy \mid X_{2n} = x). \end{aligned}$$

In this context, Bayes' rule (2.3) holds with

$$\pi_{2n}(dy) = \mathbb{P}(Y_{2n} \in dy) \quad \text{and} \quad \mathcal{K}_{2n+1}(y, dx) = \mathbb{P}(X_{2n} \in dx \mid Y_{2n} = y). \quad (2.5)$$

For odd indices we have

$$\mathcal{P}_{2n+1} := (\eta \times \mathcal{K}_{2n+1})^\flat$$

$$\iff \eta(dy) = \mathbb{P}(Y_{2n+1} \in dy) \quad \text{and} \quad \mathcal{K}_{2n+1}(y, dx) = \mathbb{P}(X_{2n+1} \in dx \mid Y_{2n+1} = y).$$

Again, Bayes' rule (2.4) holds with

$$\pi_{2n+1}(dx) = \mathbb{P}(X_{2n+1} \in dx) \quad \text{and} \quad \mathcal{K}_{2(n+1)}(x, dy) = \mathbb{P}(Y_{2n+1} \in dy \mid X_{2n+1} = x). \quad (2.6)$$

Density models

Consider locally compact Polish spaces $(\mathbb{X}, \lambda)$ and $(\mathbb{Y}, \nu)$ endowed with some locally bounded positive measures λ and ν . Also assume the marginal probability measures are given by

$$\mu(dx) = \lambda_U(dx) := e^{-U(x)} \lambda(dx) \quad \text{and} \quad \eta(dy) = \nu_V(dy) := e^{-V(y)} \nu(dy) \quad (2.7)$$

for some normalized potential functions $U : x \in \mathbb{X} \mapsto U(x) \in \mathbb{R}$ and $V : y \in \mathbb{Y} \mapsto V(y) \in \mathbb{R}$. We have

$$\mathbb{P}((X_n, Y_n) \in d(x, y)) = p_{(X_n, Y_n)}(x, y) \gamma(d(x, y)) \quad \text{with} \quad \gamma := \lambda \otimes \nu.$$

In addition

$$\begin{aligned} p_{(X_{2n}, Y_{2n})}(x, y) &= p_{X_{2n}}(x) p_{Y_{2n}|X_{2n}}(y|x) \quad \text{with} \quad p_{X_{2n}}(x) = e^{-U(x)} \\ p_{(X_{2n+1}, Y_{2n+1})}(x, y) &= p_{Y_{2n+1}}(y) p_{X_{2n+1}|Y_{2n+1}}(x|y) \quad \text{with} \quad p_{Y_{2n+1}}(y) = e^{-V(y)}. \end{aligned} \quad (2.8)$$

In this scenario, the recursions in (2.5) and (2.6) take the form

$$p_{X_{2n+1}|Y_{2n+1}} := p_{X_{2n}|Y_{2n}} \quad \text{and} \quad p_{Y_{2(n+1)}|X_{2(n+1)}} := p_{Y_{2n+1}|X_{2n+1}}. \quad (2.9)$$

2.4 Dynamic and static bridges

Consider a Markov chain Z_k on $\mathbb{X}$ with transition semigroup $M_{l,k}$ with $l \leq k$ defined by

$$M_{l,k}(z_l, dz_k) := \mathbb{P}(Z_k \in dz_k \mid Z_l = z_l).$$

We fix the time horizon n and observe that

$$\begin{aligned} \mu(dz_0) &:= \mathbb{P}(Z_0 \in dz_0) \\ \implies P(d(z_0, z_n)) &:= \mathbb{P}((Z_0, Z_n) \in d(z_0, z_n)) = (\mu \times K)(d(z_0, z_n)) \quad \text{with} \quad K = M_{0,n}. \end{aligned}$$

Pinned process

The conditional distribution of the random path $(Z_1, \dots, Z_{n-1})$ given the initial and terminal states (Z_0, Z_n) (we call this a pinned Markov chain) is given by

$$\begin{aligned} \mathbb{P}((Z_1, \dots, Z_{n-1}) \in (d(z_1, \dots, z_{n-1})) \mid (Z_0, Z_n) = (z_0, z_n)) \\ = P_{0 \rightarrow n}^{(z_0, z_n)}(d(z_1, \dots, z_{n-1})) := \prod_{1 \leq k < n} M_{k-1,k}^{(n, z_n)}(z_{k-1}, dz_k). \end{aligned}$$

with the collection of Markov transitions

$$M_{k-1,k}^{(n,z_n)}(z_{k-1}, dz_k) := \frac{M_{k-1,k}(z_{k-1}, dz_k) M_{k,n}(z_k, dz_n)}{M_{k-1,n}(z_{k-1}, dz_n)}.$$

Note that the above transitions do not depend on μ and we have

$$\begin{aligned} \mathbf{P}(d(z_0, \dots, z_n)) &:= \mathbb{P}((Z_0, \dots, Z_n) \in (d(z_0, \dots, z_n))) \\ &= P(d(z_0, z_n)) \times P_{0 \rightarrow n}^{(z_0, z_n)}(d(z_1, \dots, z_{n-1})). \end{aligned}$$

Let $\Pi_n(\mu, \eta)$ be the set of probability measures on the path space $\mathbb{X}^{n+1} := \mathbb{X} \times \mathbb{X}^n$ with marginals μ and η at time $k = 0$ and $k = n$. The dynamic Schrödinger bridge between μ and η with reference distribution $\mathbf{P}$ is given by the bridge map

$$\mathbf{P}_{\mu, \eta} := \arg \min_{\mathbf{Q} \in \Pi_n(\mu, \eta)} \mathcal{H}(\mathbf{Q} \mid \mathbf{P}).$$

Any probability measure $\mathbf{Q} \in \Pi_n(\mu, \eta)$ can be disintegrated w.r.t. the marginal distribution $Q(d(z_0, z_n))$ of the initial and terminal states, namely

$$\mathbf{Q}(d(z_0, \dots, z_n)) = Q(d(z_0, z_n)) \times Q_{0 \rightarrow n}^{(z_0, z_n)}(d(z_1, \dots, z_{n-1})).$$

This yields the entropy factorization

$$\mathcal{H}(\mathbf{Q} \mid \mathbf{P}) = \mathcal{H}(Q \mid P) + \int Q_n(d(z_0, z_n)) \mathcal{H}(Q_{0 \rightarrow n}^{(z_0, z_n)} \mid P_{0 \rightarrow n}^{(z_0, z_n)})$$

from which we conclude [3, 25, 64, 67, 87] that the static Schrödinger bridge $P_{\mu, \eta}$ and dynamic Schrödinger bridge $\mathbf{P}_{\mu, \eta}$ are connected by the formulae

$$\mathbf{P}_{\mu, \eta}(d(z_0, \dots, z_n)) = P_{\mu, \eta}(d(z_0, z_n)) \times P_{0 \rightarrow n}^{(z_0, z_n)}(d(z_1, \dots, z_{n-1})).$$

Backward conditioning formulae

To simplify notation we consider a time homogenous Markov transition $M_{k-1,k} = M$. In this context, we have the backward conditioning formulae

$$\mathbb{P}((Z_0, \dots, Z_{n-1}) \in d(z_0, \dots, z_{n-1}) \mid Z_n = z_n) = \prod_{n > k \geq 0} M_{\mu_k}^*(z_{k+1}, dz_k)$$

the dual transitions

$$M_{\mu_k}^*(z_{k+1}, dz_k) := \frac{\mu_k(dz_k) M(z_k, dz_{k+1})}{(\mu_k M)(dz_{k+1})} \quad \text{and} \quad \mu_k(dz) = \mathbb{P}(Z_k \in dz) = (\mu_{k-1} M)(dz).$$

This yields the conjugate formula

$$(\mu, K) := (\mu_0, M_{0,n}) \implies K_{\mu}^*(z_n, dz_0) = \mathbb{P}(Z_0 \in dz_0 \mid Z_n = z_n) = \frac{\mu(dz_0) K(z_0, dz_n)}{(\mu K)(dz_0)}.$$

2.5 Continuous vs discrete time models

We illustrate the conjugate formulae presented in Section 2.2 and Section 2.3 in the context of diffusion flows, due to the practical relevance of such methods.

Consider the stochastic flow $s \in [0, t] \mapsto Y_s^{(0)}(x)$ on some *fixed time horizon* t starting at $Y_0^{(0)}(x) = x \in \mathbb{X} = \mathbb{R}^d = \mathbb{Y}$ defined by the diffusion

$$dZ_s^{(0)}(x) = b_s^{(0)}(Z_s^{(0)}(x))ds + \sigma_s(Z_s^{(0)}(x))^{1/2} dB_s \quad (2.10)$$

for some drift function $b_s(x)$, some diffusion function $\sigma_s(x) \in \mathcal{S}_d^+$ and some d -dimensional Brownian motion B_s . Formally, we may slice the time interval $[0, t]_\Delta := \{s_0, \dots, s_n\}$ via some time mesh $s_{i+1} = s_i + \Delta$ from $s_0 = 0$ to $s_n = t = n\Delta$. In this notation, for $s = s_k \in [0, t]_\Delta$ with $k < n$ we compute $Z_{s+\Delta}^{(0)}(x)$ from $Z_s^{(0)}(x)$ using the formula

$$Z_{s+\Delta}^{(0)}(x) - Z_s^{(0)}(x) = b_s^{(0)}(Z_s^{(0)}(x)) \Delta + \sigma_s(Z_s^{(0)}(x))^{1/2} (B_{s+\Delta} - B_s)$$

i.e. via Euler-Maruyama time discretization.

Given some initial random variable X_0 with distribution μ independent of the Brownian motion B_s we set

$$Y_0 := Z_t^{(0)}(X_0) \implies \mathcal{P}_0(d(x, y)) := \mathbb{P}((X_0, Y_0) \in d(x, y)) = \mu(dx) \mathcal{K}_0(x, dy)$$

with the Markov transition

$$\mathcal{K}_0(x, dy) := \mathbb{P}(Z_t^{(0)}(x) \in dy) := q_t^{(0)}(x, y) dy.$$

Observe that

$$\begin{aligned} (\mu \mathcal{K}_0)(dy) &= \mathbb{P}(Y_0 \in dy) \\ &= \mathbb{P}(Z_t^{(0)}(X_0) \in dy) = p_t^{(0)}(y) dy \quad \text{with } \forall s \in [0, t] \quad p_s^{(0)}(y) := \int \mu(dx) q_s^{(0)}(x, y). \end{aligned}$$

The conjugate transition $\mathcal{K}_1$ is associated with the diffusion flow $s \in [0, t] \mapsto Z_s^{(1)}(y)$ starting at $Z_t^{(1)}(y) = y$ defined by the backward stochastic differential equation

$$-d_s Z_s^{(1)}(y) = b_s^{(1)}(Z_s^{(1)}(y)) ds + \sigma_s(Z_s^{(1)}(y))^{1/2} dB_s$$

with the drift function

$$b_s^{(1)} := -b_s^{(0)} + \nabla_{\sigma_s} \log p_s^{(0)} \quad \text{and} \quad (\nabla_{\sigma_s} \log p_s^{(0)})^j := \frac{1}{p_s^{(0)}} \sum_{1 \leq i \leq d} \partial_{x_i} \left(\sigma_s^{j,i} p_s^{(0)} \right).$$

Formally, slicing the interval as above, for $s = s_k \in [0, t]_\Delta$ with $k > 0$ we compute $Z_{s-\Delta}^{(0)}(x)$ from $Z_s^{(0)}(x)$ using the backward formula

$$Z_{s-\Delta}^{(1)}(y) - Z_s^{(1)}(y) = b_s^{(1)}(Z_s^{(1)}(y)) \Delta + \sigma_s(Z_s^{(1)}(y))^{1/2} (B_s - B_{s-\Delta})$$

We have the dual formulae

$$\mu(dx) \mathcal{K}_0(x, dy) = (\mu \mathcal{K}_0)(dy) \mathcal{K}_1(y, dx) \quad \text{and} \quad \mathcal{K}_1(y, dx) := \mathbb{P}(Z_0^{(1)}(y) \in dx).$$

A detailed proof of the above backward conjugate formula can be found in [6, 7, 78]. At the next iteration, we consider some initial random variable Y_1 with distribution η independent of the Brownian motion B_s and we set

$$X_1 := Z_0^{(1)}(Y_1) \implies \mathcal{P}_1(d(x, y)) := \mathbb{P}((X_1, Y_1) \in d(x, y)) = \eta(dy) \mathcal{K}_1(y, dx).$$

2.6 Some half-bridge entropy formulae

We have the entropy formulae

$$\mathcal{H}((\eta \times L)^\flat \mid \mu \times K) = \mathcal{H}(\eta \times L \mid (\mu K) \times K_\mu^*) = \mathcal{H}((\eta L) \times L_\eta^* \mid \mu \times K).$$

This yields the variational formulae

$$(\eta \times K_\mu^*)^\flat = \arg \min_{Q \in \Pi(\cdot, \eta)} \mathcal{H}(Q \mid \mu \times K) \quad \text{and} \quad \mu \times L_\eta^* = \arg \min_{Q \in \Pi(\mu, \cdot)} \mathcal{H}((\eta \times L)^\flat \mid Q). \quad (2.11)$$

By symmetry arguments, we also have the entropy formulae

$$\mathcal{H}(\mu \times K \mid (\eta \times L)^\flat) = \mathcal{H}((\mu K) \times K_\mu^* \mid \eta \times L) = \mathcal{H}(\mu \times K \mid (\eta L) \times L_\eta^*)$$

This yields the variational formulae

$$(\eta \times K_\mu^*)^\flat = \arg \min_{Q \in \Pi(\cdot, \eta)} \mathcal{H}(\mu \times K \mid Q) \quad \text{and} \quad \mu \times L_\eta^* = \arg \min_{Q \in \Pi(\mu, \cdot)} \mathcal{H}(Q \mid (\eta \times L)^\flat). \quad (2.12)$$

Applying the left hand side (l.h.s.) of (2.11) and (2.12) to $K = \mathcal{K}_{2n}$ we have the half-bridge equivalent formulation

$$\mathcal{P}_{2n+1} := \arg \min_{Q \in \Pi(\cdot, \eta)} \mathcal{H}(Q \mid \mathcal{P}_{2n}) = \arg \min_{Q \in \Pi(\cdot, \eta)} \mathcal{H}(\mathcal{P}_{2n} \mid Q).$$

Similarly, applying the right hand side (r.h.s.) of (2.11) and (2.12) to $L = \mathcal{K}_{2n+1}$ we also have

$$\mathcal{P}_{2(n+1)} := \arg \min_{Q \in \Pi(\mu, \cdot)} \mathcal{H}(\mathcal{P}_{2n+1} \mid Q) = \arg \min_{Q \in \Pi(\mu, \cdot)} \mathcal{H}(Q \mid \mathcal{P}_{2n+1}).$$

3 Stability of Markov operators

3.1 Overview

The time varying Gibbs-type semigroup formulation of Sinkhorn bridges (1.9) and the entropy formula (1.10) show that the convergence analysis of Sinkhorn bridges rely on the stability analysis of a class of time varying Markov processes sharing a common invariant target measure at each individual step. This section gives a summary on the stability analysis of Markov operators. Section 3.2 and Section 3.3 discuss the different classes of entropy, divergences and the Kantorovich criteria used in the article. Section 3.4 review an operator-theoretic framework based on contraction coefficients and Lyapunov principles. Section 3.5 is dedicated to transportation cost inequalities including local log-Sobolev and Talagrand quadratic cost inequalities.

3.2 Entropy and divergence criteria

A measure μ_1 on $\mathbb{X}$ is said to be absolutely continuous with respect to another measure μ_2 on $\mathbb{X}$, and we write $\mu_1 \ll \mu_2$, if $\mu_2(A) = 0$ implies $\mu_1(A) = 0$ for any measurable subset

$A \subset \mathbb{X}$. Whenever $\mu_1 \ll \mu_2$, we denote by $d\mu_1/d\mu_2$ the Radon-Nikodym derivative of μ_1 w.r.t. μ_2 . Consider a convex function

$$\begin{aligned} \Phi : \quad \mathbb{R}_+^2 &\mapsto \mathbb{R} \cup \{+\infty\} \\ (u, v) &\rightsquigarrow \Phi(u, v) \quad \text{s.t. } \forall a \in \mathbb{R}_+ \quad \Phi(au, av) = a \Phi(u, v) \text{ and } \Phi(1, 1) = 0. \end{aligned} \quad (3.1)$$

The Φ -entropy between two probability measures μ_1, μ_2 on $\mathbb{X}$ is defined for any dominating measure γ (such that $\gamma \gg \mu_1$ and $\gamma \gg \mu_2$) by the formula

$$\mathcal{H}_\Phi(\mu_1, \mu_2) := \int \Phi \left(\frac{d\mu_1}{d\gamma}(x), \frac{d\mu_2}{d\gamma}(x) \right) \gamma(dx). \quad (3.2)$$

By homogeneity arguments, the definition (3.2) does not depend on the choice of the dominating measure γ . Note that

$$\text{if } \Phi(u, v) = \Phi_0(u, v) := \frac{1}{2} |u - v| \quad \text{then} \quad \mathcal{H}_{\Phi_0}(\mu_1, \mu_2) = \|\mu_1 - \mu_2\|_{\text{tv}}. \quad (3.3)$$

In the above display, $\|\mu_1 - \mu_2\|_{\text{tv}}$ stands for the total variation distance between the probability measures μ_1 and μ_2 defined by

$$\begin{aligned} \|\mu_1 - \mu_2\|_{\text{tv}} &:= \frac{1}{2} \sup \{ |(\mu_1 - \mu_2)(f)| : f \in \mathcal{B}(\mathbb{X}) ; \|f\| \leq 1 \} \\ &= \sup \{ |(\mu_1 - \mu_2)(f)| : f \in \mathcal{B}(\mathbb{X}) ; \text{osc}(f) \leq 1 \} \end{aligned} \quad (3.4)$$

with the oscillation $\text{osc}(f)$ of the function $f \in \mathcal{B}(\mathbb{X})$ defined by

$$\text{osc}(f) := \sup \{ |f(x_1) - f(x_2)| : (x_1, x_2) \in \mathbb{X}^2 \}.$$

Whenever $\Phi(1, 0) = \infty$, the Φ -entropy coincides with the ϕ -divergence in the sense of Csiszár [38], i.e., we also have

$$\mathcal{H}_\Phi(\mu_1, \mu_2) = \int \phi \left(\frac{d\mu_1}{d\mu_2}(x) \right) \mu_2(dx) \quad \text{where} \quad \phi(u) := \Phi(u, 1)$$

and we have used the convention $\mathcal{H}_\Phi(\mu_1, \mu_2) = \infty$ when $\mu_1 \not\ll \mu_2$. Therefore, the Φ -entropy generalizes the classical relative entropy (a.k.a. Kullback-Leibler divergence or I -divergence). More precisely, choosing the functions

$$\Phi(u, v) = u \log(u/v) \quad \text{and} \quad \phi(u) = u \log u$$

we recover the classical formulae

$$\mathcal{H}_{\Phi_1}(\mu_1, \mu_2) = \mathcal{H}(\mu_1 \mid \mu_2) := \mu_1 \left(\log \left(\frac{d\mu_1}{d\mu_2} \right) \right).$$

The ϕ -divergence criteria discussed above also includes $\mathbb{L}_p$ -norms, Havrda-Charvat entropies, α -divergences, Jensen-Shannon divergence, squared Hellinger integrals and Kakutani-Hellinger, as well as Jeffreys and Rényi divergences, see for instance [38, 53, 89, 96] and references therein.

3.3 Kantorovich semi-distances

Consider a semi-distance φ on a Polish space $\mathbb{X}$; a semi-distance is essentially a distance function that is l.s.c. except that the triangle inequality may be violated. Some examples of semi-distances that we consider include the discrete metric $\varphi_d(x_1, x_2) := 1_{x_1 \neq x_2}$ and, for a given l.s.c. function g , lower bounded away from zero, the weighted discrete metric defined by the formulae

$$\varphi_g(x_1, x_2) := \varphi_d(x_1, x_2) w_g(x_1, x_2) \quad \text{with} \quad w_g(x_1, x_2) = g(x_1) + g(x_2). \quad (3.5)$$

The Kantorovich semi-distance between two measures $\mu_1, \mu_2 \in \mathbf{Prob}(\mathbb{X})$ is defined by

$$D_\varphi(\mu_1, \mu_2) := \inf_{Q \in \Pi(\mu_1, \mu_2)} Q(\varphi). \quad (3.6)$$

The assumption that $\mathbb{X}$ is a complete and separable metric space and the fact that φ is l.s.c. and non-negative ensure the existence of an optimal transport plan [119, Theorem 4.1]. Moreover, the assumption that $\varphi(x_1, x_2) = 0 \Leftrightarrow x_1 = x_2$ ensure that $D_\varphi(\mu_1, \mu_2) = 0$ if and only if $\mu_1 = \mu_2$. The Kantorovich semi-distance is related to several well-known measures of distance between probability distributions. Specifically, when $(\mathbb{X}, \varphi)$ is a Polish space the quantity $D_\varphi(\mu_1, \mu_2)$ is the Kantorovich distance (or Wasserstein 1-distance) between μ_1 and μ_2 , whilst for $p \geq 1$ the Wasserstein p -distance $\mathcal{D}_{\varphi, p}(\mu_1, \mu_2)$ is defined by $\mathcal{D}_{\varphi, p}(\mu_1, \mu_2) = D_{\varphi^p}(\mu_1, \mu_2)^{1/p}$. We also have the formulae

$$D_{\varphi_d}(\mu_1, \mu_2) = \|\mu_1 - \mu_2\|_{\text{tv}} \quad \text{as well as} \quad D_{\varphi_g}(\mu_1, \mu_2) = \|\mu_1 - \mu_2\|_g$$

with the g -weighted total variation norm

$$\begin{aligned} \|\mu_1 - \mu_2\|_g &:= |(\mu_1 - \mu_2)(g)| \\ &= \sup \{ |(\mu_1 - \mu_2)(f)| : f \in \mathcal{B}_g(\mathbb{X}) ; \|f\|_g \leq 1 \} \\ &= \sup \{ |(\mu_1 - \mu_2)(f)| : f \in \mathcal{B}_g(\mathbb{X}) ; \text{osc}_g(f) \leq 1 \}, \end{aligned} \quad (3.7)$$

where $\text{osc}_g(f)$ stands for the φ_g -oscillation of the function $f \in \mathcal{B}_g(\mathbb{X})$ defined by the Lipschitz constant

$$\text{osc}_g(f) := \sup \{ |f(x_1) - f(x_2)| / \varphi_g(x_1, x_2) : (x_1, x_2) \in \mathbb{X}^2 ; x_1 \neq x_2 \}.$$

When $\mathbb{X} = \mathbb{R}^d$ and $\varphi(x_1, x_2) = \|x_1 - x_2\|_2$ is the Euclidian norm we simplify notation and we write $\mathcal{D}_p$ instead of $\mathcal{D}_{\varphi, p}$; that is

$$\mathcal{D}_p(\mu_1, \mu_2) := \inf_{Q \in \Pi(\mu_1, \mu_2)} \left(\int \|x_1 - x_2\|_2^p Q(d(x_1, x_2)) \right)^{1/p}. \quad (3.8)$$

For a more thorough discussion on Kantorovich criteria and weighted total variation norms we refer to [51, 54] and references therein.

3.4 Contraction coefficients

3.4.1 Dobrushin coefficient

Given a convex function Φ satisfying (3.1), the Dobrushin Φ -contraction coefficient $\text{dob}_\Phi(K)$ of some Markov transition $K(x, dy)$ from $\mathbb{X}$ into $\mathbb{Y}$ is defined by

$$\text{dob}_\Phi(K) := \sup \frac{\mathcal{H}_\Phi(\mu_1 K, \mu_2 K)}{\mathcal{H}_\Phi(\mu_1, \mu_2)}, \quad (3.9)$$

where the supremum is taken over all probability measures μ_1, μ_2 such that $\mathcal{H}_\Phi(\mu_1, \mu_2) > 0$. Here, again, the terminology Φ -Dobrushin coefficient comes from the fact that choosing $\Phi(u, v) = \Phi_0(u, v) = \frac{1}{2}|u - v|$ we recover the standard Dobrushin contraction coefficient, i.e.,

$$\begin{aligned} \text{dob}_{\Phi_0}(K) = \text{dob}(K) &= \sup_{(x_1, x_2) \in \mathbb{X}^2} \|\delta_{x_1} K - \delta_{x_2} K\|_{\text{tv}} \\ &= \sup \{\text{osc}(K(f)) : f \in \mathcal{B}(\mathbb{Y}) ; \text{osc}(f) \leq 1\}. \end{aligned} \quad (3.10)$$

We also recall that the Dobrushin coefficient is a universal contraction coefficient, that is we have

$$\text{dob}_\Phi(K) \leq \text{dob}(K) := \text{dob}_{\Phi_0}(K). \quad (3.11)$$

A proof of the above estimate is provided in [31] for finite state spaces and in [53] for general measurable spaces. For further details on the Dobrushin contraction coefficient we refer the reader to [52, 53, 58] and [45, Chapter 4].

3.4.2 Kantorovich Lipschitz norm

Consider a Markov integral operator K and some other l.s.c. function h , lower bounded away from zero on $\mathbb{Y}$ and such that $\|K(h)\|_g = \|K(h)/g\| < \infty$. This condition ensures that the right-action linear operator $f \in \mathcal{B}_h(\mathbb{Y}) \mapsto K(f) \in \mathcal{B}_g(\mathbb{X})$ is continuous with operator norm defined by

$$\|K\|_{g,h} := \|K(h)/g\| = \sup \{\|K(f)\|_g : f \in \mathcal{B}_h(\mathbb{Y}) ; \|f\|_h \leq 1\} < \infty. \quad (3.12)$$

The Lipschitz constant or Kantorovich norm [59] of the integral left-action operator $\mu \in \text{Prob}_g(\mathbb{X}) \mapsto \mu K \in \text{Prob}_h(\mathbb{Y})$ is defined by

$$\text{lip}_{g,h}(K) := \sup \frac{D_{\varphi_h}(\mu_1 K, \mu_2 K)}{D_{\varphi_g}(\mu_1, \mu_2)} = \sup \frac{\|\mu_1 K - \mu_2 K\|_h}{\|\mu_1 - \mu_2\|_g} \quad (3.13)$$

where the supremum is taken over all $(\mu_1, \mu_2) \in \text{Prob}_g(\mathbb{X})^2$ such that $\|\mu_1 - \mu_2\|_g > 0$. We have the equivalent dual formulation

$$\text{lip}_{g,h}(K) = \sup \{\text{osc}_g(K(f)) : f \in \mathcal{B}_h(\mathbb{Y}) ; \text{osc}_h(f) \leq 1\} = \sup \frac{D_{\varphi_h}(\delta_{x_1} K, \delta_{x_2} K)}{D_{\varphi_g}(\delta_{x_1}, \delta_{x_2})} \quad (3.14)$$

where the supremum is taken over all states $(x_1, x_2) \in \mathbb{X}^2$ such that $\varphi_g(x_1, x_2) > 0$. For a more detailed discussion on these equivalent formulations see [51] or [54, Section 8.2].

3.4.3 Lyapunov methodologies

Consider a Markov transition $L(y, dx)$ from $\mathbb{Y}$ into $\mathbb{X}$. Assume that

$$K(h) \leq \epsilon g + c \quad \text{and} \quad L(g) \leq \epsilon h + c \quad (3.15)$$

for some constants $\epsilon \in]0, 1[$ and $c > 0$ and some l.s.c. functions g, h lower bounded away from zero. Also assume there exists a map $\iota : l \in [0, \infty[\mapsto \iota(l) \in]0, 1[$, such that for any $w_g(x_1, x_2) \leq l$ and $w_h(y_1, y_2) \leq l$ we have

$$\|\delta_{x_1} K - \delta_{x_2} K\|_{\text{tv}} \vee \|\delta_{y_1} L - \delta_{y_2} L\|_{\text{tv}} \leq 1 - \iota(l). \quad (3.16)$$

Condition (3.16) is equivalent to the local minorization condition

$$\delta_{x_1} K \wedge \delta_{x_2} K \geq \iota(l) \eta \quad \text{and} \quad \delta_{y_1} L \wedge \delta_{y_2} L \geq \iota(l) \mu \quad (3.17)$$

for some probability measures (μ, η) on $(\mathbb{X}, \mathbb{Y})$ that may depend on the parameters l and the states (x_1, x_2) and (y_1, y_2) in the sub-level sets $w_g(x_1, x_2) \leq l$ and $w_h(y_1, y_2) \leq l$ (cf. for instance [9, 48, 54]). We have

$$\{w_g \leq l\} \subset \{g \leq l\}^2 \quad \text{and} \quad \{w_h \leq l\} \subset \{h \leq l\}^2.$$

Thus, (3.17) is met as soon as

$$\forall (x, y) \in (\{g \leq l\} \times \{h \leq l\}) \quad \delta_x K \geq \iota(l) \eta \quad \text{and} \quad \delta_y L \geq \iota(l) \mu. \quad (3.18)$$

We recall the notation $\mathcal{B}_\infty(\mathbb{X})$ and $\mathcal{B}_0(\mathbb{X})$ from section 1.5. It should be noted that, in (3.15), the functions g, h are not necessarily proper Lyapunov functions, in the sense that g and h are not assumed to belong to the sub-algebrae $\mathcal{B}_\infty(\mathbb{X})$ and $\mathcal{B}_\infty(\mathbb{Y})$ (recall the collection of locally upper bounded and uniformly positive functions that growth at infinity (in the sense that their sub-level sets are compact)). When g, h are bounded (3.17) are minorization conditions on the whole state spaces. Conversely, when $(g, h) \in (\mathcal{B}_\infty(\mathbb{X}) \times \mathcal{B}_\infty(\mathbb{Y}))$ condition (3.17) are local minorization conditions on compact sub-level sets.

We note that if a function $h > 0$ satisfies

$$K(h)/g \leq \theta \in \mathcal{B}_0(\mathbb{X}) \quad (3.19)$$

then, for any $0 < \epsilon < \|\theta\| \wedge 1$, we have

$$K(h) \leq \epsilon 1_{\{\theta < \epsilon\}} g + 1_{\{\theta \geq \epsilon\}} \theta g \leq \epsilon g + c \quad \text{with} \quad c := \|\theta\| \sup_{x \in \{\theta \geq \epsilon\}} g(x). \quad (3.20)$$

Consequently, the l.h.s. drift condition in (3.15) is satisfied.

Note the rescaling properties

$$\begin{aligned} \bar{g} &:= \frac{1}{2} + \frac{\epsilon}{2c} g \quad \text{and} \quad \bar{h} := \frac{1}{2} + \frac{\epsilon}{2c} h \\ \implies K(\bar{h}) &\leq \epsilon \bar{g} + \frac{1}{2} \quad \text{and} \quad L(\bar{g}) \leq \epsilon \bar{h} + \frac{1}{2}. \end{aligned} \quad (3.21)$$

In addition, for any $a \geq 0$, we also have

$$\frac{1}{2} + a \bar{g} = \frac{a+1}{2} \left(1 + \frac{\epsilon}{c} \frac{a}{a+1} g \right) \quad \text{and} \quad \frac{1}{2} + a \bar{h} = \frac{a+1}{2} \left(1 + \frac{\epsilon}{c} \frac{a}{a+1} h \right).$$

The following theorem is a slight variation of contraction estimates presented in [9, 51] (itself based on the weighted norm contraction analysis presented in [48, 54]).

Theorem 3.1. *Assume (3.15) and (3.16) are satisfied. Then there exists some $a > 0$ and $\varrho \in]0, 1[$ that only depends on the parameters (ϵ, c) and the function $\iota(r)$ such that*

$$\text{lip}_{g_a, h_a}(K) \vee \text{lip}_{h_a, g_a}(L) \leq \varrho \quad \text{with} \quad g_a := \frac{1}{2} + a g \quad \text{and} \quad h_a := \frac{1}{2} + a h,$$

In addition, we have

$$\text{lip}_{g_a, g_a}(KL) \vee \text{lip}_{h_a, h_a}(LK) \leq \text{lip}_{g_a, h_a}(K) \text{lip}_{h_a, g_a}(L) \leq \varrho^2.$$

Observe that

$$\varphi_{g_a}(x_1, x_2) := \varphi_d(x_1, x_2) (1 + a w_g(x_1, x_2)).$$

Assume there exists a semi-distance ψ on $\mathbb{X}$ such that

$$\psi(x_1, x_2) \leq w_g(x_1, x_2) \implies a \psi(x_1, x_2) \leq \varphi_{g_a}(x_1, x_2) \implies a D_\psi(\mu_1, \mu_2) \leq \|\mu_1 - \mu_2\|_{g_a}.$$

In this situation, Theorem 3.1 yields the Kantorovich semi-distance exponential decays

$$a D_\psi(\mu_1(KL)^n, \mu_2(KL)^n) \leq \varrho^{2n} \|\mu_1 - \mu_2\|_{g_a}. \quad (3.22)$$

When $(\mathbb{X}, \psi)$ is a Polish space, the above estimates yield the exponential decays w.r.t. the Wasserstein 1-distance $\mathcal{D}_\psi$. By [51, Theorem 2.3], under the assumptions of Theorem 3.1, there exists some $b > 0$ such that

$$\psi_{a,b} := \varphi_{g_a} + b \psi \implies \text{lip}_{\psi_{a,b}, \psi_{a,b}}(KL) \leq 1 - (1 - \text{lip}_{g_a, g_a}(KL))^2.$$

Example 3.2. *For Lyapunov functions on $\mathbb{X} = \mathbb{R}^d$ of the form*

$$g(x) := c_1 \exp(c_2 \|x\|^p)$$

for constants $c_1, c_2 > 0$ and some parameter $p > 0$ we have

$$\begin{aligned} w_g(x_1, x_2) &\geq 2c_1 e^{c_2 \|x_1\|^{p/2}} e^{c_2 \|x_2\|^{p/2}} \geq 2c_1 \left(\exp\left(\frac{c_2}{2} (\|x_1\|^p + \|x_2\|^p)\right) - 1 \right) \\ &\geq \psi_{\exp}(x_1, x_2) := 2c_1 \left(\exp\left(\frac{c_2}{2^{1 \vee p}} \|x_1 - x_2\|^p\right) - 1 \right). \end{aligned}$$

The first estimate comes from the fact that $u^2 + v^2 \geq 2uv$ for any $u, v \in \mathbb{R}$. The last estimate follows from the fact that

$$\|x_1\|^p + \|x_2\|^p \geq 2^{-(p-1)+} \|x_1 - x_2\|^p.$$

Also note that for any $q \geq 1$ we have

$$\psi_{\exp}(x_1, x_2) \geq \psi_{p,q}(x_1, x_2) := c_{p,q} \|x_1 - x_2\|^{pq} \quad \text{with} \quad c_{p,q} := \frac{2c_1}{q!} \left(\frac{c_2}{2^{1 \vee p}} \right)^q.$$

In this case, in terms of the Wasserstein metric (3.8) we have

$$a c_{p,q} \mathcal{D}_{pq}(\mu_1, \mu_2)^{pq} \leq \|\mu_1 - \mu_2\|_{g_a}.$$

3.5 Transportation cost inequality

3.5.1 Log-Sobolev constants

We say that a probability measure μ_2 satisfies the quadratic transportation cost inequality $\mathbb{T}_2(\rho)$ with constant $\rho > 0$ if for any probability μ_1 we have

$$\frac{1}{2} \mathcal{D}_2(\mu_1, \mu_2)^2 \leq \rho \mathcal{H}(\mu_1 \mid \mu_2). \quad (3.23)$$

We say that a Markov transition K from $\mathbb{X}$ into $\mathbb{Y}$ satisfies the quadratic transportation cost inequality $\mathbb{T}_2(\rho)$ when $\delta_x K$ satisfies a quadratic transportation cost inequality $\mathbb{T}_2(\rho)$ for any $x \in \mathbb{X}$.

Lemma 3.3 ([49, 50]). *The Markov transition K satisfies the $\mathbb{T}_2(\rho)$ inequality if and only if for any probability distribution μ and any Markov transition L we have*

$$\mathcal{D}_2(\mu L, \mu K)^2 \leq \int \mu(dx) \mathcal{D}_2(\delta_x L, \delta_x K)^2 \leq 2\rho \mathcal{H}(\mu \times L \mid \mu \times K). \quad (3.24)$$

In addition, we have the bias estimate

$$\int \mu(dx) \|\text{mean}_K(x) - \text{mean}_L(x)\|_2^2 \leq 2\rho \mathcal{H}(\mu \times L \mid \mu \times K)$$

as well as the covariance entropic bounds

$$\int \mu(dx) \|\text{cov}_K(x) - \text{cov}_L(x)\|_F \leq 4\rho \mathcal{H}(\mu \times L \mid \mu \times K) + c_{\mu, L} (8\rho \mathcal{H}(\mu \times L \mid \mu \times K))^{1/2} \quad \text{with} \quad c_{\mu, L}^2 := \mu(\text{Tr}(\text{cov}_L))$$

When $\mathbb{X} = \mathbb{R}^d$ and $d\mu_1/d\mu_2$ is differentiable the (relative) Fisher information of μ_1 with respect to μ_2 is defined by

$$\mathcal{J}(\mu_1 \mid \mu_2) = \mu_1(\|\nabla \log(d\mu_1/d\mu_2)\|_2^2)$$

We also use the convention $\mathcal{J}(\mu_1 \mid \mu_2) = \infty$ when $\mu_1 \not\ll \mu_2$.

A probability measure μ_2 satisfies the log-Sobolev inequality $LS(\rho)$ if for any probability μ_1 we have

$$\mathcal{H}(\mu_1 \mid \mu_2) \leq \frac{\rho}{2} \mathcal{J}(\mu_1 \mid \mu_2). \quad (3.25)$$

In addition, a Markov transition $K(x, dy)$ from $\mathbb{R}^d$ into itself satisfies the log-Sobolev inequality $LS(\rho)$ when $\delta_x K$ satisfies a log-Sobolev inequality $LS(\rho)$ for any $x \in \mathbb{R}^d$. Finally, a Markov transition K satisfies the local log-Sobolev inequality $LS_{\text{loc}}(\rho)$ if for any $x_1, x_2 \in \mathbb{R}^d$ we have the inequalities

$$\mathcal{H}(\delta_{x_1} K \mid \delta_{x_2} K) \leq \frac{\rho}{2} \mathcal{J}(\delta_{x_1} K \mid \delta_{x_2} K). \quad (3.26)$$

A theorem by Otto and Villani ([99, Theorem 1]) ensures that the log-Sobolev inequality $LS(\rho)$ implies the quadratic transportation cost inequality $\mathbb{T}_2(\rho)$.

We also say that K satisfies the Fisher-Lipschitz inequality $\mathbb{J}(\kappa)$ with constant $\kappa > 0$ if for any $x_1, x_2 \in \mathbb{R}^d$ we have the inequalities

$$\mathcal{J}(\delta_{x_1} K \mid \delta_{x_2} K) \leq \kappa^2 \|x_1 - x_2\|_2^2. \quad (3.27)$$

3.5.2 Some illustrations

Several criteria with explicit estimates of the log-Sobolev constants of Gibbs measures with non uniformly convex potentials are available in the literature under rather mild conditions, including the celebrated Bakry-Emery criteria on Riemannian manifolds [12]. It is clearly out of the scope of the present article to review these estimates, see for instance [13, 20, 21, 94, 98, 118, 119] and references therein.

Consider a Boltzmann-Gibbs probability measure on $\mathbb{R}^d$ of the form

$$\lambda_U(dx) = e^{-U(x)} dx$$

for some potential function $U(x)$ such that

$$\nabla^2 U(x) \geq a \mathbf{1}_{\|x\|_2 \geq l} I - b \mathbf{1}_{\|x\|_2 < l} I \quad (3.28)$$

for some $a > 0$, $b \geq 0$ and $l \geq 0$. In this context there exists some $\bar{U}$ with bounded oscillations such that $\nabla^2(U + \bar{U}) \geq (a/2) I$. By the Holley-Stroock perturbation lemma [82], λ_U satisfies the $LS(\rho)$ -inequality (3.25) with some parameter $\rho = \rho(a, b, l)$ that depends on (a, b, l) . For an explicit construction of such function we refer to [93]. As an example, the one-dimensional double-well potential function $U(x) = x^4 - a x^2$, with $a > 0$ are convex at infinity and thus satisfy the $LS(\rho)$ -inequality for some $\rho > 0$.

The local Hessian condition (3.28) is often termed “strongly convexity outside a ball” or in short “convex-at-infinity”. To avoid confusion, we underline that this condition strongly differs from the strong convexity of U on the non convex domain $\{x \in \mathbb{R}^d : \|x\|_2 \geq r\}$ discussed in [88, 101, 122]. It is also well known (see for instance Corollary 9.3.2 and Corollary 5.7.2 in [13]) that

$$\begin{aligned} \exists \sigma \in \mathcal{S}_d^+ \quad \text{such that} \quad \forall x \in \mathbb{R}^d \quad \nabla^2 U(x) \geq \sigma^{-1} \\ \implies \lambda_U \text{ satisfy the } LS(\rho) \text{ inequality with } \rho = \|\sigma\|_2. \end{aligned} \quad (3.29)$$

We also refer to [94] for a variety of log-Sobolev criteria when the curvature is not positive everywhere based on Holley-Stroock and Aida-Shigekawa perturbation arguments [2, 82].

We will now illustrate the transportation cost inequalities discussed above in the context of Gaussian models.

Denote by $\nu_{m,\sigma}$ the Gaussian distribution on $\mathbb{X} = \mathbb{R}^d$ with mean $m \in \mathbb{R}^d$ and covariance matrix $\sigma \in \mathcal{S}_d^+$. In addition, let gauss_σ denote the probability density function of the distribution $\nu_{0,\sigma}$, with covariance matrix $\sigma \in \mathcal{S}_d^+$. In this context, we have

$$\nu_{m,\sigma} = \lambda_U \quad \text{with} \quad U(x) = -\log \text{gauss}_\sigma(x - m) \implies \nabla^2 U(x) = \sigma^{-1}.$$

For any parameters $(m, \bar{\sigma}) \in (\mathbb{R}^d \times \mathcal{S}_d^+)$ we recall that

$$2 \mathcal{H}(\nu_{m,\sigma} \mid \nu_{\bar{m},\bar{\sigma}}) = D(\sigma \mid \bar{\sigma}) + \|\bar{\sigma}^{-1/2} (m - \bar{m})\|_2^2 \quad (3.30)$$

with the Burg (or log-det) divergence

$$D_{\text{burg}}(\sigma \mid \bar{\sigma}) := \text{Tr}(\sigma \bar{\sigma}^{-1} - I) - \log \det(\sigma \bar{\sigma}^{-1}). \quad (3.31)$$

Let $(\alpha, \beta, \tau) \in (\mathbb{R}^d \times \mathcal{G}l_d \times \mathcal{S}_d^+)$, we have the linear Gaussian transition

$$\begin{aligned} K(x, dy) &= e^{-W(x, y)} dy \quad \text{with} \quad W(x, y) = -\log \text{gauss}_\tau(y - (\alpha + \beta x)) \\ \implies \nabla_2 W(x, y) &= \tau^{-1}(y - (\alpha + \beta x)) \quad \text{and} \quad \nabla_2^2 W(x, y) = \tau^{-1} \geq \|\tau\|_2^{-1} I. \end{aligned} \quad (3.32)$$

The curvature estimate in the above display ensures that K satisfies the $LS(\rho)$ -inequality with $\rho = \|\tau\|_2$. Observe that

$$2 \mathcal{H}(\delta_{x_1} K \mid \delta_{x_2} K) = (x_1 - x_2)' \left((\tau^{-1/2} \beta)' (\tau^{-1/2} \beta) \right) (x_1 - x_2).$$

On the other hand, we have

$$\begin{aligned} \mathcal{J}(\delta_{x_1} K \mid \delta_{x_2} K) &= \int K(x_1, dy) \|\nabla_2 W(x_1, y) - \nabla_2 W(x_2, y)\|^2 \\ &= (x_1 - x_2)' \left((\tau^{-1/2} \beta)' \tau^{-1} (\tau^{-1/2} \beta) \right) (x_1 - x_2). \end{aligned}$$

We recover the fact that K satisfies the $LS_{\text{loc}}(\|\tau\|_2)$ -inequality and the $\mathbb{J}(\|\tau^{-1} \beta\|_2)$ inequality.

3.5.3 A stability theorem

For any probability measures μ_1 and μ_2 on $\mathbb{X}$ and a Markov transition $K(x, dy)$ from $\mathbb{X}$ into $\mathbb{Y}$ we have

$$\begin{aligned} \mathcal{H}(\mu_1 \mid \mu_2) &= \mathcal{H}(\mu_1 \times K \mid \mu_2 \times K) = \mathcal{H}((\mu_1 K) \times K_{\mu_1}^* \mid (\mu_2 K) \times K_{\mu_2}^*) \\ &= \mathcal{H}(\mu_1 K \mid \mu_2 K) + \int (\mu_1 K)(dy) \mathcal{H}(\delta_y K_{\mu_1}^* \mid \delta_y K_{\mu_2}^*) \end{aligned}$$

and hence that

$$\mathcal{H}(\mu_1 K \mid \mu_2 K) \leq \mathcal{H}(\mu_1 \mid \mu_2). \quad (3.33)$$

In information theory, the above inequality is also known as the data-processing inequality. For any probability measures μ and η on $\mathbb{X}$ and $\mathbb{Y}$, from the convexity of the function $h(u) = u \log u$ we have

$$h\left(\frac{d\mu K}{d\eta}(y)\right) = h\left(\int \mu(dx) \frac{d\delta_x K}{d\eta}(y)\right) \leq \int \mu(dx) h\left(\frac{d\delta_x K}{d\eta}(y)\right).$$

This implies that

$$\begin{aligned} \mathcal{H}(\mu K \mid \eta) &= \int \eta(dy) h\left(\frac{d\mu K}{d\eta}(y)\right) \\ &\leq \int \mu(dx) \int \eta(dy) h\left(\frac{d\delta_x K}{d\eta}(y)\right) = \int \mu(dx) \mathcal{H}(\delta_x K \mid \eta). \end{aligned}$$

We conclude that

$$\mathcal{H}(\mu K \mid \eta) \leq \int \mu(dx) \mathcal{H}(\delta_x K \mid \eta). \quad (3.34)$$

Let μ and η be two probability measures on $\mathbb{X}$ and let $K(x, dy)$ be Markov transition from $\mathbb{X}$ into $\mathbb{Y}$. From the definition of the bridge map given in (1.1), for any reference measure of the form $P := (\mu \times K)$ with some given probability measures μ on $\mathbb{X}$ we have

$$P_{\mu, \mu K} = \arg \min_{Q \in \Pi(\mu, \mu K)} \mathcal{H}(Q|P) = \mu \times K \quad \text{and} \quad P_{\mu, \eta K} := \arg \min_{Q \in \Pi(\mu, \eta K)} \mathcal{H}(Q|P).$$

We are now in position to state and prove the following stability theorem.

Theorem 3.4 ([49]). *Assume that $\mathbb{X} = \mathbb{Y} = \mathbb{R}^d$ and the Markov transition K satisfies the local log-Sobolev inequality $LS_{\text{loc}}(\rho)$ and the Fisher-Lipschitz inequality $\mathbb{J}(\kappa)$ inequality. In this situation, for any reference measure of the form $P := (\mu \times K)$ for some probability measures μ on $\mathbb{R}^d$, and for any probability measure η on $\mathbb{R}^d$ have*

$$\mathcal{H}(\eta K | \mu K) \leq \mathcal{H}(P_{\mu, \eta K} | P_{\mu, \mu K}) \leq \kappa^2 \rho \frac{1}{2} \mathcal{D}_2(\mu, \eta)^2. \quad (3.35)$$

In addition, we have the following estimates:

- When η satisfies the $\mathbb{T}_2(\rho')$ inequality we have

$$\mathcal{H}(P_{\mu, \eta K} | P_{\mu, \mu K}) \leq \kappa^2 \rho \rho' \mathcal{H}(\mu | \eta). \quad (3.36)$$

- When μ satisfies the $\mathbb{T}_2(\rho')$ inequality we have

$$\mathcal{H}(P_{\mu, \eta K} | P_{\mu, \mu K}) \leq \kappa^2 \rho \rho' \mathcal{H}(\eta | \mu). \quad (3.37)$$

- When K satisfies the $\mathbb{T}_2(\rho')$ inequality we have

$$\mathcal{D}_2(\eta K, \mu K)^2 \leq 2\rho' \mathcal{H}(P_{\mu, \eta K} | P_{\mu, \mu K}) \leq \kappa^2 \rho \rho' \mathcal{D}_2(\mu, \eta)^2. \quad (3.38)$$

Proof. Applying (3.34) to a Markov transition T we check that

$$\begin{aligned} \mathcal{H}(\mu \times (TK) | \mu \times K) &= \int \mu(dz) \mathcal{H}((\delta_z T)K | \delta_z K) \\ &\leq \int (\mu \times T)^\flat(d(x, z)) \mathcal{H}(\delta_x K | \delta_z K). \end{aligned}$$

Choosing T such that $\mu T = \eta$ we check that

$$(\mu \times T)^\flat \in \Pi(\eta, \mu) \quad \text{and} \quad (\mu \times (TK)) \in \Pi(\mu, \eta K).$$

On the other hand, by (3.26) we have

$$\mathcal{H}(\delta_{x_1} K | \delta_{x_2} K) \leq \frac{\rho}{2} \kappa^2 \|x_1 - x_2\|_2^2.$$

For any T such that $\mu T = \eta$ we have

$$P := \mu \times K \implies \mathcal{H}(P_{\mu, \eta K} | P_{\mu, \mu K}) \leq \frac{\rho}{2} \kappa^2 \int \mu(dz) T(z, dx) \|z - x\|_2^2.$$

We end the proof of (3.35) choosing the optimal 2-Wasserstein coupling transition. The estimates (3.36) and (3.37) are immediate consequences of the definition of the $\mathbb{T}_2(\rho')$ inequality. The estimate (3.38) is a consequence of Lemma 3.3 and (3.35). $\blacksquare$

4 A brief review on Sinkhorn bridges

4.1 Overview

Entropic optimal transport for density models on Polish spaces is often presented as an optimal transport problem with an entropic regularization. The entropic regularization problem is solved by Sinkhorn iterations for small values of the regularization parameter to solve the initial optimal transport problem. A brief presentation of these models is provided in Section 4.2. Recall that the Sinkhorn bridges introduced in (1.4) are defined sequentially by conjugate formulae and adjoint integral operators (cf. Section 2.2). This probabilistic interpretation allows one to connect the convergence of Sinkhorn equations with the stability analysis of Markov semigroups. In Section 4.3 we briefly review some elementary monotone properties of Sinkhorn semigroup based on the non expansive property of Markov operators (3.33) w.r.t. the relative entropy criteria. In the context of density models, Schrödinger and Sinkhorn bridges can alternatively be defined by a sequence of entropic-potential functions. These entropic potentials are defined in terms of a system of recursive equations. These dual-type formulations are presented in Section 4.4. Section 4.5 and Section 4.6 provide some entropic projection properties of bridge maps and present some pivotal commutation-type integral transport formulae. Section 4.7 underlines some links between the gradient and Hessian of entropic-potential functions and the conditional mean and covariance of Sinkhorn transitions.

4.2 Entropy-regularized transport

Consider the marginal measures $(\mu, \eta) = (\lambda_U, \nu_V)$ on locally compact Polish spaces $(\mathbb{X}, \lambda)$ and $(\mathbb{Y}, \nu)$ defined in (2.7). Denote K a Markov transition from $\mathbb{X}$ into $\mathbb{Y}$ of the form

$$K(x, dy) = e^{-W(x, y)} \nu(dy) \quad (4.1)$$

for some transition potential function $W : (x, y) \in (\mathbb{X} \times \mathbb{Y}) \mapsto W(x, y) \in \mathbb{R}$. In this context, given the reference measure $P := (\mu \times K)$, for any $Q \in \Pi(\mu, \eta)$ we check that

$$\frac{dQ}{dP} = \frac{d(\mu \otimes \eta)}{dP} \times \frac{dQ}{d(\mu \otimes \eta)} \implies \mathcal{H}(Q | P) = Q(W) + \mathcal{H}(Q | \mu \otimes \eta) + \mathcal{H}(\eta | \nu).$$

Therefore the problem (1.1) is equivalent to the optimal entropic transport problem

$$P_{\mu, \eta} := \arg \min_{Q \in \Pi(\mu, \eta)} \mathcal{H}(Q | P) = \arg \min_{Q \in \Pi(\mu, \eta)} [Q(W) + \mathcal{H}(Q | \mu \otimes \eta)]. \quad (4.2)$$

Choosing a potential transition of the form $W(x, y) = \mathcal{W}(x, y)/t + c_t$, for some parameter $t > 0$ and some constant c_t , the problem (1.1) is now equivalent to the optimal entropic transport problem

$$P_{\mu, \eta} := \arg \min_{Q \in \Pi(\mu, \eta)} [Q(\mathcal{W}) + t \mathcal{H}(Q | \mu \otimes \eta)]. \quad (4.3)$$

In the optimal transport literature the function $\mathcal{W}$ is sometimes called the cost function and the parameter $t > 0$ is interpreted as a regularization parameter. In this context, the “unregularized” problem corresponding to $t = 0$ coincides with the Kantorovich optimal

transport problem with cost function $\mathcal{W}$. The quadratic cost model on $\mathbb{X} = \mathbb{R}^d = \mathbb{Y}$ with the Lebesgue measure $\nu(dy) = dy$ is defined by

$$W(x, y) = \frac{1}{t} \mathcal{W}(x, y) + \frac{d}{2} \log(2\pi t) \quad \text{with} \quad \mathcal{W}(x, y) = \frac{1}{2} \|x - y\|^2 \quad (4.4)$$

In this situation, $K(x, dy)$ corresponds to the heat equation transition semigroup on a given time horizon $t > 0$.

We remark that $(\mu \otimes \eta) \in \Pi(\mu, \eta)$ and

$$\mathcal{H}(\mu \otimes \eta | P) < \infty \iff (\mu \otimes \eta)(W) < \infty.$$

In contrast with Kantorovich criteria or related Wasserstein-type semi-distances, note that even when $\mathbb{X} = \mathbb{Y}$ and the same marginal $\mu = \eta$ the identity-map deterministic coupling has infinite entropy; that is we have that

$$Q(d(x, y)) := \mu(dx) \delta_x(dy) \in \Pi(\mu, \eta) \implies \mathcal{H}(Q | P) = +\infty.$$

This establishes that the Schrödinger bridge $P_{\mu, \eta}$ strongly depends on the reference probability distribution $P = (\mu \times K)$, equivalently on the cost function W for density models with reference transition K given by (4.1).

4.3 Some monotone properties

Applying (1.8) we have the Markov transport equations

$$(\pi_{2(n+1)}, \eta) = (\mu \mathcal{K}_{2(n+1)}, \pi_{2n+1} \mathcal{K}_{2(n+1)}) \quad \text{and} \quad (\mu, \pi_{2n+1}) = (\pi_{2n} \mathcal{K}_{2n+1}, \eta \mathcal{K}_{2n+1}). \quad (4.5)$$

Combining (4.5) with (3.33) we check that

$$\begin{aligned} \mathcal{H}(\pi_{2(n+1)} | \eta) &\leq \mathcal{H}(\mu | \pi_{2n+1}) \leq \mathcal{H}(\pi_{2n} | \eta) \\ \mathcal{H}(\eta | \pi_{2(n+1)}) &\leq \mathcal{H}(\pi_{2n+1} | \mu) \leq \mathcal{H}(\eta | \pi_{2n}). \end{aligned}$$

In the same vein, we have the Markov transport equations

$$(\pi_{2n+1}, \mu) = (\eta \mathcal{K}_{2n+1}, \pi_{2n} \mathcal{K}_{2n+1}) \quad \text{and} \quad (\eta, \pi_{2n}) = (\pi_{2n-1} \mathcal{K}_{2n}, \mu \mathcal{K}_{2n}). \quad (4.6)$$

Combining (4.6) with (3.33) we check that

$$\begin{aligned} \mathcal{H}(\pi_{2n+1} | \mu) &\leq \mathcal{H}(\eta | \pi_{2n}) \leq \mathcal{H}(\pi_{2n-1} | \mu) \\ \mathcal{H}(\mu | \pi_{2n+1}) &\leq \mathcal{H}(\pi_{2n} | \eta) \leq \mathcal{H}(\mu | \pi_{2n-1}). \end{aligned}$$

We summarize the above discussion with the following proposition.

Proposition 4.1. *For any $n \geq 0$ we have*

$$\begin{aligned} \mathcal{H}(\eta | \pi_{2(n+1)}) &\leq \mathcal{H}(\pi_{2n+1} | \mu) \leq \mathcal{H}(\eta | \pi_{2n}) \\ \mathcal{H}(\pi_{2(n+1)} | \eta) &\leq \mathcal{H}(\mu | \pi_{2n+1}) \leq \mathcal{H}(\pi_{2n} | \eta). \end{aligned}$$

4.4 Sinkhorn potentials

Consider the marginal measures $(\mu, \eta) = (\lambda_U, \nu_V)$ and the Markov transition K defined in (2.7) and (4.1). Note that

$$K(x, dy) = k(x, y) \nu(dy) \quad \text{with the density} \quad k(x, y) = e^{-W(x, y)}.$$

We introduce the backward-type (non necessarily Markov) integral operator

$$K^\flat(y, dx) := k^\flat(y, x) \lambda(dx) \tag{4.7}$$

with the density

$$k^\flat(y, x) = e^{-W^\flat(y, x)} = k(x, y) \iff W^\flat(y, x) = W(x, y).$$

Recalling that $\gamma := \lambda \otimes \nu$, we have

$$\mathcal{P}_0(d(x, y)) = p_0(x, y) \gamma(d(x, y))$$

with the density

$$p_0(x, y) := e^{-U_0(x)} k(x, y) e^{-V_0(y)} \quad \text{and the potential functions} \quad (U_0, V_0) = (U, 0).$$

Note that

$$\mathcal{K}_0(x, dy) = \frac{K(x, dy) e^{-V_0(y)}}{K(e^{-V_0})(x)} = K(x, dy)$$

Using (2.8) and (2.9) we check that

$$\mathcal{P}_1(d(x, y)) = p_1(x, y) \gamma(d(x, y))$$

with

$$p_1(x, y) = e^{-V(y)} \frac{p_0(x, y)}{\int p_0(\bar{x}, y) \lambda(d\bar{x})} = e^{-U_0(x)} k(x, y) \frac{e^{-V(y)}}{K^\flat(e^{-U_0})(y)}.$$

This implies that

$$p_1(x, y) := e^{-U_1(x)} k(x, y) e^{-V_1(y)}$$

with

$$U_1 = U_0 \quad \text{and} \quad V_1 = V + \log K^\flat(e^{-U_0}) \implies \mathcal{K}_1(y, dx) = \frac{K^\flat(y, dx) e^{-U_1(x)}}{K^\flat(e^{-U_1})(y)}.$$

In the same vein, we have

$$\mathcal{P}_2(d(x, y)) = p_2(x, y) \gamma(d(x, y))$$

with

$$p_2(x, y) = e^{-U(x)} \frac{p_1(x, y)}{\int p_1(x, \bar{y}) \nu(d\bar{y})} = \frac{e^{-U(x)}}{K(e^{-V_1})(x)} k(x, y) e^{-V_1(y)}.$$

This implies that

$$p_2(x, y) := e^{-U_2(x)} k(x, y) e^{-V_2(y)}$$

with

$$U_2 = U + \log K(e^{-V_1}) \quad \text{and} \quad V_2 = V_1 \implies \mathcal{K}_2(x, dy) = \frac{K(x, dy) e^{-V_2(y)}}{K(e^{-V_2})(x)}.$$

Iterating the above procedure we check the following proposition.

Proposition 4.2. *Sinkhorn bridges are given for any $n \geq 0$ by*

$$\mathcal{P}_n(d(x, y)) = p_n(x, y) \gamma(d(x, y)) \quad \text{with} \quad p_n(x, y) := e^{-U_n(x)} k(x, y) e^{-V_n(y)}. \quad (4.8)$$

The potentials (U_n, V_n) are defined for any $n \geq 0$ by the recursions

$$\begin{aligned} U_{2n+1} &= U_{2n} & V_{2n+1} &= V + \log K^\flat(e^{-U_{2n}}) \\ U_{2(n+1)} &= U + \log K(e^{-V_{2n+1}}) & V_{2(n+1)} &= V_{2n+1} \end{aligned} \quad (4.9)$$

with the initial condition $(U_0, V_0) = (U, 0)$.

By Proposition 4.2 Sinkhorn transitions take the following form

$$\begin{aligned} \mathcal{K}_{2n}(x, dy) &= K_{V_{2n}}(x, dy) := \frac{K(x, dy) e^{-V_{2n}(y)}}{K(e^{-V_{2n}})(x)} \\ \mathcal{K}_{2n+1}(y, dx) &= K_{U_{2n+1}}^\flat(y, dx) := \frac{K^\flat(y, dx) e^{-U_{2n+1}(x)}}{K^\flat(e^{-U_{2n+1}})(y)}. \end{aligned} \quad (4.10)$$

Rewriting (4.10) in a slightly different form we check the following proposition.

Proposition 4.3. *For any $n \geq 0$ we have*

$$\mathcal{K}_{2n}(x, dy) = e^{-W_{2n}(x, y)} \nu(dy) \quad \text{and} \quad \mathcal{K}_{2n+1}(y, dx) = e^{-W_{2n+1}^\flat(y, x)} \lambda(dx)$$

with the transition potential functions

$$\begin{aligned} U(x) + W_{2n}(x, y) &= U_{2n}(x) + W(x, y) + V_{2n}(y) \\ V(y) + W_{2n+1}^\flat(y, x) &= V_{2n+1}(y) + W^\flat(y, x) + U_{2n+1}(x) \end{aligned} \quad (4.11)$$

with the potential functions (U_n, V_n) as in (4.9).

Using (4.8) and recalling that $(U_0, V_0) = (U, 0)$ we check that

$$V_{2(l+1)} = V_{2l} - \log \left(\frac{d\eta}{d\pi_{2l}} \right) \quad \text{and} \quad U_{2(l+1)} = U_{2l} - \log \left(\frac{d\mu}{d\pi_{2l+1}} \right) \quad (4.12)$$

This yields the following series formulation.

Proposition 4.4. *Sinkhorn potentials (4.9) are given for any $n \geq 0$ by the formulae*

$$V_{2n} = V_0 - \sum_{0 \leq l < n} \log \left(\frac{d\eta}{d\pi_{2l}} \right) \quad \text{and} \quad U_{2n} = U_0 - \sum_{0 \leq l < n} \log \left(\frac{d\mu}{d\pi_{2l+1}} \right).$$

In addition, we have

$$\eta(V_{2(n+1)}) \leq \eta(V_{2n}) \leq \eta(V_0) = 0 \quad \text{and} \quad \mu(U_{2(n+1)}) \leq \mu(U_{2n}) \leq \mu(U).$$

Condition $\mathcal{H}(P_{\mu,\eta}|P) < \infty$ ensures the convergence of the series (1.10) and Sinkhorn entropic potentials functions $(U_n, V_n) \rightarrow_{n \rightarrow \infty} (\mathbb{U}, \mathbb{V})$ with the solution of the fixed point equation associated with the recursion (4.9) defined by the Schrödinger system

$$\begin{aligned} \mathbb{U} &= U + \log K(e^{-\mathbb{V}}) \quad \text{and} \quad \mathbb{V} = V + \log K^{\flat}(e^{-\mathbb{U}}) \\ \iff \frac{e^{-U}}{K(e^{-\mathbb{V}})} &= e^{-\mathbb{U}} \quad \text{and} \quad \frac{e^{-V}}{K^{\flat}(e^{-\mathbb{U}})} = e^{-\mathbb{V}} \end{aligned} \quad (4.13)$$

In this context, the bridge takes the form

$$P_{\mu,\eta}(d(x, y)) = \lambda_U(dx) K_{\mathbb{V}}(x, dy) = \nu_V(dy) K_{\mathbb{U}}^{\flat}(y, dx)$$

for some entropic potential functions $(\mathbb{U}, \mathbb{V})$ and the transitions

$$K_{\mathbb{V}}(x, dy) := \frac{K(x, dy) e^{-\mathbb{V}(y)}}{K(e^{-\mathbb{V}})(x)} \quad \text{and} \quad K_{\mathbb{U}}^{\flat}(y, dx) := \frac{K^{\flat}(y, dx) e^{-\mathbb{U}(x)}}{K^{\flat}(e^{-\mathbb{U}})(y)}. \quad (4.14)$$

An alternative description of Schrödinger bridges is provided in Theorem 4.2 in [97], see also Section 5.3.2 and Theorem 5.29 in the present article for a description of the entropic potential functions $(\mathbb{U}, \mathbb{V})$ in terms of absolutely convergent series. An explicit description of the entropic potential functions in the context of linear-Gaussian models is also provided in the article [3].

4.5 Entropic projections

Using (2.3) and (2.4) for any $n > p$ we have the product formulae

$$\begin{aligned} \frac{d\mathcal{P}_{2n}}{d\mathcal{P}_{2p}}(x, y) &= \left[\prod_{p \leq l < n} \frac{d\mathcal{P}_{2l+1}}{d\mathcal{P}_{2l}}(x, y) \right] \left[\prod_{p \leq l < n} \frac{d\mathcal{P}_{2(l+1)}}{d\mathcal{P}_{2l+1}}(x, y) \right] \\ &= \left[\prod_{p \leq l < n} \frac{d\eta}{d\pi_{2l}}(y) \right] \left[\prod_{p \leq l < n} \frac{d\mu}{d\pi_{2l+1}}(x) \right]. \end{aligned} \quad (4.15)$$

This yields for any $Q \in \Pi(\mu, \eta)$ and $n > p \geq 0$ the formulae

$$Q \left(\log \frac{d\mathcal{P}_{2n}}{d\mathcal{P}_{2p}} \right) = \mathcal{H}(Q|\mathcal{P}_{2p}) - \mathcal{H}(Q|\mathcal{P}_{2n}) = \sum_{p \leq l < n} [\mathcal{H}(\eta|\pi_{2l}) + \mathcal{H}(\mu|\pi_{2l+1})]. \quad (4.16)$$

Applying the above to $p = 0$ and recalling that $\mathcal{P}_0 = P$, this yields the formula

$$\mathcal{H}(Q|P) = \mathcal{H}(Q|\mathcal{P}_{2n}) + \sum_{0 \leq l < n} [\mathcal{H}(\eta|\pi_{2l}) + \mathcal{H}(\mu|\pi_{2l+1})]. \quad (4.17)$$

Observe that

$$\begin{aligned} \mathcal{H}(Q|\mathcal{P}_{2n}) &= \mathcal{H}(Q|\mathcal{P}_{2n+1}) + Q \left(\log \frac{d\mathcal{P}_{2n+1}}{d\mathcal{P}_{2n}} \right) = \mathcal{H}(Q|\mathcal{P}_{2n+1}) + \mathcal{H}(\eta|\pi_{2n}) \\ \mathcal{H}(Q|\mathcal{P}_{2n+1}) &= \mathcal{H}(Q|\mathcal{P}_{2(n+1)}) + Q \left(\log \frac{d\mathcal{P}_{2(n+1)}}{d\mathcal{P}_{2n+1}} \right) = \mathcal{H}(Q|\mathcal{P}_{2(n+1)}) + \mathcal{H}(\mu|\pi_{2n+1}). \end{aligned} \quad (4.18)$$

This implies that

$$\mathcal{H}(Q|P) = \mathcal{H}(Q|\mathcal{P}_{2n+1}) + \mathcal{H}(\eta|\pi_{2n}) + \sum_{0 \leq l < n} [\mathcal{H}(\eta|\pi_{2l}) + \mathcal{H}(\mu|\pi_{2l+1})]. \quad (4.19)$$

Combining (4.17) with (4.19) and (1.1) we find that the Schrödinger bridge $P_{\mu,\eta}$ with reference measure P defined in (1.1) can be computed from any Sinkhorn reference bridge $\mathcal{P}_n$. Equivalently, the bridge map $P \mapsto P_{\mu,\eta}$ is constant along the flow of Sinkhorn bridges $\mathcal{P}_n$. More formally we have the following proposition.

Proposition 4.5. *For any $n \geq 0$ we have*

$$(\mathcal{P}_n)_{\mu,\eta} = \arg \min_{Q \in \Pi(\mu,\eta)} \mathcal{H}(Q|\mathcal{P}_n) = P_{\mu,\eta}. \quad (4.20)$$

Applying (4.16) to $Q = P_{\mu,\eta}$ we have

$$\mathcal{H}(P_{\mu,\eta}|\mathcal{P}_{2p}) = \mathcal{H}(P_{\mu,\eta}|\mathcal{P}_{2n}) + \sum_{p \leq l < n} [\mathcal{H}(\eta|\pi_{2l}) + \mathcal{H}(\mu|\pi_{2l+1})]. \quad (4.21)$$

We further assume the series on the r.h.s. of (4.21) converges. By Proposition 4.1, this condition is met if and only if the sequence of entropies $\mathcal{H}(\eta|\pi_{2l})$ or $\mathcal{H}(\mu|\pi_{2l+1})$ is summable.

Whenever

$$\lim_{n \rightarrow \infty} \mathcal{H}(P_{\mu,\eta}|\mathcal{P}_{2n}) = 0, \quad (4.22)$$

letting $n \rightarrow \infty$ in (4.21), for any $p \geq 0$ we have

$$\mathcal{H}(P_{\mu,\eta}|\mathcal{P}_{2p}) = \sum_{l \geq p} [\mathcal{H}(\eta|\pi_{2l}) + \mathcal{H}(\mu|\pi_{2l+1})]. \quad (4.23)$$

Conversely when (4.23) is satisfied (4.22) clearly holds. We summarize the above discussion with the following proposition.

Proposition 4.6. *We have the equivalence (1.10) as soon as the sequence of entropies $\mathcal{H}(\eta|\pi_{2l})$ or $\mathcal{H}(\mu|\pi_{2l+1})$ is summable.*

We end this section with some elementary monotone properties. Applying (4.18) to the bridge $Q = P_{\mu,\eta}$ and using Proposition 4.1 we check that

$$\begin{aligned} \mathcal{H}(P_{\mu,\eta}|\mathcal{P}_{2(n+1)}) &= \mathcal{H}(P_{\mu,\eta}|\mathcal{P}_{2n+1}) - \mathcal{H}(\mu|\pi_{2n+1}) \\ &= \mathcal{H}(P_{\mu,\eta}|\mathcal{P}_{2n}) - [\mathcal{H}(\eta|\pi_{2n}) + \mathcal{H}(\mu|\pi_{2n+1})] \\ &\leq \mathcal{H}(P_{\mu,\eta}|\mathcal{P}_{2n}) - [\mathcal{H}(\pi_{2n+1}|\mu) + \mathcal{H}(\pi_{2(n+1)}|\eta)]. \end{aligned} \quad (4.24)$$

Similarly, we have

$$\begin{aligned} \mathcal{H}(P_{\mu,\eta}|\mathcal{P}_{2n+1}) &= \mathcal{H}(P_{\mu,\eta} | \mathcal{P}_{2n}) - \mathcal{H}(\eta | \pi_{2n}) \\ &= \mathcal{H}(P_{\mu,\eta}|\mathcal{P}_{2n-1}) - [\mathcal{H}(\mu|\pi_{2n-1}) + \mathcal{H}(\eta|\pi_{2n})] \\ &\leq \mathcal{H}(P_{\mu,\eta}|\mathcal{P}_{2n-1}) - [\mathcal{H}(\pi_{2n}|\eta) + \mathcal{H}(\pi_{2n+1}|\mu)]. \end{aligned} \quad (4.25)$$

4.6 Commutation formulae

Observe that

$$\begin{aligned} \frac{d\mathcal{P}_{2(l+1)}}{d\mathcal{P}_{2l}}(x, y) &= \frac{d\mathcal{P}_{2(l+1)}}{d\mathcal{P}_{2l+1}}(x, y) \frac{d\mathcal{P}_{2l+1}}{d\mathcal{P}_{2l}}(x, y) = \frac{d\mu}{d\pi_{2l+1}}(x) \frac{d\eta}{d\pi_{2l}}(y) \\ \iff \mathcal{K}_{2(l+1)}(x, dy) &= \frac{d\mu}{d\pi_{2l+1}}(x) \mathcal{K}_{2l}(x, dy) \frac{d\eta}{d\pi_{2l}}(y). \end{aligned}$$

Similarly, we have

$$\begin{aligned} \frac{d\mathcal{P}_{2l+1}}{d\mathcal{P}_{2l-1}}(x, y) &= \frac{d\mathcal{P}_{2l+1}}{d\mathcal{P}_{2l}}(x, y) \frac{d\mathcal{P}_{2l}}{d\mathcal{P}_{2l-1}}(x, y) = \frac{d\eta}{d\pi_{2l}}(y) \frac{d\mu}{d\pi_{2l-1}}(x) \\ \iff \mathcal{K}_{2l+1}(y, dx) &= \frac{d\eta}{d\pi_{2l}}(y) \mathcal{K}_{2l-1}(y, dx) \frac{d\mu}{d\pi_{2l-1}}(x). \end{aligned}$$

These computations allow us to deduce the following result.

Proposition 4.7. *For any $l \geq 0$ we have the commutation formulae*

$$\mathcal{K}_{2l} \left(\frac{d\eta}{d\pi_{2l}} \right) = \frac{d\pi_{2l+1}}{d\mu} \quad \text{and} \quad \mathcal{K}_{2(l+1)} \left(\frac{d\pi_{2l}}{d\eta} \right) = \frac{d\mu}{d\pi_{2l+1}}.$$

In addition, for any $l \geq 1$ we have

$$\mathcal{K}_{2l-1} \left(\frac{d\mu}{d\pi_{2l-1}} \right) = \frac{d\pi_{2l}}{d\eta} \quad \text{and} \quad \mathcal{K}_{2l+1} \left(\frac{d\pi_{2l-1}}{d\mu} \right) = \frac{d\eta}{d\pi_{2l}}.$$

In terms of the forward-backward Gibbs-type transitions (1.9) we have

$$\frac{d\pi_{2l}}{d\eta} = \mathcal{K}_{2l-1} \left(\frac{d\mu}{d\pi_{2l-1}} \right) = \mathcal{K}_{2l-1} \mathcal{K}_{2l} \left(\frac{d\pi_{2(l-1)}}{d\eta} \right) = \mathcal{S}_{2l} \left(\frac{d\pi_{2(l-1)}}{d\eta} \right).$$

In addition, we have

$$\frac{d\pi_{2l+1}}{d\mu} = \mathcal{K}_{2l} \left(\frac{d\eta}{d\pi_{2l}} \right) = \mathcal{K}_{2l} \left(\mathcal{K}_{2l+1} \left(\frac{d\pi_{2l-1}}{d\mu} \right) \right) = \mathcal{S}_{2l+1} \left(\frac{d\pi_{2l-1}}{d\mu} \right).$$

This yields the following result.

Proposition 4.8. *For any $l \geq 1$ we have*

$$\mathcal{S}_{2l} \left(\frac{d\pi_{2(l-1)}}{d\eta} \right) = \frac{d\pi_{2l}}{d\eta} \quad \text{and} \quad \mathcal{S}_{2l+1} \left(\frac{d\pi_{2l-1}}{d\mu} \right) = \frac{d\pi_{2l+1}}{d\mu}.$$

4.7 Gradient and Hessian formulae

Consider the marginal measures (μ, η) and the transition K discussed in Section 4.4 for some potential functions (U, V) and W on the state spaces $\mathbb{X} = \mathbb{R}^d = \mathbb{Y}$ equipped with the Lebesgue measure $\lambda(dz) = dz = \nu(dz)$. In addition, assume (U, V) in (2.7) are twice differentiable and W is the linear-Gaussian transition potential defined in (3.32), for some

parameters $(\alpha, \beta, \tau) \in (\mathbb{R}^d \times \mathcal{G}l_d \times \mathcal{S}_d^+)$. In this scenario the gradient and the Hessian defined in (1.14) reduce to

$$\begin{aligned} W_x(y) &= \chi'((\alpha + \beta x) - y) \quad \text{and} \quad W_y^b(x) = \tau^{-1}(y - (\alpha + \beta x)) \\ \implies W_{x,x}(y) &= \chi' \beta \quad \text{and} \quad W_{y,y}^b(x) = \tau^{-1} \quad \text{with} \quad \chi := \tau^{-1} \beta. \end{aligned} \tag{4.26}$$

Therefore, the conditional covariances (1.16) and (1.17) associated with Sinkhorn transitions (4.10) take the form

$$\text{cov}_{\mathcal{K}_{2n}, W}(x) = \chi' \text{cov}_{\mathcal{K}_{2n}}(x) \chi \quad \text{and} \quad \text{cov}_{\mathcal{K}_{2n+1}, W^b}(y) = \chi \text{cov}_{\mathcal{K}_{2n+1}}(y) \chi'.$$

Combining (1.15) with (4.9) and (4.10) one has the following result.

Proposition 4.9. *For any $n \geq 0$ we have the gradient formulae*

$$\begin{aligned} \nabla U_{2n}(x) &= \nabla U(x) - \chi' \{(\alpha + \beta x) - \text{mean}_{\mathcal{K}_{2n}}(x)\} \\ \nabla V_{2n+1}(y) &= \nabla V(y) - \tau^{-1} \{y - (\alpha + \beta \text{mean}_{\mathcal{K}_{2n+1}}(y))\} \end{aligned} \tag{4.27}$$

with the conditional expectations $\text{mean}_{\mathcal{K}_n}$ defined in (1.17). In addition, we have

$$\begin{aligned} \nabla^2 U_{2n}(x) &= \nabla^2 U(x) - \chi' \beta + \chi' \text{cov}_{\mathcal{K}_{2n}}(x) \chi \\ \nabla^2 V_{2n+1}(y) &= \nabla^2 V(y) - \tau^{-1} + \chi \text{cov}_{\mathcal{K}_{2n+1}}(y) \chi'. \end{aligned} \tag{4.28}$$

We remark that (1.15) and (4.13) yield

$$\begin{aligned} \nabla \mathbb{U}(x) &= \nabla U(x) - \chi' \{(\alpha + \beta x) - \text{mean}_{K_{\mathbb{V}}}(x)\} \\ \nabla^2 \mathbb{U}(x) &= \nabla^2 U(x) - \chi' \beta + \chi' \text{cov}_{K_{\mathbb{V}}}(x) \chi \end{aligned}$$

with the bridge transition $K_{\mathbb{V}}$ defined in (4.14). Recalling that $\mathcal{K}_{2n} = K_{V_{2n}}$ we have

$$\nabla(\mathbb{U} - U_{2n}) = \chi'(\text{mean}_{K_{\mathbb{V}}} - \text{mean}_{K_{V_{2n}}}) \quad \text{and} \quad \nabla^2(\mathbb{U} - U_{2n}) = \chi'(\text{cov}_{K_{\mathbb{V}}} - \text{cov}_{K_{V_{2n}}}) \chi.$$

The convergence of conditional expectations and covariances of Sinkhorn transitions towards the conditional expectations and covariances of Schrödinger bridges is discussed in [50], see also Theorem 5.13 in the present article. We conclude this discussion by noting that applying Proposition 4.9 along with (4.11) we obtain the following interplay between the Hessian of Sinkhorn log-densities and conditional covariance functions.

Proposition 4.10. *For any $n \geq 0$ we have the decompositions*

$$\begin{aligned} \nabla_2^2 W_{2n+1}^b(y, x) &= \nabla^2 U(x) + \chi' \text{cov}_{\mathcal{K}_{2n}}(x) \chi \\ \nabla_2^2 W_{2(n+1)}(x, y) &= \nabla^2 V(y) + \chi \text{cov}_{\mathcal{K}_{2n+1}}(y) \chi'. \end{aligned} \tag{4.29}$$

5 Stability analysis

5.1 Overview

In Section 4 we have presented different formulations of Sinkhorn bridges, including geometric entropic projection methodologies, operator-theoretic framework with Markov integral transport formulae and commutation-type formulae, as well as dual formulations in terms of potential functions. The stability of Sinkhorn bridges discussed in this section combines these different formulations with the stability theory of Markov operators provided in Section 3.

5.2 Linear decay rates

Using (4.17) and the monotone properties stated in Proposition 4.1 for any $Q \in \Pi(\mu, \eta)$ and $n \geq 0$ we have the estimate

$$\mathcal{H}(Q|P) \geq \sum_{0 \leq l \leq n} (\mathcal{H}(\eta|\pi_{2l}) + \mathcal{H}(\mu|\pi_{2l+1})) \geq (n+1) (\mathcal{H}(\eta|\pi_{2n}) + \mathcal{H}(\mu|\pi_{2n+1})).$$

This yields an elementary and very short proof of the following rather well known theorem.

Theorem 5.1. *For any $n \geq 1$ we have*

$$\mathcal{H}(\eta|\pi_{2n}) + \mathcal{H}(\mu|\pi_{2n+1}) \leq \frac{1}{n+1} \mathcal{H}(P_{\mu,\eta}|P)$$

and, in addition,

$$\lim_{n \rightarrow \infty} n \mathcal{H}(\eta | \pi_{2n}) = \lim_{n \rightarrow \infty} n \mathcal{H}(\mu | \pi_{2n+1}) = 0.$$

The last assertion is a direct consequence of the convergence of the series (4.19). Sub-linear rates have been developed in the articles [5, 22, 61]. Linear rates with the robust constant $\text{Ent}(P_{\mu,\eta} | P)$ on possibly non-compact spaces were first obtained by Léger in [86] using elegant gradient descent and Bregman divergence techniques, see also the recent articles [42, 84]. A refined convergence rate at least one order faster has also been developed in [73].

5.3 Bounded cost functions

5.3.1 Universal contraction rates

The contraction analysis developed in this section is taken from [4]. Consider the marginal measures $(\mu, \eta) = (\lambda_U, \nu_V)$ and the Markov transition K defined in (2.7) and (4.1). We further assume that $\text{osc}(W) < \infty$. In this case, using (4.10) we check that

$$\begin{aligned} \frac{d\delta_{x_1} \mathcal{K}_{2n}}{d\delta_{x_2} \mathcal{K}_{2n}}(y) \frac{d\delta_{x_2} \mathcal{K}_{2n}}{d\delta_{x_1} \mathcal{K}_{2n}}(z) &= \exp((W(x_2, y) - W(x_1, y)) + (W(x_1, z) - W(x_2, z))) \quad (5.1) \\ &\geq \epsilon_W := \exp(-2 \text{osc}(W)). \end{aligned}$$

The time-homogeneous property (5.1) is sometimes called “structure conservation”. A first proof of geometric convergence based on these constant cross-product ratios is due to Fienberg in the seminal article [66].

The case when one has bounded transition potentials, W , typically arise on finite or compact spaces. For instance, Markov processes on compact manifolds, such as diffusions on bounded manifolds reflected at the boundary, have Markov transitions of the form (4.1) with a uniformly bounded transition potential W that depends on the time parameter [9, 10, 95, 117]. The convergence analysis of Sinkhorn bridges for uniformly bounded transition potential W is often developed using Hilbert projective metrics [26, 43, 69, 57]. As remarked in [32, Section 3], the Hilbert metric dominates all ϕ -divergences, thus we expect rather strong conditions to ensure the contraction of Sinkhorn transitions w.r.t. the Hilbert metric, see for instance [17, 26, 43, 69]. Thus contraction estimates w.r.t. the Hilbert metric yield exponential decays w.r.t. ϕ -divergences, including the total variation norm, the relative

entropy and Wasserstein distances of any order. Nevertheless for the same reason (i.e. the domination of all ϕ -divergences) they can rarely be used to derive contraction estimates with respect to some ϕ -divergence criteria (unless the criteria is equivalent to the Hilbert metric).

To extend our discussion, we observe that for any $n \geq 0$ we have

$$\begin{aligned} \delta_{x_1} \mathcal{K}_{2n}(dy) \delta_{x_2} \mathcal{K}_{2n}(dz) &\geq \epsilon_W \delta_{x_2} \mathcal{K}_{2n}(dy) \delta_{x_1} \mathcal{K}_{2n}(dz) \\ \implies \delta_{x_1} \mathcal{K}_{2n} &\geq \epsilon_W \delta_{x_2} \mathcal{K}_{2n} \implies \text{dob}(\mathcal{K}_{2n}) \leq 1 - \epsilon_W. \end{aligned} \quad (5.2)$$

In a similar manner, for any $n \geq 0$ we check that

$$\delta_{y_1} \mathcal{K}_{2n+1} \geq \epsilon_W \delta_{y_2} \mathcal{K}_{2n+1} \implies \text{dob}(\mathcal{K}_{2n+1}) \leq 1 - \epsilon_W. \quad (5.3)$$

Without further work, the Markov transport equations presented in (4.5) and (4.6) yield the following theorem.

Theorem 5.2. *For any convex function Φ satisfying (3.1) we have the contraction estimates*

$$\begin{aligned} \mathcal{H}_\Phi(\pi_{2(n+1)}, \eta) &\leq (1 - \epsilon_W) \mathcal{H}_\Phi(\mu, \pi_{2n+1}) \leq (1 - \epsilon_W)^2 \mathcal{H}_\Phi(\pi_{2n}, \eta), \\ \mathcal{H}_\Phi(\eta, \pi_{2(n+1)}) &\leq (1 - \epsilon_W) \mathcal{H}_\Phi(\pi_{2n+1}, \mu) \leq (1 - \epsilon_W)^2 \mathcal{H}_\Phi(\eta, \pi_{2n}). \end{aligned} \quad (5.4)$$

The above inequalities remain valid when we replace (η, π_{2n}) by (μ, π_{2n+1}) .

Proposition 5.3. *There exists some $c > 0$ such that for any $n \geq 1$ we have*

$$\left\| \frac{d\pi_{2n}}{d\eta} - 1 \right\| \vee \left\| \frac{d\eta}{d\pi_{2n}} - 1 \right\| \vee \left\| \log \frac{d\eta}{d\pi_{2n}} \right\| \leq c (1 - \epsilon_W)^{2n}.$$

The above estimates remain valid when we replace (π_{2n}, η) by (π_{2n+1}, μ) .

Proof. By (5.3) for any $n \geq 1$ and $x \in \mathbb{X}$ and $y_1, y_2 \in \mathbb{Y}$ we have

$$\epsilon_W \delta_{y_1} \mathcal{S}_{2n} \leq \delta_{y_2} \mathcal{S}_{2n} = \int \delta_{y_2} \mathcal{K}_{2n-1}(dz) \delta_z \mathcal{K}_{2n} \leq \epsilon_W^{-1} \delta_{y_1} \mathcal{S}_{2n}.$$

Using (1.8) and (1.9), integrating y_1 w.r.t. η and y_2 w.r.t. $\pi_{2(n-1)}$ this implies that

$$\begin{aligned} \epsilon_W \eta &\leq \pi_{2n} \leq \epsilon_W^{-1} \eta \\ \implies \frac{d\eta}{d\pi_{2n}} \vee \frac{d\pi_{2n}}{d\eta} &\leq \epsilon_W^{-1} \quad \text{and similarly} \quad \frac{d\mu}{d\pi_{2n+1}} \vee \frac{d\pi_{2n+1}}{d\mu} \leq \epsilon_W^{-1}. \end{aligned}$$

By (1.8) and Proposition 4.8 for any $k, l \geq 0$ we have

$$\frac{d\pi_{2(l+k)}}{d\eta}(y) - 1 = (\delta_y - \eta) \mathcal{S}_{2(l+k)} \cdots \mathcal{S}_{2(l+1)} \left(\frac{d\pi_{2l}}{d\eta} \right).$$

On the other hand, (5.3) and (5.2) ensure that for any $n \geq 1$ we have $\text{dob}(\mathcal{S}_n) \leq (1 - \epsilon_W)^2$.

Using (3.4) and (3.9) for any $l \geq 1$ and $k \geq 0$ we check that

$$\left| \frac{d\pi_{2(l+k)}}{d\eta}(y) - 1 \right| \leq 2\epsilon_W^{-1} \|(\delta_y - \eta) \mathcal{S}_{2(l+k)} \cdots \mathcal{S}_{2(l+1)}\|_{\text{tv}} \leq 2\epsilon_W^{-1} (1 - \epsilon_W)^{2k}.$$

In the same vein, we have

$$\begin{aligned} \frac{d\pi_{2(l+k)+1}}{d\mu}(x) - 1 &= (\delta_x - \mu) \mathcal{S}_{2(l+k)+1} \cdots \mathcal{S}_{2(l+1)+1} \left(\frac{d\pi_{2l+1}}{d\mu} \right) \\ \implies \left| \frac{d\pi_{2(l+k)+1}}{d\mu}(x) - 1 \right| &\leq 2\epsilon_W^{-1} (1 - \epsilon_W)^{2k}. \end{aligned}$$

By Proposition 4.7 we have

$$\frac{d\eta}{d\pi_{2l}} - 1 = \mathcal{K}_{2l+1} \left(\frac{d\pi_{2l-1}}{d\mu} - 1 \right) \quad \text{and} \quad \frac{d\mu}{d\pi_{2l+1}} - 1 = \mathcal{K}_{2(l+1)} \left(\frac{d\pi_{2l}}{d\eta} - 1 \right).$$

Finally, note that for any function $f(z) > 0$ we have

$$\begin{aligned} |\log f(z)| &= \log(1 + (f(z) - 1)) \mathbf{1}_{f(z) \geq 1} + \log\left(1 + \frac{1}{f(z)} - 1\right) \mathbf{1}_{1/f(z) \geq 1} \\ &\leq (f(z) - 1) \mathbf{1}_{f(z) \geq 1} + \left(\frac{1}{f(z)} - 1\right) \mathbf{1}_{1/f(z) \geq 1}. \end{aligned} \tag{5.5}$$

The last assertion follows from the fact that $\log(1 + u) \leq u$, for any $1 + u > 0$. $\blacksquare$

5.3.2 Schrödinger bridges

By Proposition 4.4 and Proposition 5.3 there exists some constant $c > 0$ such that for any $n \geq 0$ we have

$$\|V_{2n} - \mathbb{V}\| \vee \|U_{2n} - \mathbb{U}\| \leq c (1 - \epsilon_W)^{2n}$$

with the uniformly absolutely-convergent series

$$\mathbb{V} = V_0 - \sum_{l \geq 0} \log \left(\frac{d\eta}{d\pi_{2l}} \right) \quad \text{and} \quad \mathbb{U} = U_0 - \sum_{l \geq 0} \log \left(\frac{d\mu}{d\pi_{2l+1}} \right). \tag{5.6}$$

In addition, we have

$$\mathbb{P}(d(x, y)) := e^{-\mathbb{U}(x)} e^{-W(x, y)} e^{-\mathbb{V}(y)} \gamma(d(x, y)) \implies \|\mathcal{P}_{2n} - \mathbb{P}\|_{\text{tv}} \leq c (1 - \epsilon_W)^{2n}.$$

We check this claim using the fact that for any probability measures of the form

$$p_u(dz) := e^{-u(z)} \gamma(dz) \quad \text{and} \quad p_v(dz) := e^{-v(z)} \gamma(dz)$$

for some real valued functions $u(z), v(z)$ we have

$$\begin{aligned} \|p_u - p_v\|_{\text{tv}} &= \frac{1}{2} \int |e^{-u(z)} - e^{-v(z)}| \gamma(dz) \\ &\leq \frac{1}{2} \|u - v\| \int (e^{-u(z)} + e^{-v(z)}) \gamma(dz) = \|u - v\|. \end{aligned}$$

In addition, using (4.9) one has

$$\mathbb{V} = V + \log K^b(e^{-\mathbb{U}}) \quad \text{and} \quad \mathbb{U} = U + \log K(e^{-\mathbb{V}}) \implies \mathbb{P} = P_{\mu, \eta}.$$

Also note that

$$\mathcal{H}(P_{\mu,\eta} \mid \mathcal{P}_{2n}) = \mu(U_{2n} - \mathbb{U}) + \eta(V_{2n} - \mathbb{V}).$$

We summarize the above discussion with the following theorem.

Theorem 5.4. *There exists some $c > 0$ such that for any $n \geq 0$ we have*

$$\mathcal{H}(P_{\mu,\eta} \mid \mathcal{P}_{2n}) \vee \|P_{\mu,\eta} - \mathcal{P}_{2n}\|_{\text{tv}} \leq c (1 - \epsilon_W)^{2n}.$$

In addition, we have the entropic series expansion

$$\mathcal{H}(P_{\mu,\eta} \mid \mathcal{P}_{2n}) = \sum_{l \geq n} [\mathcal{H}(\mu \mid \pi_{2l+1}) + \mathcal{H}(\eta \mid \pi_{2l})].$$

5.4 Gaussian bridges

5.4.1 Linear Gaussian models

Consider the state spaces $\mathbb{X} = \mathbb{Y} = \mathbb{R}^d$, for some integer $d \geq 1$ and the Lebesgue measure $\lambda(dz) = \nu(dz) = dz$. Hereafter, points x in the Euclidean space $\mathbb{X} = \mathbb{R}^d$ are represented by d -dimensional column vectors (or, equivalently, by $d \times 1$ matrices). When there is no possible confusion, we use the notation $\|\cdot\|$ for any equivalent matrix or vector norm.

Consider the target Gaussian probability measures $(\mu, \eta) = (\lambda_U, \lambda_V) = (\nu_{m,\sigma}, \nu_{\bar{m},\bar{\sigma}})$ discussed in (3.30). Equivalently, the log-density functions (U, V) are defined by

$$U(x) = -\log \text{gauss}_\sigma(x - m) \quad \text{and} \quad V(y) := -\log \text{gauss}_{\bar{\sigma}}(y - \bar{m}). \quad (5.7)$$

Denote by G a d -dimensional centred Gaussian random variable (r.v.) with unit covariance. Then we have that the linear Gaussian transition K is defined by (3.32) takes the form

$$K(x, dy) := e^{-W(x,y)} dy = \mathcal{K}_0(x, dy) := \mathbb{P}(\mathcal{Z}_0(x) \in dy) \quad (5.8)$$

with the random map

$$\begin{aligned} \mathcal{Z}_0(x) &:= \alpha + \beta x + \tau^{1/2} G \\ \implies \pi_0 &:= \mu K = \nu_{m_0, \sigma_0} \quad \text{with} \quad m_0 := \alpha + \beta m \quad \text{and} \quad \sigma_0 := \beta \sigma \beta' + \tau. \end{aligned} \quad (5.9)$$

5.4.2 Sinkhorn bridges

We can rewrite $\mathcal{Z}_0(x)$ into the following alternative form

$$\mathcal{Z}_0(x) := m_0 + \beta_0 (x - m) + \tau_0^{1/2} G \quad \text{with} \quad (\beta_0, \tau_0) := (\beta, \tau).$$

The conjugate formula (2.1) is defined via

$$K_\mu^*(y, dx) = \mathcal{K}_1(y, dx) := \mathbb{P}(\mathcal{Z}_1(y) \in dx) \quad (5.10)$$

with the random map $\mathcal{Z}_1(y) := m + \beta_1 (y - m_0) + \tau_1^{1/2} G$

defined in terms of the gain matrix

$$\beta_1 := \sigma \beta' \sigma_0^{-1} \quad \text{and} \quad \tau_1^{-1} := \sigma^{-1} + \beta' \tau^{-1} \beta,$$

In signal processing and filtering literature, this explicit formula is often referred as the Bayesian update formula for the Kalman filter.

Applying the matrix inversion lemma to the covariance matrix σ_0 defined in (5.9) we have

$$\begin{aligned} \sigma_0^{-1} &= \tau^{-1} - \tau^{-1} \beta (\sigma^{-1} + \beta' \tau^{-1} \beta)^{-1} \beta' \tau^{-1} = \tau^{-1} - \tau^{-1} \beta \tau_1 \beta' \tau^{-1} \\ \implies \beta_1 &= \sigma \beta' \tau^{-1} - \sigma (\beta' \tau^{-1} \beta) \tau_1 \beta' \tau^{-1} = \sigma \beta' \tau^{-1} - \sigma (\tau_1^{-1} - \sigma^{-1}) \tau_1 \beta' \tau^{-1} \\ \implies \beta_1 &= \tau_1 \beta' \tau^{-1}. \end{aligned}$$

Then one can obtain $\mathcal{Z}_1(y)$ as

$$\mathcal{Z}_1(y) = m_1 + \beta_1 (y - \bar{m}) + \tau_1^{1/2} G \quad \text{with} \quad m_1 = m + \beta_1 (\bar{m} - m_0)$$

This implies that $\pi_1 = \nu_{\bar{m}, \bar{\sigma}} \mathcal{K}_1 = \nu_{m_1, \sigma_1}$ with the covariance matrix

$$\sigma_1 := \beta_1 \bar{\sigma} \beta_1' + \tau_1. \quad (5.11)$$

The term (1.5) can be expressed using the regression formula

$$(\mathcal{K}_1)_\eta^*(x, dy) = \mathcal{K}_2(x, dy) := \mathbb{P}(\mathcal{Z}_2(x) \in dy) \quad (5.12)$$

$$\mathcal{Z}_2(x) := \bar{m} + \beta_2 (x - m_1) + \tau_2^{1/2} G \quad (5.13)$$

where

$$\beta_2 := \bar{\sigma} \beta_1' \sigma_1^{-1} \quad \text{and} \quad \tau_2^{-1} := \bar{\sigma}^{-1} + \beta_1' \tau_1^{-1} \beta_1.$$

Here $\mathcal{Z}_2(x)$ is given by

$$\mathcal{Z}_2(x) = m_2 + \beta_2 (x - m) + \tau_2^{1/2} G \quad \text{with} \quad m_2 = \bar{m} + \beta_2 (m - m_1).$$

Applying the matrix inversion lemma to the covariance matrix σ_1 defined in (5.11) we have

$$\begin{aligned} \sigma_1^{-1} &= \tau_1^{-1} - \tau_1^{-1} \beta_1 (\bar{\sigma}^{-1} + \beta_1' \tau_1^{-1} \beta_1)^{-1} \beta_1' \tau_1^{-1} = \tau_1^{-1} - \tau_1^{-1} \beta_1 \tau_2 \beta_1' \tau_1^{-1} \\ \implies \beta_2 &= \bar{\sigma} \beta_1' \tau_1^{-1} - \bar{\sigma} (\beta_1' \tau_1^{-1} \beta_1) \tau_2 \beta_1' \tau_1^{-1} = \bar{\sigma} \beta_1' \tau_1^{-1} - \bar{\sigma} (\tau_2^{-1} - \bar{\sigma}^{-1}) \tau_2 \beta_1' \tau_1^{-1} \\ \implies \beta_2 &= \tau_2 \beta_1' \tau_1^{-1} = \tau_2 \tau^{-1} \beta. \end{aligned}$$

Iterating the Bayes update formulae (5.10), (5.12), (5.13) (see section 4 and [3, Appendix C]) the Sinkhorn transitions can be expressed as

$$\mathcal{K}_{2n}(x, dy) = \mathbb{P}(\mathcal{Z}_{2n}(x) \in dy) \quad \text{and} \quad \mathcal{K}_{2n+1}(y, dx) = \mathbb{P}(\mathcal{Z}_{2n+1}(y) \in dx)$$

with linear-Gaussian random maps defined, for any $n \geq 0$, by

$$\begin{aligned} \mathcal{Z}_{2n}(x) &= m_{2n} + \beta_{2n}(x - m) + \tau_{2n}^{1/2} G \\ \mathcal{Z}_{2n+1}(y) &= m_{2n+1} + \beta_{2n+1}(y - \bar{m}) + \tau_{2n+1}^{1/2} G. \end{aligned} \quad (5.14)$$

The parameters in (5.14) are described sequentially by the following theorem.

Theorem 5.5 ([3]). *The gain β_n and covariance matrices τ_n in the Bayes update formulae (5.14) are defined for any $n \geq 0$ by the recursion*

$$\begin{aligned}\beta_{2n} &= \tau_{2n} \tau^{-1} \beta & \beta_{2n+1} &= \tau_{2n+1} \beta' \tau^{-1}. \\ \tau_{2n+1}^{-1} &= \sigma^{-1} + \beta'_{2n} \tau_{2n}^{-1} \beta_{2n} & \tau_{2(n+1)}^{-1} &= \bar{\sigma}^{-1} + \beta'_{2n+1} \tau_{2n+1}^{-1} \beta_{2n+1}\end{aligned}\tag{5.15}$$

The parameters m_n in (5.14) are also defined, for any $n \geq 0$, by the recursion

$$\begin{aligned}m_{2n+1} &= m + \beta_{2n+1}(\bar{m} - m_{2n}), \\ m_{2(n+1)} &= \bar{m} + \beta_{2(n+1)}(m - m_{2n+1})\end{aligned}\tag{5.16}$$

with initial condition $m_0 = \alpha_0 + \beta_0 m$, where, again, $(\alpha_0, \beta_0, \tau_0) = (\alpha, \beta, \tau)$ are parameters of the reference linear Gaussian transition (5.8).

Note that $\pi_n = \nu_{m_n, \sigma_n}$ with the covariance matrices

$$\sigma_{2n} := \beta_{2n} \sigma \beta'_{2n} + \tau_{2n} \quad \text{and} \quad \sigma_{2n+1} := \beta_{2n+1} \bar{\sigma} \beta'_{2n+1} + \tau_{2n+1}.$$

Example 5.6. *In the context of stochastic flows discussed in (2.10) with a linear drift $b_s(x) = Ax$ and an homogeneous diffusion function $\sigma_s(x) = \Sigma$, for some $A \in \mathbb{R}^{d \times d}$ and $\Sigma \in \mathcal{S}_d^+$, for any given time horizon $t > 0$ one has*

$$T_t(x, dy) = \mathbb{P}(X_t(x) \in dy) = e^{-\mathcal{W}_t(x, y)} dy \quad \text{with} \quad \mathcal{W}_t(x, y) := -\log \text{gauss}_{\tau(t)}(y - \beta(t)x)\tag{5.17}$$

and the parameters

$$(\beta(t), \tau(t)) = \left(e^{tA}, \int_0^t e^{sA} \Sigma e^{sA'} ds \right).$$

In this context we can choose the reference transition (5.8) with $W = \mathcal{W}_t$.

5.4.3 Riccati equations

Consider the rescaled covariance matrices

$$v_{2n} := \bar{\sigma}^{-1/2} \tau_{2n} \bar{\sigma}^{-1/2} \quad \text{and} \quad v_{2n+1} := \sigma^{-1/2} \tau_{2n+1} \sigma^{-1/2}\tag{5.18}$$

Using (5.15) we check that

$$\begin{aligned}\tau_{2n+1}^{-1} &= \sigma^{-1} + (\tau_{2n} \tau^{-1} \beta)' \tau_{2n}^{-1} (\tau_{2n} \tau^{-1} \beta) = \sigma^{-1} + (\tau^{-1} \beta)' \tau_{2n}^{-1} (\tau^{-1} \beta) \\ \tau_{2(n+1)}^{-1} &= \bar{\sigma}^{-1} + (\tau_{2n+1} \beta' \tau^{-1})' \tau_{2n+1}^{-1} (\tau_{2n+1} \beta' \tau^{-1}) = \bar{\sigma}^{-1} + (\tau^{-1} \beta)' \tau_{2n+1}^{-1} (\tau^{-1} \beta)'\end{aligned}$$

This yields the recursions

$$v_{2n+1}^{-1} = I + \gamma' v_{2n} \gamma \quad \text{and} \quad v_{2(n+1)}^{-1} = I + \gamma v_{2n+1} \gamma'\tag{5.19}$$

with

$$\gamma := \bar{\sigma}^{1/2} \tau^{-1} \beta \sigma^{1/2}.$$

Next, observe that

$$u^{-1} = I + \gamma' v \gamma \implies (I + \gamma u \gamma')^{-1} = \left(I + (\varpi + v)^{-1} \right)^{-1}, \quad \text{with } \varpi := (\gamma \gamma')^{-1}. \quad (5.20)$$

To verify this claim, we use the matrix inversion lemma to prove that

$$\begin{aligned} u &= I - \gamma' (v^{-1} + \gamma \gamma')^{-1} \gamma. \\ \implies \gamma u \gamma' &= \gamma \gamma' - \gamma \gamma' (v^{-1} + \gamma \gamma')^{-1} \gamma \gamma' = \left((\gamma \gamma')^{-1} + v \right)^{-1} = (\varpi + v)^{-1} \end{aligned}$$

Applying the above to $(u, v) = (v_{2n+1}, v_{2n})$ and then to $(u, v) = (v_{2n}, v_{2n-1})$ one has.

Theorem 5.7 ([3]). *For any $n \geq 1$ we have*

$$v_{2n} = \text{Ricc}_{\varpi} (v_{2(n-1)}) \quad \text{and} \quad v_{2n+1} = \text{Ricc}_{\overline{\varpi}} (v_{2n-1}), \quad (5.21)$$

with the Riccati maps $(\text{Ricc}_{\varpi}, \text{Ricc}_{\overline{\varpi}})$ defined in (1.21) and the matrices

$$(\varpi^{-1}, \overline{\varpi}^{-1}) := (\gamma \gamma', \gamma' \gamma). \quad (5.22)$$

The exponential decays of Riccati matrices towards the positive fixed points of the Riccati maps follows the contraction analysis discussed in (1.26). In addition, Corollary 4.4 in [3] yields for any $n \geq 1$ the uniform estimates

$$\overline{\sigma}^{1/2} (I + \varpi^{-1})^{-1} \overline{\sigma}^{1/2} \leq \tau_{2n} \leq \overline{\sigma} \quad \text{and} \quad \sigma^{1/2} (I + \overline{\varpi}^{-1})^{-1} \sigma^{1/2} \leq \tau_{2n+1} \leq \sigma. \quad (5.23)$$

5.4.4 Schrödinger bridges

Note that the transitions of the Gibbs-loop process introduced in (1.9) for even indices can be rewritten as

$$\mathcal{S}_{2n}(y_1, dy_2) := \mathbb{P} (Z_{2n}^{\circ}(y_1) \in dy_2)$$

with the random maps

$$\mathcal{Z}_{2n}^{\circ}(y) := \overline{m} + \beta_{2n}^{\circ}(y - \overline{m}) + (\tau_{2n}^{\circ})^{1/2} G$$

defined in terms of the parameters

$$\beta_{2n}^{\circ} := \beta_{2n} \beta_{2n-1}, \quad \text{and} \quad \tau_{2n}^{\circ} := \tau_{2n} + \beta_{2n} \tau_{2n-1} \beta_{2n}'.$$

Consider the directed matrix products

$$\beta_{2n,0}^{\circ} := \beta_{2n}^{\circ} \beta_{2(n-1)}^{\circ} \cdots \beta_2^{\circ} \quad (5.24)$$

Then we have

$$m_{2n} - \overline{m} = \beta_{2n}^{\circ} (m_{2(n-1)} - \overline{m}) \quad \text{and} \quad \sigma_{2n} - \overline{\sigma} = \beta_{2n,0}^{\circ} (\sigma_0 - \overline{\sigma}) (\beta_{2n,0}^{\circ})'. \quad (5.25)$$

The l.h.s. assertion in (5.25) is a direct consequence of (5.16) and the detailed proof of the covariance formula (5.25) is provided in Appendix C in [3]. By [3, Lemma 4.9], we have

$$0 \leq \overline{\sigma}^{-1/2} \beta_{2n}^{\circ} \overline{\sigma}^{1/2} = I - v_{2n} \leq (I + \varpi)^{-1}.$$

This shows that

$$\lim_{n \rightarrow \infty} \|\beta_{2n,0}^\circ\| = 0 \quad \text{and therefore} \quad (m_{2n}, \sigma_{2n}) \longrightarrow_{n \rightarrow \infty} (\bar{m}, \bar{\sigma})$$

More refined estimates are provided in Section 5.4.5.

In terms of the rescaled matrices (5.18) combining (5.14) with (5.15) we check that

$$\mathcal{Z}_{2n}(x) = m_{2n} + \tau_{2n} \tau^{-1} \beta (x - m) + \tau_{2n}^{1/2} G \quad \text{with} \quad \tau_{2n} = \bar{\sigma}^{1/2} v_{2n} \bar{\sigma}^{1/2}$$

with the solution v_{2n} of the Riccati equation (5.21). Taking the limit as $n \rightarrow \infty$ one has the following theorem.

Theorem 5.8 ([3]). *The Schrödinger bridge from $\mu = \nu_{m,\sigma}$ to $\eta = \nu_{\bar{m},\bar{\sigma}}$ with respect to the reference measure $P := (\mu \times K)$ is given by the formula*

$$P_{\mu,\eta}(d(x, y)) = \mu(dx) \mathbb{P}(Z(x) \in dy)$$

with the random map

$$Z(x) = \bar{m} + \varsigma \tau^{-1} \beta (x - m) + \varsigma^{1/2} G \tag{5.26}$$

and the parameters $\varsigma := \bar{\sigma}^{1/2} r \bar{\sigma}^{1/2}$ and the matrix $r = \text{Ricc}_\varpi(r)$ as in (1.25).

The transport property stems from the equivalences

$$\begin{aligned} (\varsigma \tau^{-1} \beta) \sigma (\varsigma \tau^{-1} \beta)' + \varsigma_\theta = \bar{\sigma} &\iff \varsigma + \varsigma \left(\bar{\sigma}^{1/2} \varpi \bar{\sigma}^{1/2} \right)^{-1} \varsigma = \bar{\sigma} \\ &\iff r \varpi^{-1} r + r = I \iff (1.23). \end{aligned}$$

When $\beta = I$ and τ is diagonal the above formula reduces to [19, Eq (2)]. Theorem 5.8 can be seen as a simplification and a generalization of Schrödinger bridge formulae recently presented in the series of articles [1, 16, 44, 83, 90] when the drift matrix β is arbitrary and τ is an arbitrary positive definite matrix.

5.4.5 A convergence theorem

We only consider Sinkhorn bridges $\mathcal{P}_{2n}$ associated with even indices, the analysis of the bridges associated with odd indices follows the same line of arguments, thus it is skipped (further details can be found in Section 5 in [3]).

Recall that Sinkhorn random maps $\mathcal{Z}_{2n}(x)$ defined in (5.14) and the random map $Z(x)$ of the Schrödinger bridge defined in (5.26) take the form

$$\begin{aligned} \mathcal{Z}_{2n}(x) &= m_{2n} + \tau_{2n} \tau^{-1} \beta (x - m) + \tau_{2n}^{1/2} G \quad \text{with} \quad \tau_{2n} := \bar{\sigma}^{1/2} v_{2n} \bar{\sigma}^{1/2} \\ Z(x) &= \bar{m} + \varsigma \tau^{-1} \beta (x - m) + \varsigma^{1/2} G \quad \text{with} \quad \varsigma := \bar{\sigma}^{1/2} r \bar{\sigma}^{1/2}. \end{aligned} \tag{5.27}$$

In addition, the matrix flow v_{2n} satisfies the Riccati equation

$$v_{2n} = \text{Ricc}_\varpi(v_{2(n-1)}) \longrightarrow_{n \rightarrow \infty} \text{Ricc}_\varpi(r) = r. \tag{5.28}$$

with the fixed point $\text{Ricc}_\varpi(r) = r$ of the Riccati map given in closed form in (1.25). The next result is a consequence of (1.26) and the Ando-Hemmen inequality (1.20).

Theorem 5.9 ([3]). *There exists some $c < \infty$ such that for any $n \geq 0$ we have*

$$\|\tau_{2n} - \varsigma\| \vee \|\tau_{2n}^{1/2} - \varsigma^{1/2}\| \leq c (1 + \lambda_{\min}(r + \varpi))^{-2n} \|\tau - \varsigma\|$$

with (ϖ, r, ς) as in (5.22) and (5.26). In addition, we have

$$\|\bar{\sigma}^{-1/2} \beta_{2n,0}^\circ \bar{\sigma}^{1/2}\| \leq c (1 + \lambda_{\min}(r + \varpi))^{-n}.$$

Theorem 5.9 can be used to derive a variety of quantitative estimates. For instance, using (3.30) we check the entropy formula

$$2\mathcal{H}(\mathcal{P}_{2n} \mid P_{\mu,\eta}) = D_{\text{burg}}(\tau_{2n} \mid \varsigma) + \|\varsigma^{-1/2} (m_{2n} - \bar{m})\|_2^2 + \|\varsigma^{-1/2}(\tau_{2n} - \varsigma) (\tau^{-1} \beta \sigma^{1/2})\|_F^2 \quad (5.29)$$

with the Burg distance D_{burg} defined in (3.31) and the Schrödinger bridge $P_{\mu,\eta}$ defined in (5.26). For a more detailed discussion we refer the reader to [3, Section 5.2]. We also note that by [3, Lemma B.1] (see also [55, Section 11]) we have

$$\|\tau_{2n} - \varsigma\|_F \leq \frac{1}{2\|\varsigma^{-1}\|_F} \implies D_{\text{burg}}(\tau_{2n} \mid \varsigma) \leq \frac{5}{2} \|\varsigma^{-1}\|_F \|\tau_{2n} - \varsigma\|_F.$$

Corollary 5.10. *There exists some finite $c < \infty$ and some $n_0 \geq 0$ such that for any $n \geq n_0$ we have the entropy estimates*

$$\mathcal{H}(\mathcal{P}_{2n} \mid P_{\mu,\eta}) \leq c (1 + \lambda_{\min}(r + \varpi))^{-2n} (\|\tau - \varsigma\| + \|m_0 - \bar{m}\|^2).$$

Corollary 5.11. *For any $p \geq 1$ there exists some finite $c < \infty$ such that for any $n \geq 0$ we have the p -Wasserstein estimates*

$$\mathcal{D}_p(\mathcal{P}_{2n} \mid P_{\mu,\eta}) \leq c (1 + \lambda_{\min}(r + \varpi))^{-2n} \|\tau - \varsigma\|_F + c (1 + \lambda_{\min}(r + \varpi))^{-n} \|m_0 - \bar{m}\|_F.$$

5.5 Log-concave-at-infinity models

Consider the marginal measures (μ, η) the Markov transition K discussed in Section 4.4 on the state spaces $\mathbb{X} = \mathbb{R}^d = \mathbb{Y}$ equipped with the Lebesgue measure $\lambda(dz) = dz = \nu(dz)$. Also assume that the log density W is differentiable. In this case, by (4.11) for any $n \geq 0$ we have

$$\begin{aligned} \nabla_2 W_{2n}(x_1, y) - \nabla_2 W_{2n}(x_2, y) &= \nabla_2 W(x_1, y) - \nabla_2 W(x_2, y) \\ \nabla_2 W_{2n+1}^b(y_1, x) - \nabla_2 W_{2n+1}^b(y_2, x) &= \nabla_2 W^b(y_1, x) - \nabla_2 W^b(y_2, x). \end{aligned} \quad (5.30)$$

To simplify the presentation, we restrict our attention to linear Gaussian transition potential of the form (3.32) and assume that U and V satisfy the strongly convex outside a ball condition (3.28). This condition ensures that the probability measures (μ, η) satisfy the $LS(\rho)$ inequality and thus the $\mathbb{T}_2(\rho)$ inequality for some $\rho > 0$. More general situations are discussed in [49].

We have that

$$\begin{aligned} \nabla_2 W(x_1, y) - \nabla_2 W(x_2, y) &= \tau^{-1} \beta (x_2 - x_1) \\ \nabla_2 W^b(y_1, x) - \nabla_2 W^b(y_2, x) &= \beta' \tau^{-1} (y_2 - y_1). \end{aligned}$$

Thus, for any $n \geq 0$ we have

$$\mathcal{J}(\delta_{z_1} \mathcal{K}_n \mid \delta_{z_2} \mathcal{K}_n) \leq \kappa^2 \|z_1 - z_2\|_2^2 \quad \text{with} \quad \kappa := \|\tau^{-1} \beta\|_2. \quad (5.31)$$

This shows that Sinkhorn transitions $\mathcal{K}_n$ satisfy the $\mathbb{J}(\|\tau^{-1} \beta\|_2)$ inequality.

Example 5.12. *In the context of the linear-Gaussian transitions discussed in (2.10):*

- *When $A = 0$ we have*

$$\beta(t) = I \quad \text{and} \quad \tau(t) = t \Sigma \implies \kappa(t) := \|\tau(t)^{-1} \beta(t)\|_2 = \frac{1}{t} \|\Sigma^{-1}\|_2. \quad (5.32)$$

- *Whenever A is Hurwitz there exists some parameters $c, a > 0$ such that*

$$\begin{aligned} \|\beta(t)\|_2 &\leq c e^{-at} \quad \text{and} \quad \forall t \geq t_0 \quad \tau(t_0) \leq \tau(t) \leq \tau^+ := \int_0^\infty \beta(s) \Sigma \beta(s)' ds \\ \implies \exists c_0 > 0 : \forall t \geq t_0 \quad \kappa(t) &:= \|\tau(t)^{-1} \beta(t)\|_2 \leq c_0 e^{-at}. \end{aligned} \quad (5.33)$$

5.5.1 Curvature estimates

Assume that (U, V) are twice differentiable. Using (1.15) and (4.9) for any $n \geq 0$ we also check that

$$\begin{aligned} \nabla^2 V_{2n+1}(y) &= \nabla^2 V(y) - \mathcal{K}_{2n+1}(W_{y,y}^b)(y) + \text{cov}_{\mathcal{K}_{2n+1}, W^b}(y) \\ \nabla^2 U_{2(n+1)}(x) &= \nabla^2 U(x) - \mathcal{K}_{2(n+1)}(W_{x,x})(x) + \text{cov}_{\mathcal{K}_{2(n+1)}, W}(x). \end{aligned} \quad (5.34)$$

Combining the above with (4.11) we have

$$\nabla_2^2 W_{2(n+1)}(x, y) = \nabla^2 V(y) + W_{y,y}^b(x) - \mathcal{K}_{2n+1}(W_{y,y}^b)(y) + \text{cov}_{\mathcal{K}_{2n+1}, W^b}(y)$$

with the conditional covariance functions $\text{cov}_{\mathcal{K}_{2n+1}, W^b}$ defined in (1.16). One also has

$$\nabla_2^2 W_1^b(y, x) = W_{x,x}(y) + \nabla^2 U(x).$$

In addition, for $n \geq 1$ we have

$$\nabla_2^2 W_{2n+1}^b(y, x) = \nabla^2 U(x) + W_{x,x}(y) - \mathcal{K}_{2n}(W_{x,x})(x) + \text{cov}_{\mathcal{K}_{2n}, W}(x).$$

For linear Gaussian transition potential of the form (3.32) the above decompositions resume to (4.29). In this situation, for any $n \geq 1$ we have

$$\nabla_2^2 W_{2n-1}^b(y, x) \geq \nabla^2 U(x) \quad \text{and} \quad \nabla_2^2 W_{2n}(x, y) \geq \nabla^2 V(y). \quad (5.35)$$

Thus, for linear Gaussian transition (3.32) potentials (U, V) satisfying the convex-at-infinity condition (3.28), the Markov transitions $\mathcal{K}_n$ with $n \geq 1$ satisfy the $LS(\rho)$ and thus the $\mathbb{T}_2(\rho)$ inequality. In the context of linear-Gaussian models, compare (5.35) with (5.14) and the uniform estimates (5.23).

Applying Lemma 3.3 to Sinkhorn transition $\mathcal{K}_{2n} = K_{V_{2n}}$ and the Schrödinger bridge transition $L = K_{\mathbb{V}}$ defined respectively in (4.10) and (4.14) one can prove the following result.

Theorem 5.13 ([50]). *For any $n \geq 1$, we have the bias and covariance entropic bounds estimate*

$$\int \mu(dx) \|\text{mean}_{K_{V_{2n}}}(x) - \text{mean}_{K_V}(x)\|_2^2 \leq 2\rho \mathcal{H}(P_{\mu,\eta} \mid \mathcal{P}_{2n})$$

as well as the covariance entropic bounds

$$\begin{aligned} & \int \mu(dx) \|\text{cov}_{K_{V_{2n}}}(x) - \text{cov}_{K_V}(x)\|_F \\ & \leq 4\rho \mathcal{H}(P_{\mu,\eta} \mid \mathcal{P}_{2n}) + c (8\rho \mathcal{H}(P_{\mu,\eta} \mid \mathcal{P}_{2n}))^{1/2} \quad \text{with } c^2 := \mu(\text{Tr}(\text{cov}_{K_V})) \end{aligned}$$

5.5.2 Contraction inequalities

Using (1.3) and (4.20) for any $n \geq 0$ we check that

$$P_{\mu,\eta} = (\mathcal{P}_{2n})_{\mu,\eta} \quad \text{and} \quad P_{\eta,\mu}^\flat = (\mathcal{P}_{2n+1}^\flat)_{\eta,\mu}.$$

Recalling that

$$\pi_{2n} \mathcal{K}_{2n+1} = \mu \quad \text{and} \quad \pi_{2n-1} \mathcal{K}_{2n} = \eta$$

we also have the bridge map formulae

$$\begin{aligned} \mathcal{P}_{2n} &= \mu \times \mathcal{K}_{2n} = (\mathcal{P}_{2n})_{\mu,\mu\mathcal{K}_{2n}} & P_{\mu,\eta} &= (\mathcal{P}_{2n})_{\mu,\eta} = (\mathcal{P}_{2n})_{\mu,\pi_{2n-1}\mathcal{K}_{2n}} \\ \mathcal{P}_{2n+1}^\flat &= \eta \times \mathcal{K}_{2n+1} = (\mathcal{P}_{2n+1}^\flat)_{\eta,\eta\mathcal{K}_{2n+1}} & P_{\eta,\mu}^\flat &= (\mathcal{P}_{2n+1}^\flat)_{\eta,\mu} = (\mathcal{P}_{2n+1}^\flat)_{\eta,\pi_{2n}\mathcal{K}_{2n+1}}. \end{aligned}$$

Recall that the probability measures (μ, η) as well as the collection of Sinkhorn transitions $\mathcal{K}_n$ satisfy for any $n \geq 1$ the $LS(\rho)$ and thus the $\mathbb{T}_2(\rho)$ inequalities for some common constant ρ . However, by (5.31), Sinkhorn transitions $\mathcal{K}_n$ also satisfy the Fisher-Lipschitz inequality (3.27) with $\kappa := \|\tau^{-1}\beta\|_2$.

Next we present some direct consequences of Theorem 3.4:

- Applying (3.37) to $(P, K, \eta) = (\mathcal{P}_{2n}, \mathcal{K}_{2n}, \pi_{2n-1})$ we check that

$$\mathcal{H}(\eta \mid \pi_{2n}) \leq \mathcal{H}(P_{\mu,\eta} \mid \mathcal{P}_{2n}) \leq \epsilon_\kappa(\rho) \mathcal{H}(\pi_{2n-1} \mid \mu) \quad \text{with } \epsilon_\kappa(\rho) := (\kappa\rho)^2 \quad (5.36)$$

Thus, by (4.24) we have

$$\begin{aligned} \mathcal{H}(P_{\mu,\eta} \mid \mathcal{P}_{2n}) &\leq \mathcal{H}(P_{\mu,\eta} \mid \mathcal{P}_{2(n-1)}) - \mathcal{H}(\pi_{2n-1} \mid \mu) \\ &\leq \mathcal{H}(P_{\mu,\eta} \mid \mathcal{P}_{2(n-1)}) - \epsilon_\kappa(\rho)^{-1} \mathcal{H}(P_{\mu,\eta} \mid \mathcal{P}_{2n}) \end{aligned} \quad (5.37)$$

Similarly, by (3.38) we have the 2-Wasserstein estimates

$$\mathcal{D}_2(\pi_{2n}, \eta) \leq \kappa\rho \mathcal{D}_2(\pi_{2n-1}, \mu). \quad (5.38)$$

- Applying (3.37) and (3.38) to $(P, K, \eta) = (\mathcal{P}_{2n+1}^\flat, \mathcal{K}_{2n+1}, \pi_{2n})$ we check that

$$\mathcal{H}(\mu \mid \pi_{2n+1}) \leq \mathcal{H}(P_{\mu,\eta} \mid \mathcal{P}_{2n+1}) \leq \epsilon_\kappa(\rho) \mathcal{H}(\pi_{2n} \mid \eta) \quad (5.39)$$

Thus, by (4.25) we have

$$\begin{aligned} \mathcal{H}(P_{\mu,\eta} \mid \mathcal{P}_{2n+1}) &\leq \mathcal{H}(P_{\mu,\eta} \mid \mathcal{P}_{2n}) - \mathcal{H}(\pi_{2n} \mid \eta) \\ &\leq \mathcal{H}(P_{\mu,\eta} \mid \mathcal{P}_{2n}) - \epsilon_\kappa(\rho)^{-1} \mathcal{H}(P_{\mu,\eta} \mid \mathcal{P}_{2n+1}). \end{aligned} \quad (5.40)$$

By (3.38) we have the 2-Wasserstein estimates

$$\mathcal{D}_2(\pi_{2n+1}, \mu) \leq \kappa \rho \mathcal{D}_2(\pi_{2n}, \eta). \quad (5.41)$$

Using (5.37) and (5.40) one has the following theorem.

Theorem 5.14. *For any $n \geq 0$ we have the entropic exponential decays*

$$\mathcal{H}(P_{\mu, \eta} | \mathcal{P}_{2n}) \vee \mathcal{H}(P_{\mu, \eta} | \mathcal{P}_{2n+1}) \leq (1 + 1/\epsilon_\kappa(\rho))^{-n} \mathcal{H}(P_{\mu, \eta} | \mathcal{P}_0) \quad (5.42)$$

with $\epsilon_\kappa(\rho)$ as in (5.36). In addition, when $\epsilon_\kappa(\rho) < 1$ we have Wasserstein exponential decays

$$\mathcal{D}_2(\pi_{2n}, \eta) \leq \epsilon_\kappa(\rho)^n \mathcal{D}_2(\pi_0, \eta) \quad \text{and} \quad \mathcal{D}_2(\pi_{2n+1}, \mu) \leq \epsilon_\kappa(\rho)^n \mathcal{D}_2(\pi_1, \mu). \quad (5.43)$$

More refined estimates are provided in [49], see also Section 5.5.3.

Example 5.15. *In the context of the linear-Gaussian transitions discussed in (2.10) and in Example 5.12 we have*

$$(5.32) \implies \epsilon_{\kappa(t)}(\rho)^{1/2} = \kappa(t)\rho = \frac{\rho}{t} \|\Sigma^{-1}\|_2 \quad \text{and} \quad (5.33) \implies \epsilon_{\kappa(t)}(\rho)^{1/2} \leq c_0 \rho e^{-at}.$$

5.5.3 Some quantitative estimates

For instance, combining (3.29) with (5.35) we also check that

$$\begin{aligned} & \exists \sigma, \bar{\sigma} \in \mathcal{S}_d^+ \quad \text{such that} \quad \forall x \in \mathbb{R}^d \quad \nabla^2 U(x) \geq \sigma^{-1} \quad \text{and} \quad \nabla^2 V(x) \geq \bar{\sigma}^{-1} \\ \implies & \begin{cases} \lambda_U \text{ and } \mathcal{K}_{2n+1} \text{ satisfy the } LS(\rho) \text{ inequality with } \rho = \|\sigma\|_2 \\ \lambda_V \text{ and } \mathcal{K}_{2n} \text{ satisfy the } LS(\bar{\rho}) \text{ inequality with } \bar{\rho} = \|\bar{\sigma}\|_2. \end{cases} \end{aligned} \quad (5.44)$$

More generally, assume (U, V) satisfying the convex-at-infinity conditions

$$\nabla^2 U(x) \geq a \mathbf{1}_{\|x\|_2 \geq l} I - b \mathbf{1}_{\|x\|_2 < l} I \quad \text{and} \quad \nabla^2 V(x) \geq \bar{a} \mathbf{1}_{\|x\|_2 \geq \bar{l}} I - \bar{b} \mathbf{1}_{\|x\|_2 < \bar{l}} I$$

for some parameters $a, \bar{a} > 0$, $b, \bar{b} \geq 0$ and $l, \bar{l} \geq 0$. In this context, combining (3.28) with (5.35) there exists some parameters $\rho(a, b, l) \wedge \rho(\bar{a}, \bar{b}, \bar{l}) > 0$ such that

$$\begin{cases} \lambda_U \text{ and } \mathcal{K}_{2n+1} \text{ satisfy the } LS(\rho) \text{ inequality with } \rho = \rho(a, b, l) \\ \lambda_V \text{ and } \mathcal{K}_{2n} \text{ satisfy the } LS(\bar{\rho}) \text{ inequality with } \bar{\rho} := \rho(\bar{a}, \bar{b}, \bar{l}). \end{cases} \quad (5.45)$$

Arguing as above we check that (5.36) and (5.39) and thus (5.42) hold with the parameter

$$\epsilon_\kappa(\rho) := \kappa^2 \rho(a, b, l) \rho(\bar{a}, \bar{b}, \bar{l}). \quad (5.46)$$

The estimates (5.38) and resp. (5.41) also holds with the parameter $\rho = \rho(\bar{a}, \bar{b}, \bar{l})$ and resp. $\rho = \rho(a, b, l)$, thus (5.43) also holds with $\epsilon_\kappa(\rho)$ as in (5.46).

One can also show that the entropic exponential decays presented in (5.42) can be refined. We set

$$I_n := \mathcal{H}(P_{\mu, \eta} | \mathcal{P}_n) \quad \text{and} \quad \varepsilon := \epsilon_\kappa(\rho).$$

Combining (4.24) with Proposition 4.1 and (5.39) we check that

$$I_{2n} = I_{2n-1} - \mathcal{H}(\mu \mid \pi_{2n-1}) \leq I_{2n-1} - \mathcal{H}(\pi_{2n} \mid \eta) \leq I_{2n-1} - \varepsilon^{-1} I_{2n+1}$$

Now, using (4.25) along with Proposition 4.1 and (5.36) we have

$$I_{2n+1} = I_{2n} - \mathcal{H}(\eta \mid \pi_{2n}) \leq I_{2n} - \mathcal{H}(\pi_{2n+1} \mid \mu) \leq I_{2n} - \varepsilon^{-1} I_{2(n+1)}.$$

In summary we have proved that for any $n \geq 0$ the second order linear homogeneous recurrence relation

$$I_n \geq I_{n+1} + \varepsilon^{-1} I_{n+2} \geq (1 + \varepsilon^{-1}) I_{n+2}. \quad (5.47)$$

These linear recurrence equations are rather well known, for instance following the proof of [49, Theorem 6.9], for any $n \geq 0$ we have

$$\mathcal{H}(P_{\mu,\eta} \mid \mathcal{P}_{2(n+1)}) \leq (1 + 1/\varepsilon + 1/\bar{\varepsilon})^{-n} \mathcal{H}(P_{\mu,\eta} \mid \mathcal{P}_0)$$

with the parameter

$$1 + 1/\bar{\varepsilon} := \frac{1 + \sqrt{1 + 4/\varepsilon}}{2} \implies (1 + 1/\bar{\varepsilon})^2 = 1 + 1/\varepsilon + 1/\bar{\varepsilon} > 1 + 1/\varepsilon.$$

5.5.4 Strongly convex models

Assume there exist some positive definite matrices $\sigma \geq \sigma_-$ and $\bar{\sigma} \geq \bar{\sigma}_-$ such that

$$\sigma^{-1} \leq \nabla^2 U \leq \sigma_-^{-1} \quad \text{and} \quad \bar{\sigma}^{-1} \leq \nabla^2 V \leq \bar{\sigma}_-^{-1}.$$

This rather strong convexity condition [28, 33, 49, 34] is often used to improve the curvature estimates in (4.29) and (5.35). This section underlines some links between these studies and the Riccati matrix difference equations discussed in Section 5.4.3 in the context of linear-Gaussian models.

It should be noted that for linear Gaussian transition potential of the form (3.32) the decompositions (4.29) yield for any $n \geq 0$ the estimates

$$\begin{aligned} \sigma^{-1} + \chi' \text{cov}_{\mathcal{K}_{2n}}(x) \chi &\leq \nabla_2^2 W_{2n+1}^b(y, x) \leq \sigma_-^{-1} + \chi' \text{cov}_{\mathcal{K}_{2n}}(x) \chi \\ \bar{\sigma}^{-1} + \chi \text{cov}_{\mathcal{K}_{2n+1}}(y) \chi' &\leq \nabla_2^2 W_{2(n+1)}(x, y) \leq \bar{\sigma}_-^{-1} + \chi \text{cov}_{\mathcal{K}_{2n+1}}(y) \chi'. \end{aligned}$$

Applying (1.19) to the Markov transition $\mathcal{K}_1$ we have

$$\begin{aligned} \text{cov}_{\mathcal{K}_0}(x) = \tau &\implies \tau_1^{-1} := \sigma^{-1} + \chi' \tau \chi \leq \nabla_2^2 W_1^b \leq \tau_{1-}^{-1} := \sigma_-^{-1} + \chi' \tau \chi \\ &\implies \tau_{1-} \leq \text{cov}_{\mathcal{K}_1} \leq \tau_1 \end{aligned}$$

with χ as in (4.26). Applying (1.19) to the Markov transition $\mathcal{K}_2$, this implies that

$$\begin{aligned} \tau_2^{-1} := \bar{\sigma}^{-1} + \chi \tau_{1-} \chi' &\leq \nabla_2^2 W_2 \leq \tau_{2-}^{-1} := \bar{\sigma}_-^{-1} + \chi \tau_1 \chi' \\ \implies \tau_{2-} &\leq \text{cov}_{\mathcal{K}_2} \leq \tau_2. \end{aligned}$$

Iterating the procedure, we have the covariance estimates

$$\tau_{n-} \leq \text{cov} \mathcal{K}_n \leq \tau_n$$

as well as the Riccati difference matrix inequalities

$$\begin{aligned} \tau_{2n+1}^{-1} &:= \sigma^{-1} + \chi' \tau_{(2n)-} \chi \leq \nabla_2^2 W_{2n+1}^b \leq \tau_{(2n+1)-}^{-1} := \sigma_-^{-1} + \chi' \tau_{2n} \chi \\ \tau_{2(n+1)}^{-1} &:= \bar{\sigma}^{-1} + \chi \tau_{(2n+1)-} \chi' \leq \nabla_2^2 W_{2(n+1)} \leq \tau_{2(n+1)-}^{-1} := \bar{\sigma}_-^{-1} + \chi \tau_{2n+1} \chi'. \end{aligned}$$

For instance, if we set

$$\bar{\tau}_{2n} := \bar{\sigma}^{-1/2} \tau_{2n} \bar{\sigma}^{-1/2} \quad \text{and} \quad \bar{\tau}_{(2n+1)-} := \sigma_-^{-1/2} \tau_{(2n+1)-} \sigma_-^{-1/2}$$

we obtain

$$\begin{aligned} \bar{\tau}_{2(n+1)} &= (I + \gamma_- \bar{\tau}_{(2n+1)-} \gamma_-')^{-1} \leq I \\ \bar{\tau}_{(2n+1)-} &= (I + \gamma_-' \bar{\tau}_{2n} \gamma_-)^{-1} \quad \text{with} \quad \gamma_- := \bar{\sigma}^{1/2} \chi \sigma_-^{1/2}. \end{aligned}$$

Arguing as in (5.20) we conclude that

$$\bar{\tau}_{2(n+1)} = \text{Ric}_{\varpi_-}(\bar{\tau}_{2n}) \quad \text{with} \quad \varpi_-^{-1} := \gamma_- \gamma_-'.$$

The monotone properties (1.24) of Riccati maps provide several uniform upper bounds of the flow of covariance matrices $\bar{\tau}_{2n}$. These upper bounds also yield curvature estimates for the Hessian matrices $\nabla_2^2 W_{2n}$. For a more detailed discussion one can see [49, Appendix D].

5.6 Lyapunov techniques

5.6.1 Regularity conditions

Consider the marginal measures $(\mu, \eta) = (\lambda_U, \nu_V)$ and the integral operators K and K^b defined in (2.7) and (4.1) and (4.7). We assume U, V, W are locally bounded and $\|e^{-W}\| < \infty$. In addition U, V have compact sub-level sets. These conditions ensures that (U, V, W) are lower bounded by real numbers (U_*, V_*, W_*) so that (e^{-U}, e^{-V}) are uniformly bounded. We again recall the notation $\mathcal{B}_\infty(\mathbb{X})$ and $\mathcal{B}_0(\mathbb{X})$ from section 1.5. We remark that for any $\delta > 0$ we have $(e^{\delta U}, e^{\delta V}) \in (\mathcal{B}_\infty(\mathbb{X}) \times \mathcal{B}_\infty(\mathbb{Y}))$ and set

$$\mathcal{U}_\delta := \delta U - W^V \quad \text{and} \quad \mathcal{V}_\delta := \delta V - W_U \quad (5.48)$$

with the functions

$$W^V(x) := \int \nu_V(dy) W(x, y) \quad \text{and} \quad W_U(y) := \int \lambda_U(dx) W(x, y).$$

Consider the following condition:

(A) : *There exist $\delta \in]0, 1/2[$ and functions $(\theta_u, \theta_v) \in (\mathcal{B}_0(\mathbb{X}) \times \mathcal{B}_0(\mathbb{Y}))$ such that*

$$\lambda_{(1-2\delta)U}(1) \vee \nu_{(1-2\delta)V}(1) < \infty \quad \text{as well as} \quad e^{-\mathcal{U}_\delta} \leq \theta_u \quad \text{and} \quad e^{-\mathcal{V}_\delta} \leq \theta_v. \quad (5.49)$$

Condition (A) is slightly weaker than condition $\mathcal{H}'_\delta(U, V)$ discussed in [4]. Note that (A) is met with $(\theta_u, \theta_v) = (e^{-\mathcal{U}_\delta}, e^{-\mathcal{V}_\delta})$ as soon as the functions $(\mathcal{U}_\delta, \mathcal{V}_\delta)$ are locally bounded with compact sub-level sets. Intuitively, condition (A) ensure that marginal-potentials (U, V) dominate at infinity the (integrated) transition-potentials (W_U, W^V) . In terms of potential energy conversion, the marginal potentials should be large enough to compensate high energy transition losses.

Remark 5.16. Assume that $(\mathcal{U}_\delta, \mathcal{V}_\delta)$ are lower bounded by functions $(\mathcal{U}_\delta^-, \mathcal{V}_\delta^-)$. Note that in this case for any $l > 0$ we have

$$\{\mathcal{U}_\delta \leq l\} \subset \{\mathcal{U}_\delta^- \leq l\} \quad \text{and} \quad \{\mathcal{V}_\delta \leq l\} \subset \{\mathcal{V}_\delta^- \leq l\}.$$

In this scenario, we can choose $(\theta_u, \theta_v) := (e^{-\mathcal{U}_\delta^-}, e^{-\mathcal{V}_\delta^-}) \in (\mathcal{B}_0(\mathbb{X}) \times \mathcal{B}_0(\mathbb{Y}))$ as soon as the functions $(\mathcal{U}_\delta^-, \mathcal{V}_\delta^-)$ are locally bounded with compact sub-level sets.

Assume that $W(x, y) \leq a_W(x) + b_W(y)$ for some locally bounded functions (a_W, b_W) such that $\lambda_U(a_W) \vee \nu_V(b_W) < \infty$. Then, we have

$$\mathcal{U}_\delta^-(x) := \delta U(x) - a_W(x) - \nu_V(b_V) \leq \mathcal{U}_\delta(x) \leq \delta U(x) - W_\star.$$

This shows that $(\mathcal{U}_\delta, \mathcal{U}_\delta^-)$ and similarly $(\mathcal{V}_\delta, \mathcal{V}_\delta^-)$ with $\mathcal{V}_\delta^- := \delta V - b_W$ are locally bounded. Finally, $(\mathcal{U}_\delta^-, \mathcal{V}_\delta^-)$ have compact compact sub-level sets if and only if $(\delta U - a_W, \delta V - b_W)$ have compact compact sub-level sets.

Remark 5.17. When $\|W\| < \infty$ we have $w := \|W^V\| \vee \|W_U\| < \infty$. In this case, we have

$$e^{-\mathcal{U}_\delta} \leq \theta_u = e^w e^{-\delta U} \in \mathcal{B}_0(\mathbb{X}) \quad \text{and} \quad e^{-\mathcal{V}_\delta} \leq \theta_v = e^w e^{-\delta V} \in \mathcal{B}_0(\mathbb{Y})$$

The following examples are taken from [4, 51].

Example 5.18. Consider a Beta distribution on the open unit interval $\mathbb{X} :=]0, 1[$ with shape parameters $a_u, b_u > 0$ defined by

$$\lambda_U(dx) = c_u x^{a_u} (1-x)^{b_u} \lambda(dx) \quad \text{with} \quad \lambda(dx) = 1_{]0, 1[}(x) dx$$

and some normalizing constant $c_u > 0$. In this situation, for any $\delta > 0$ we have

$$\implies U(x) = a_u \log \frac{1}{x} + b_u \log \frac{1}{1-x} + \log \frac{1}{c_u} \quad \text{and} \quad e^{-\delta U(x)} = c_u^\delta x^{\delta a_u} (1-x)^{\delta b_u}$$

By Remark 5.17, when W is bounded we have $w := \|W^V\| < \infty$ as well as the estimate

$$e^{-\mathcal{U}_\delta} \leq \theta_u = c_{u,w} x^{\delta a_u} (1-x)^{\delta b_u} \in \mathcal{B}_0(\mathbb{X}) \quad \text{with} \quad c_{u,w} := e^w c_u^\delta$$

Thus, condition (A) is met for any $\delta \in]0, 1/2[$ as soon as W is bounded and (λ_U, ν_V) are both Beta distributions on $\mathbb{X} =]0, 1[= \mathbb{Y}$ with shape parameters $a_u, b_u > 0$ and $a_v, b_v > 0$.

Example 5.19. Assume that there exist some parameters $a_1, b_1 \in \mathbb{R}$, $a_2, b_2 > 0$ and $q \geq p > 0$ such that for any $x, y \in \mathbb{X} = \mathbb{Y} = \mathbb{R}^d$ we have

$$W(x, y) \leq a_1 + a_2 (\|x\|^p + \|y\|^p) \quad \text{and} \quad U(x) \wedge V(x) \geq b_1 + b_2 \|x\|^q.$$

In this case, there exists some constant $c \in \mathbb{R}$ such that

$$\mathcal{U}_\delta(x) \wedge \mathcal{V}_\delta(x) \geq c + \delta b_2 \|x\|^q - a_2 \|x\|^p.$$

We conclude that (5.49) is satisfied for any $\delta \in]0, 1/2[$ as soon as $q > p$, and for any $\delta \in]a_2/b_2, 1/2[$ as soon as $b_2 > 2a_2$ and $q = p$. Indeed, in both situations we have

$$e^{-\mathcal{U}_\delta(x)} \vee e^{-\mathcal{V}_\delta(x)} \leq \theta(x) := e^{-c} \exp \left(-\delta b_2 \|x\|^q \left(1 - \frac{a_2}{\delta b_2} \frac{1}{\|x\|^{q-p}} \right) \right) \xrightarrow{\|x\| \rightarrow \infty} 0$$

A wealth of examples satisfying these regularity conditions is presented in [4, Section 5] ranging from polynomial growth potentials and heavy tailed marginals on general normed spaces to more sophisticated boundary state space models. The latter constitutes semi-circle transitions, Beta, Weibull, exponential marginals as well as semi-compact models. More precisely, in [4, Proposition 5.3] the Gaussian case is considered with extensions to general mixtures ([4, Proposition 5.4]) and particularly Gaussian mixtures ([4, Proposition 5.5]). The mixture results help one to consider kernel-type density estimators of complex data distributions arising in generative modeling.

5.6.2 A technical lemma

Consider the integral operators

$$K_V(x, dy) := e^{-W(x,y)} \nu_V(dy) \quad \text{and} \quad K_U^b(y, dx) := e^{-W^b(y,x)} \lambda_U(dx)$$

and set

$$W_{V,U} := \log K_V(\exp(W_U)) \quad \text{and} \quad W^{U,V} := \log K_U^b(\exp(W^V)).$$

Lemma 5.20. *Assume (A). Then we have that*

$$\|\exp W_{V,U}\| \vee \|\exp W^{U,V}\| < \infty \quad \text{and} \quad \|\exp W_{(1-\delta)V,U}\| \vee \|\exp W^{(1-\delta)U,V}\| < \infty. \quad (5.50)$$

In addition, for any $n \geq 0$ we have

$$\begin{aligned} -W^V + \eta(V) &\leq \log K(\exp(-V_{2n})) \\ -W_U &\leq \log K^b(\exp(-U_{2n})) \leq -\eta(V) + W^{U,V}. \end{aligned} \quad (5.51)$$

For any $n \geq 1$ we also have

$$\begin{aligned} U - W^V + \eta(V) &\leq U_{2n} \leq U + W_{V,U} \\ V - W_U &\leq V_{2n} \leq -\eta(V) + V + W^{U,V} \quad \text{and} \quad \log K(\exp(-V_{2n})) \leq W_{V,U}. \end{aligned} \quad (5.52)$$

Proof. Recalling that U is locally bounded with compact sub-level sets, we have

$$\lambda_{(1-\delta)U}(1) = \lambda_{(1-2\delta)U}(e^{-\delta U}) \leq \|e^{-\delta U}\| \lambda_{(1-2\delta)U}(1) < \infty \quad \text{and} \quad \nu_{(1-\delta)V}(1) < \infty.$$

When $c := \|e^{-W}\| < \infty$ for any $\delta \in]0, 1[$ we have

$$\begin{aligned} \exp(W_{V,U}) &\leq c \nu_V(e^{W_U}) = c \nu_{(1-\delta)V}(e^{-\mathcal{V}_\delta}) < \infty \implies \|e^{W_{V,U}}\| < \infty \\ \exp(W^{U,V}) &\leq c \lambda_U(e^{W^V}) = c \lambda_{(1-\delta)U}(e^{-\mathcal{U}_\delta}) < \infty \implies \|e^{W^{U,V}}\| < \infty. \end{aligned}$$

Similiarly, one has

$$\begin{aligned}\exp(W_{(1-\delta)V,U}) &= K_{(1-\delta)V}(e^{W_U}) \leq c \nu_{(1-\delta)V}(e^{W_U}) = c \nu_{(1-2\delta)V}(e^{-V_\delta}) \\ \exp(W^{(1-\delta)U,V}) &= K_{(1-\delta)U}^\flat(e^{W^V}) \leq c \lambda_{(1-\delta)U}(e^{W^V}) = c \lambda_{(1-2\delta)U}(e^{-U_\delta}).\end{aligned}$$

Applying Jensen's inequality and Proposition 4.4, for any $n \geq 0$ we have

$$\begin{aligned}\log K(\exp(-V_{2n}))(x) &= \log \int \eta(dy) \exp(-W(x,y) + V(y) - V_{2n}(y)) \\ &\geq -W^V(x) + \eta(V) - \eta(V_{2n}) \geq -W^V(x) + \eta(V).\end{aligned}$$

Note that

$$U_{2n} = U + \log K(e^{-V_{2n}}) \geq U - W^V + \eta(V).$$

In addition, for any $n \geq 0$ we check that

$$\begin{aligned}\log K^\flat(\exp(-U_{2n}))(y) &= \log \int \mu(dx) \exp(-W(x,y) + U(x) - U_{2n}(x)) \\ &\geq -W_U(y) + \mu(U) - \mu(U_{2n}) \geq -W_U(y)\end{aligned}$$

as well as

$$\log K^\flat(\exp(-U_{2n})) \leq -\eta(V) + \log K^\flat(\exp(-(U - W^V))) = -\eta(V) + W^{U,V}.$$

By (4.9) we also have

$$V - W_U \leq V_{2(n+1)} = V + \log K^\flat(e^{-U_{2n}}) \leq V - \eta(V) + W^{U,V}.$$

In a similar manner, we have

$$U - W^V + \eta(V) \leq U_{2(n+1)} = U + \log K(e^{-V_{2(n+1)}}) \leq U + W_{V,U}.$$

For any $n \geq 1$

$$\exp(-W^V + \eta(V)) \leq K(\exp(-V_{2n})) \leq K(\exp(-(V - W_U))) = K_V(\exp(W_U)).$$

This ends the proof of (5.52). ■

5.6.3 Some uniform estimates

In this section we assume (A) is satisfied and we set

$$(g, h) := (\exp(\delta U), \exp(\delta V)) \tag{5.53}$$

Proposition 5.21. *There exists some constant $c > 0$ such that, for any $n \geq 1$, we have*

$$\frac{d\pi_{2n}}{d\eta} \vee \frac{d\eta}{d\pi_{2n}} \leq c \exp(W_U) \quad \text{and} \quad \frac{d\pi_{2n+1}}{d\mu} \vee \frac{d\mu}{d\pi_{2n+1}} \leq c \exp(W^V). \tag{5.54}$$

We also have the uniform estimates

$$\|d\pi_{2n}/d\eta\|_h \vee \|d\eta/d\pi_{2n}\|_h \leq c \quad \text{and} \quad \|d\pi_{2n+1}/d\mu\|_g \vee \|d\mu/d\pi_{2n+1}\|_g \leq c \tag{5.55}$$

Proof. Using (4.12) and (5.52) for any $n \geq 0$ we have

$$-(W^V + W_{V,U} - \eta(V)) \leq U_{2(n+1)} - U_{2n} = \log \left(\frac{d\pi_{2n+1}}{d\mu} \right) \leq W^V + W_{V,U} - \eta(V)$$

as well as

$$-(W_U + W^{U,V} - \eta(V)) \leq V_{2(n+1)} - V_{2n} = \log \left(\frac{d\pi_{2n}}{d\eta} \right) \leq W_U + W^{U,V} - \eta(V).$$

We end the proof of (5.54) taking the exponential and using (5.50). The last assertion follows from the fact that

$$e^{-\delta V} \left(\frac{d\pi_{2n}}{d\eta} \vee \frac{d\eta}{d\pi_{2n}} \right) \leq c \exp(-\mathcal{V}_\delta) \quad \text{and} \quad e^{-\delta U} \left(\frac{d\pi_{2n+1}}{d\mu} \vee \frac{d\mu}{d\pi_{2n+1}} \right) \leq c \exp(-\mathcal{U}_\delta).$$

This ends the proof of the proposition. $\blacksquare$

Lemma 5.22. *There exists $a > 0$ such that for any $n \geq 1$ we have*

$$\mathcal{K}_{2n}(h)/g \leq a e^{-\mathcal{U}_\delta} \quad \text{and} \quad \mathcal{K}_{2n+1}(g)/h \leq a e^{-\mathcal{V}_\delta}.$$

In addition, there exists some $\iota > 0$ such that for any $n \geq 1$ we have

$$\iota K_V(x, dy) \leq \mathcal{K}_{2n}(x, dy) \quad \text{and} \quad \iota K_U^\flat(y, dx) \leq \mathcal{K}_{2n+1}(y, dx).$$

Proof. Combining (4.10) with (5.50), (5.51) and (5.52) there exists some $c_0 > 0$ such that for any $n \geq 1$ we have

$$\frac{1}{c_0} K_V(x, dy) \leq \mathcal{K}_{2n}(x, dy) = \frac{K_V(x, dy) e^{(V-V_{2n})(y)}}{K(e^{-V_{2n}})(x)} \leq c_0 e^{W^V(x)} K_V(x, dy) e^{W_U(y)}.$$

This implies that

$$e^{-\delta U} \mathcal{K}_{2n}(e^{\delta V}) \leq c_0 e^{-\mathcal{U}_\delta} K_{(1-\delta)V}(e^{W_U}) \leq c_1 e^{-\mathcal{U}_\delta} \quad \text{with} \quad c_1 := c_0 \|\exp W_{(1-\delta)V,U}\|.$$

Then, recalling that $U_{2n+1} = U_{2n}$ we have that there exists some $c_0 > 0$ such that for any $n \geq 1$ we have

$$\frac{1}{c_0} K_U^\flat(y, dx) \leq \mathcal{K}_{2n+1}(y, dx) = \frac{K_U^\flat(y, dx) e^{(U-U_{2n})(x)}}{K^\flat(e^{-U_{2n}})(y)} \leq c_0 e^{W_U(y)} K_U^\flat(y, dx) e^{W^V(x)}.$$

This implies that

$$e^{-\delta V} \mathcal{K}_{2n+1}(e^{\delta U}) \leq c_0 e^{-\mathcal{V}_\delta} K_{(1-\delta)U}^\flat(e^{W^V}) \leq c_1 e^{-\mathcal{V}_\delta} \quad \text{with} \quad c_1 := c_0 \|\exp W^{(1-\delta)U,V}\|.$$

This ends the proof of the lemma. $\blacksquare$

5.6.4 A contraction theorem

Theorem 5.23. *Assume (A). Then there exists some $a > 0$ and $\varrho \in]0, 1[$ such that for any $n \geq 1$ we have*

$$\text{lip}_{g_a, h_a}(\mathcal{K}_{2n}) \vee \text{lip}_{h_a, g_a}(\mathcal{K}_{2n+1}) \leq \varrho \quad \text{with} \quad g_a := \frac{1}{2} + a g \quad \text{and} \quad h_a := \frac{1}{2} + a h \quad (5.56)$$

and (g, h) as in (5.53). In addition, we have the uniform estimates

$$\sup_{n \geq 1} (\pi_{2n}(h) \vee \pi_{2n+1}(g)) < \infty \quad \text{and} \quad \sup_{n \geq 1} (\|\mathcal{K}_{2n}(h)\|_g \vee \|\mathcal{K}_{2n+1}(g)\|_h) < \infty. \quad (5.57)$$

Proof. For any $\epsilon \in]0, 1[$, choosing $r > 0$ such that

$$a e^{-r} \leq \epsilon \quad \text{and set} \quad c := a \left(\sup_{\mathcal{U}_\delta(x) \leq r} \left(e^{-\mathcal{U}_\delta(x)} g(x) \right) \vee \sup_{\mathcal{V}_\delta(y) \leq r} \left(e^{-\mathcal{V}_\delta(y)} h(y) \right) \right).$$

In this notation, we have

$$\mathcal{K}_{2n}(h) \leq \epsilon g + c \quad \text{and} \quad \mathcal{K}_{2n+1}(g) \leq \epsilon h + c.$$

This implies that

$$\begin{aligned} \pi_{2n}(h) = \lambda_U \mathcal{K}_{2n}(h) &\leq \epsilon \lambda_{(1-\delta)U}(1) + c \\ \pi_{2n+1}(g) = \nu_V \mathcal{K}_{2n+1}(g) &\leq \epsilon \nu_{(1-\delta)V}(1) + c. \end{aligned} \quad (5.58)$$

We end the proof of (5.57) recalling that (U, V) are bounded below by a real number.

Since $\mathbb{X}$ and $\mathbb{Y}$ can be exhausted respectively by the compact sub-levels sets $\{g \leq l\}$ and $\{h \leq l\}$, there exists some $l_1 > 1$ such that $\lambda(\{g \leq l\}) \wedge \nu(\{h \leq l\}) > 0$ for $l = l_1$ and thus for any $l \geq l_1$. We set $j_l := \sup_{g(x) \vee h(y) \leq l} W(x, y)$. Then, for any $l \geq l_1$ and $g(x) \leq l$ we have

$$K_V(x, dy) \geq \iota_V(l) \frac{\nu_V(dy) 1_{\{h \leq l\}}(y)}{\nu_V(\{h \leq l\})} \quad \text{with} \quad \iota_V(l) := e^{-j_l} \nu_V(\{h \leq l\})$$

Note that

$$\nu_V(\{h \leq l\}) = \nu(e^{-V} 1_{V \leq \log(l)/\delta}) \geq e^{-\log(l)/\delta} \nu(\{h \leq l\}) > 0$$

In the same vein, for any $l \geq l_1$ and $h(y) \leq l$ we have

$$K_U^\flat(y, dx) \geq \iota_U(l) \frac{\lambda_U(dx) 1_{\{g \leq l\}}(x)}{\lambda_U(\{g \leq l\})} \quad \text{with} \quad \iota_U(l) := e^{-j_l} \lambda_U(\{g \leq l\}).$$

This implies that for any $g(x_1) \vee g(x_2) \leq l$ and $h(y_1) \vee h(y_2) \leq l$ we have

$$\|\delta_{x_1} \mathcal{K}_{2n} - \delta_{x_2} \mathcal{K}_{2n}\|_{tv} \vee \|\delta_{y_1} \mathcal{K}_{2n+1} - \delta_{y_2} \mathcal{K}_{2n+1}\|_{tv} \leq 1 - \iota(l)$$

with $\iota(l) := \iota(\iota_U(l) \wedge \iota_V(l))$. By Theorem 3.1, the contraction inequalities (5.56) are satisfied for some $a > 0$ and $\varrho \in]0, 1[$ that only depends on the parameters (ϵ, c) and the function $\iota(l)$ and hence we conclude. $\blacksquare$

The next corollary is a consequence of (5.56).

Corollary 5.24. *Under the assumptions of Theorem 5.23, we have the uniform contraction estimates*

$$\sup_{n \geq 1} (\text{lip}_{g_a, g_a}(\mathcal{S}_{2n+1}) \vee \text{lip}_{h_a, h_a}(\mathcal{S}_{2(n+1)})) \leq \varrho^2.$$

with $\varrho \in]0, 1[$ as in (5.56). In addition, for any $n \geq 1$ we have

$$\|\pi_{2(n+1)} - \eta\|_{h_a} \leq \varrho^2 \|\pi_{2n} - \eta\|_{h_a} \quad \text{and} \quad \|\pi_{2n+1} - \mu\|_{g_a} \leq \varrho^2 \|\pi_{2n-1} - \mu\|_{g_a}.$$

Combining Corollary 5.24 with the uniform estimates (5.58) there exists some constant $c > 0$ such that for any $n \geq 1$ we have

$$\|\pi_{2n+1} - \mu\|_{g_a} \vee \|\pi_{2(n+1)} - \eta\|_{h_a} \leq c \varrho^{2n}. \quad (5.59)$$

Following the proof of (3.22) using (5.59) we obtain the following Kantorovich semi-distance exponential decays.

Corollary 5.25. *Assume the assumptions of Theorem 5.23 are met and there exists semi-distance ψ_g and ψ_h on $\mathbb{X}$ and $\mathbb{Y}$ such that*

$$\psi_g(x_1, x_2) \leq w_g(x_1, x_2) \quad \text{and} \quad \psi_h(y_1, y_2) \leq w_h(y_1, y_2).$$

In this situation, for any $n \geq 1$ we have

$$D_{\psi_g}(\pi_{2n+1}, \mu) \vee D_{\psi_h}(\pi_{2(n+1)}, \eta) \leq (c/a) \varrho^{2n}$$

with $\varrho \in]0, 1[$ as in (5.56) and the Kantorovich semi-distances (D_{ψ_g}, D_{ψ_h}) defined in (3.6).

Example 5.26. *Consider the Beta marginal distribution λ_U on $\mathbb{X} :=]0, 1[$ discussed in Example 5.18. In this case, recalling that for any $p > 0$ and $u, v \geq 1$ we have*

$$2uv \geq u + v \quad \text{and} \quad |u - v|^p \leq (u + v)^p \leq 2^{(p-1)+} (u^p + v^p)$$

and one has

$$g(x) := e^{\delta U(x)} = \frac{1}{c^\delta} \left(\frac{1}{x}\right)^{\delta a} \left(\frac{1}{1-x}\right)^{\delta b} \geq \frac{1}{2c^\delta} \left(\left(\frac{1}{x}\right)^{p_a} + \left(\frac{1}{1-x}\right)^{p_b} \right)$$

with $(p_a, p_b) := (\delta a, \delta b)$. Therefore

$$w_g(x_1, x_2) \geq \psi_g(x_1, x_2) := c_{a,b} \left(\left| \frac{1}{x_1} - \frac{1}{x_2} \right|^{p_a} + \left| \frac{1}{1-x_1} - \frac{1}{1-x_2} \right|^{p_b} \right) \quad \text{for some } c_{a,b} > 0.$$

Example 5.27. *Consider the marginal distribution λ_U on $\mathbb{X} := \mathbb{R}^d$ discussed in Example 5.19 for some parameters (b_1, b_2, q) . In this case we have*

$$g(x) := e^{\delta U(x)} \geq c \exp(2b\|x\|^q) \quad \text{with } (c, b) := (e^{b_1}, b_2/2).$$

Following Example 3.2 several semi-distances ψ_g can be chosen, for instance we have

$$\begin{aligned} w_g(x_1, x_2) &\geq 2c (\exp(b(\|x_1\|^q + \|x_2\|^q)) - 1) \\ &= \psi_g(x_1, x_2) := 2c \left(\exp\left(\frac{b}{2^{(q-1)+}} \|x_1 - x_2\|^q\right) - 1 \right). \end{aligned}$$

5.6.5 Entropy estimates

Proposition 5.28. *Under the assumptions of Theorem 5.23, there exists some $c > 0$ such that for any $n \geq 1$ we have*

$$\left\| \frac{d\pi_{2(n+1)}}{d\eta} - 1 \right\|_h \vee \left\| \frac{d\eta}{d\pi_{2(n+1)}} - 1 \right\|_h \vee \left\| \log \frac{d\eta}{d\pi_{2(n+1)}} \right\|_h \leq c \varrho^{2(n+1)}$$

with $\varrho \in]0, 1[$ as in (5.56). The above estimates remain valid when we replace $(\pi_{2(n+1)}, \eta, h)$ by (π_{2n+1}, μ, g) .

Proof. Arguing as in the proof of Proposition 5.3 for any $k, l \geq 1$ we have

$$\left| \frac{d\pi_{2(l+k)}}{d\eta}(y) - 1 \right| \leq \|d\pi_{2l}/d\eta\|_{h_a} \left\| (\delta_y - \eta) \mathcal{S}_{2(l+k)} \cdots \mathcal{S}_{2(l+1)} \right\|_{h_a}.$$

Note that

$$\|f\|_{h_a} \leq (2\|f\|) \wedge (\|f\|_h/a).$$

By (5.55) there exists some constant $c > 0$ such that for any $k \geq 1$ we have

$$\left| \frac{d\pi_{2k}}{d\eta}(y) - 1 \right| \leq c \varrho^{2k} (1 + h(y)) \implies \left\| \frac{d\pi_{2k}}{d\eta} - 1 \right\|_h \leq c \varrho^{2k}.$$

In addition, for any $k, l \geq 1$ we have

$$\begin{aligned} \left| \frac{d\pi_{2(l+k)+1}}{d\mu}(x) - 1 \right| &\leq \|d\pi_{2l+1}/d\mu\|_{g_a} \left\| (\delta_x - \mu) \mathcal{S}_{2(l+k)+1} \cdots \mathcal{S}_{2(l+1)+1} \right\|_{g_a} \\ \implies \forall k \geq 1 \quad \left\| \frac{d\pi_{2k+1}}{d\mu} - 1 \right\|_g &\leq c \varrho^{2k}. \end{aligned}$$

By Proposition 4.7 we have

$$\begin{aligned} \frac{d\eta}{d\pi_{2(k+1)}} - 1 &= \mathcal{K}_{2(k+1)+1} \left(\frac{d\pi_{2k+1}}{d\mu} - 1 \right) \quad \text{and} \quad \frac{d\mu}{d\pi_{2k+1}} - 1 = \mathcal{K}_{2(k+1)} \left(\frac{d\pi_{2k}}{d\eta} - 1 \right) \\ \implies \left\| \frac{d\eta}{d\pi_{2(k+1)}} - 1 \right\|_h &\leq \left\| \frac{d\pi_{2k+1}}{d\mu} - 1 \right\|_g \|\mathcal{K}_{2(k+1)+1}(g)\|_h. \end{aligned}$$

Using (5.5) and (5.57) we conclude that for any $k \geq 2$ we have

$$\left\| \frac{d\pi_{2k}}{d\eta} - 1 \right\|_h \vee \left\| \frac{d\eta}{d\pi_{2k}} - 1 \right\|_h \vee \left\| \log \frac{d\eta}{d\pi_{2k}} \right\|_h \leq c \varrho^{2k}.$$

This ends the proof of the proposition. ■

Proposition 5.28 ensures the convergence of the series (5.6) in the weighted Banach spaces $\mathcal{B}_g(\mathbb{X})$ and $\mathcal{B}_h(\mathbb{Y})$. Combining Proposition 4.4 with (5.6) and using Proposition 4.6 we obtain the following theorem.

Theorem 5.29. *Under the assumptions of Theorem 5.23, there exists some constant $c > 0$ such that for any $n \geq 1$ we have*

$$\|\mathbb{U} - U_{2n}\|_g \vee \|\mathbb{V} - V_{2n}\|_h \vee \mathcal{H}(P_{\mu, \eta} \mid \mathcal{P}_{2n}) \leq c \varrho^{2n} \quad (5.60)$$

with $\varrho \in]0, 1[$ as in (5.56) and with (U_{2n}, V_{2n}) and $(\mathbb{U}, \mathbb{V})$ as in Proposition 4.4 and (5.6).

References

- [1] M. Agueh, G. Carlier. Barycenters in the Wasserstein space. *SIAM*, vol. 43, no. 2, pp. 904–924 (2011).
- [2] S. Aida, I. Shigekawa. Logarithmic Sobolev inequalities and spectral gaps: Perturbation theory. *J. Funct. Anal.*, vol. 126, no. 2, pp. 448–475 (1994).
- [3] O.D. Akyildiz, P. Del Moral, J. Miguez. Gaussian entropic optimal transport: Schrödinger bridges and the Sinkhorn algorithm. *ArXiv preprint arXiv:2412.18432* (2024). *Foundations of Data Science*. doi: 10.3934/fods.2026004 (2026).
- [4] O.D. Akyildiz, P. Del Moral, J. Miguez. On the contraction properties of Sinkhorn semigroups. *ArXiv preprint 2503.09887* (2025). *Foundations of Data Science*, to appear (2026).
- [5] J. Altschuler, J. Niles-Weed, P. Rigollet, Near-linear time approximation algorithms for optimal transport via Sinkhorn iteration, *Advances in Neural Information Processing Systems*, pp. 1964–1974 (2017).
- [6] B.D.O. Anderson. Reverse-time diffusion equation models. *Stochastic Processes and their Applications*. vol.12, no. 3. pp. 313–326 (1982).
- [7] B.D.O. Anderson, A.N. Bishop, Del Moral Moral, C. Palmier. Backward nonlinear smoothing diffusions. *Theory of Probability & Its Applications*, vol. 66, no. 2, pp. 245–262 (2021).
- [8] T. Ando and J. L. van Hemmen. An inequality for trace ideals. *Commun. Math. Phys.*, vol. 76, pp. 143–148 (1980).
- [9] M. Arnaudon, P. Del Moral, E. Ouhabaz. A Lyapunov approach to stability of positive semigroups: An overview with illustrations. *Stochastic Analysis and Applications*, vol. 42, no. 1, pp. 121–200 (2024).
- [10] D. G. Aronson, Bounds for the fundamental solution of a parabolic equation, *Bulletin of American Math. Soc.*, vol. 73, pp. 890–896 (1967).
- [11] A. Bain, D. Crisan. *Fundamentals of Stochastic Filtering*. Springer New York (2009).
- [12] D. Bakry B. Emery. Diffusions hypercontractives. *Séminaire de Probabilités*, vol. XIX 1983/84: Proceedings. Springer, pp. 177–206 (1985).
- [13] D. Bakry, I. Gentil, M. Ledoux. *Analysis and geometry of Markov diffusion operators*. Vol. 103. Cham, Springer (2014).
- [14] J.D. Benamou, G. Carlier, M. Cuturi, L. Nenna, G. Peyré. Iterative Bregman projections for regularized transportation problems. *SIAM Journal on Scientific Computing*, vol. 37, no. 2A, pp. 1111–A1138 (2015).
- [15] G. Birkhoff. Extensions of Jentzsch’s theorem. *Transactions of the American Mathematical Society*, vol. 85, no.1, pp. 219–227 (1957).

- [16] R. Bojilov, A. Galichon. Matching in closed-form: equilibrium, identification, and comparative statics. *Economic Theory*, vol. 61, no. 4, pp. 587–609 (2016).
- [17] J.M. Borwein, A. S. Lewis, R. Nussbaum. Entropy minimization, DAD problems, and doubly stochastic kernels. *Journal of Functional Analysis*, vol. 123, no. 2, pp. 264–307 (1994).
- [18] R. Brualdi, A. Seymour, V. Parter, H. Schneider. The diagonal equivalence of a non-negative matrix to a stochastic matrix. In: *Journal of Mathematical Analysis and Applications*, vol. 16, no. 1, pp. 31–50 (1966).
- [19] C. Bunne, Y.P. Hsieh, M. Cuturi, A. Krause. The Schrödinger bridge between Gaussian measures has a closed form. In *International Conference on Artificial Intelligence and Statistics*, pp. 5802–5833 (2023).
- [20] P. Cattiaux, A. Guillin. Semi log-concave Markov diffusions. *Séminaire de Probabilités*, vol. XLVI, pp. 231–292 (2014).
- [21] P. Cattiaux, A. Guillin. Functional inequalities for perturbed measures with applications to log-concave measures and to some Bayesian problems. *Bernoulli*, vol. 28, no. 4, pp. 2294–2321. (2022).
- [22] D. Chakrabarty, S. Khanna, Better and simpler error analysis of the Sinkhorn-Knopp algorithm for matrix scaling. *First Symposium on Simplicity in Algorithms (SOSA 2018)*, Raimund Seidel, ed., pp. 4:1–4:11 (2018).
- [23] S. Chen, S. Chewi, J. Li, Y. Li, A. Salim, A. Zhang. Sampling is as easy as learning the score: theory for diffusion models with minimal data assumptions. In *The Eleventh International Conference on Learning Representations* (2022).
- [24] T. Chen, G.H. Liu, E. Theodorou. Likelihood training of Schrödinger bridge using forward- backward SDEs theory. In *International Conference on Learning Representations* (2022).
- [25] Y. Chen. Modeling and control of collective dynamics: From Schrödinger bridges to optimal mass transport. *Doctoral dissertation, University of Minnesota* (2016).
- [26] Y. Chen, T. Georgiou, and M. Pavon. Entropic and displacement interpolation: a computational approach using the Hilbert metric. *SIAM J. Appl. Math.*, vol. 76, no. 6, pp. 2375–2396 (2016).
- [27] Y. Chen, T. Georgiou, and M. Pavon. Optimal transport in systems and control. *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 4, no. 1, pp. 89–113 (2021).
- [28] S. Chewi, A.A. Pooladian. An entropic generalization of caffarelli’s contraction theorem via covariance inequalities. *Comptes Rendus. Mathématique*, vol. 361, G9, pp. 1471–1482 (2023).

- [29] A. Chiarini, G. Conforti, G. Greco, L. Tamanini. A semiconcavity approach to stability of entropic plans and exponential convergence of Sinkhorn’s algorithm. ArXiv preprint arXiv:2412.09235 (2024).
- [30] L. Chizat, A. Delalande, T. Vaskevicius. Sharper exponential convergence rates for sinkhorn’s algorithm in continuous settings: L. chizat et al. Mathematical Programming, pages 1–50 (2025).
- [31] J.E. Cohen, Y. Iwasa, G. Rautu, G. M.B. Ruskai, E. Seneta, and G. Zbaganu. Relative entropy under mappings by stochastic matrices. Linear Algebra Appl., vol.179, pp. 211–235 (1993).
- [32] S.N. Cohen, E. Fausti. Hyperbolic contractivity and the Hilbert metric on probability measures. arXiv preprint arXiv:2309.02413 (2023).
- [33] G. Conforti. Weak semiconvexity estimates for Schrödinger potentials and logarithmic Sobolev inequality for schrödinger bridges. Probability Theory and Related Fields, vol. 189, no. 3, pp. 1045–1071 (2024).
- [34] G. Conforti, A. Durmus, G. Greco. Quantitative contraction rates for Sinkhorn algorithm: beyond bounded costs and compact marginals. arXiv preprint arXiv:2304.04451 (2023).
- [35] E. Cramer. Conditional iterative proportional fitting for Gaussian distributions. Journal of multivariate analysis, vol. 65, no.2, pp. 261–276 (1998).
- [36] E. Cramer, Probability measures with given marginals and conditionals: I-projections and conditional iterative proportional fitting. Statistics & Risk Modeling, vol. 18, no. 3, pp. 311–330 (2000).
- [37] E. Cramer, Wahrscheinlichkeitsmaße mit gegebenen marginalen und bedingten Verteilungen: I-Projektionen und Conditional Iterative Proportional Fitting, Ph.D. Thesis, Aachen University of Technology, Aachen (1997).
- [38] I. Csiszár. Eine informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität von Markoffschen Ketten. Magyar Tud. Akad. Mat. Kutató Int. Közl., vol. 8, pp. 85–108 (1963).
- [39] I. Csiszár. I-Divergence Geometry of Probability Distributions and Minimization Problems. Ann. Probab., vol. 3, no. 1, pp. 146–158 (1975).
- [40] M. Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In Advances in Neural Information Processing Systems, pp. 2292–2300 (2013).
- [41] V. De Bortoli, A. Galashov, J.S. Guntupalli, G. Zhou, K. Murphy, A. Gretton, A. Doucet. Distributional Diffusion Models with Scoring Rules. arXiv preprint arXiv:2502.02483 (2025).
- [42] V. de Bortoli, J. Thornton, J. Heng, A. Doucet. Diffusion Schrödinger bridge with applications to score-based generative modeling. Advances in Neural Information Processing Systems, vol. 34, pp. 17695–17709 (2021).

- [43] G. Deligiannidis, V. De Bortoli, A. Doucet. Quantitative uniform stability of the iterative proportional fitting procedure. *The Annals of Applied Probability*, vol. 34, no. 1A, pp. 501– 516 (2024).
- [44] E. del Barrio, J.M. Loubes. The statistical effect of entropic regularization in optimal transportation. *arXiv preprint arXiv:2006.05199* (2020).
- [45] P. Del Moral. *Feynman-Kac formulae*. Springer New York (2004).
- [46] P. Del Moral, A. Guionnet. On the stability of interacting processes with applications to filtering and genetic algorithms. *Annales de l’Institut Henri Poincaré Probab. & Statist.* vol. 37, no. 2, pp. 155–194 (2001).
- [47] P. Del Moral, E. Horton. A note on Riccati matrix difference equations. *SIAM Journal on Control and Optimization*, vol. 60, no. 3, pp. 1393–1409 (2022).
- [48] P. Del Moral, E. Horton, A. Jasra. On the stability of positive semigroups. *The Annals of Applied Probability*, vol. 33, no. 6A, pp. 4424–4490 (2023).
- [49] P. Del Moral. A Contraction Theory for Sinkhorn and Schrödinger Bridges via Log-Sobolev Inequalities. *ArXiv preprint arXiv:2503.15963* (2025). *Stochastic Analysis and Applications to appear* (2026).
- [50] P. Del Moral. Entropic continuity bounds for conditional covariances with applications to Schrödinger and sinkhorn bridges. *arXiv preprint arXiv:2504.18822*; *Communications in Information and Systems (CIS) special issue dedicated to Prof. Thomas Kailath.* (2025).
- [51] P. Del Moral, M. Gerber. On the Kantorovich contraction of Markov semigroups. *ArXiv preprint arXiv:2511.08111* (2025).
- [52] P. Del Moral, A. Guionnet. On the stability of interacting processes with applications to filtering and genetic algorithms. *Annales de l’Institut Henri Poincaré Probab. & Statist.*, vol.37, no. 2, pp. 155–194 (2001).
- [53] P. Del Moral, M. Ledoux, and L. Miclo. On contraction properties of Markov kernels. *Probab. Theory Relat. Fields*, vol. 126, pp. 395–420 (2003).
- [54] P. Del Moral and S. Penev. *Stochastic Processes: From Applications to Theory*. CRC-Press (2016).
- [55] P. Del Moral and J. Tugaut. On the stability and the uniform propagation of chaos properties of ensemble Kalman–Bucy filters. *The Annals of Applied Probability*, vol. 28, no. 2, pp. 790–850, (2018).
- [56] W.E. Deming and F.F. Stephan. On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. *Ann. Math. Statistics*, vol. 11, pp. 427–444 (1940).
- [57] S. Di Marino, A. Gerolin. An optimal transport approach for the Schrödinger bridge problem and convergence of Sinkhorn algorithm. *Journal of Scientific Computing* vol. 85, no. 2 : 27 (2020).

- [58] R.L. Dobrushin. Central limit theorem for non stationary Markov chains, i,ii. *Theory Probab. Appl.*, vol. 1, no. 4, pp. 66–80 and 330–385 (1956).
- [59] R.L. Dobrushin. Perturbation methods of the theory of Gibbsian fields. In *Lectures on Probability Theory and Statistics: École d’Été de Probabilités de Saint-Flour XXIV -1994*. *Lecture Notes in Mathematics* 1648, pp. 1–66. Berlin: Springer (1996).
- [60] R. Douc, E. Moulines, J.S. Rosenthal. Quantitative bounds on convergence of time-inhomogeneous Markov chains. *Ann. Appl. Probab.*, vol. 14, no. 4, pp. 1643–1665 (2004).
- [61] P. Dvurechensky, A. Gasnikov, A. Kroshnin, Computational optimal transport: Complexity by accelerated gradient descent is better than by Sinkhorn’s algorithm, *Proceedings of the 35th International Conference on Machine Learning*, Jennifer Dy and Andreas Krause, eds. *Proceedings of Machine Learning Research*, vol. 80, PMLR, 10–15, pp. 1367–1376 (2018).
- [62] A. Eberle. Reflection couplings and contraction rates for diffusions. *Probability theory and related fields*, vol. 166, no. 3, pp. 851–886 (2016).
- [63] S. Eckstein. Hilbert’s projective metric for functions of bounded growth and exponential convergence of sinkhorn’s algorithm. *Probability Theory and Related Fields*, pages 1–37, (2025).
- [64] M. Essid and M. Pavon. Traversing the Schrödinger bridge strait: Robert Fortet’s marvelous proof redux. *Journal of Optimization Theory and Applications*, vol. 181, no. 1, pp. 23–60 (2019).
- [65] W.A. Farahat, H. Asada. Estimation of state transition probabilities in asynchronous vector markov processes. *Journal of dynamic systems, measurement, and control*, vol. 134.6, no. 061003 (2012).
- [66] S.E. Fienberg. An Iterative Procedure for Estimation in Contingency Tables. *Ann. Math. Statist.*, vol. 41, no. 3, pp. 907–917 (1970).
- [67] H. Föllmer. Random fields and diffusion processes. In *Lecture Notes in Mathematics*, volume 136, p. 101–204 (1988).
- [68] R. Fortet. Résolution d’un système d’équations de Schrödinger. *J. Math. Pures Appl.*, vol. 19, pp. 83–105 (1940).
- [69] J. Franklin, J. Lorenz. On the scaling of multidimensional matrices. *Linear Algebra and its Applications*, vol. 114–115, pp. 717–735 (1989).
- [70] A. Garg, R. Oliveira. Recent progress on scaling algorithms and applications. *ArXiv preprint arXiv:1808.09669* (2018).
- [71] A. Genevay, M. Cuturi, G. Peyré. Learning generative models with Sinkhorn divergences. In *International Conference on Artificial Intelligence and Statistics*, pp. 1608–1617 (2018).

- [72] A. Genevay, L. Chizat, F. Bach, M. Cuturi, G. Peyré. Sample complexity of Sinkhorn divergences. In The 22nd international conference on artificial intelligence and statistics, pp. 1574–1583 (2019).
- [73] P. Ghosal, M. Nutz. On the Convergence Rate of Sinkhorn’s Algorithm. ArXiv preprint arXiv:2212.06000 (2022).
- [74] G. Greco, L. Tamanini. Hessian stability and convergence rates for entropic and Sinkhorn potentials via semiconcavity. arXiv:2504.11133 (2025).
- [75] G. Greco, M. Noble, G. Conforti, A. Durmus. Non asymptotic convergence bounds for sinkhorn iterates and their gradients: a coupling approach. In The Thirty Sixth Annual Conference on Learning Theory, pages 716–746. PMLR (2023).
- [76] M. Hairer and J. C. Mattingly. Yet another look at Harris’ ergodic theorem for Markov chains. Seminar on Stochastic Analysis, Random Fields and Applications VI, 109–117, Progr. Probab., 63, Birkhäuser/Springer Basel AG, Basel (2011).
- [77] R. Z. Has’minskii. Stochastic stability of differential equations, vol. 7 of Monographs and Textbooks on Mechanics of Solids and Fluids: Mechanics and Analysis. Sijthoff & Noordhoff, Alphen aan den Rijn. Translated from the Russian by D. Louvish (1980).
- [78] H.G. Haussmann and E. Pardoux. Time reversal of diffusions. The Annals of Probability. vol. 14, no. 4. pp. 1188–205 (1986).
- [79] V. Huynh, D. Phung. Optimal transport for deep generative models: State of the art and research challenges. In International Joint Conference on Artificial Intelligence 2021 (pp. 4450–4457). Association for the Advancement of Artificial Intelligence (2021).
- [80] M. Idel. A review of matrix scaling and Sinkhorn’s normal form for matrices and positive maps. ArXiv preprint arXiv:1609.06349 (2016).
- [81] N. J. Higham. Functions of Matrices : Theory and Computation, SIAM, Philadelphia, PA (2008).
- [82] R. Holley, D. Stroock. Logarithmic Sobolev inequalities and stochastic Ising models. J. Statist. Phys., vol. 46, no. 5-6, pp. 1159–1194 (1987).
- [83] H. Janati, B. Muzellec, G. Peyré, M. Cuturi Entropic optimal transport between unbalanced gaussian measures has a closed form. Advances in neural information processing systems, vol. 33, pp. 10468–10479 (2020).
- [84] M.R. Karimi, Y.P. Hsieh, A. Krause. Sinkhorn Flow as Mirror Flow: A Continuous-Time Framework for Generalizing the Sinkhorn Algorithm. International Conference on Artificial Intelligence and Statistics (2024).
- [85] J. Kruithof. Telefoonverkeersrekening (calculation of telephone traffic). De Ingenieur, vol. 52, no. 8, E15–E25 (1937).
- [86] F. Léger. A gradient descent perspective on Sinkhorn. Appl. Math. Optim., vol. 84, no. 2, pp. 1843–1855 (2021).

- [87] C. Léonard. A survey of the Schrödinger problem and some of its connections with optimal transport. *Discrete Contin. Dyn. Syst.*, vol. 34, no. 4, pp. 1533–1574 (2014).
- [88] Y.A Ma, Y. Chen, C. Jin, N. Flammarion, M.I. Jordan. Sampling can be faster than optimization. *Proceedings of the National Academy of Sciences*, vol. 116, no. 42, pp. 20881–20885 (2019).
- [89] V. Majernik. Entropy-A universal concept in sciences. *Natural Science*, vol. 6, pp. 552–564 (2014).
- [90] A. Mallasto, A. Gerolin, H.Q. Minh. Entropy-regularized 2-Wasserstein distance between Gaussian measures. *Information Geometry*, vol. 5, no. 1 pp. 289–323 (2022).
- [91] S. P. Meyn, R. L. Tweedie. Stability of Markovian processes. I. Criteria for discrete-time chains. *Adv. in Appl. Probab.*, vol. 24, no. 3, pp. 542–574 (1992).
- [92] S. P. Meyn, R. L. Tweedie. Markov chains and stochastic stability. *Communications and Control Engineering Series*. Springer-Verlag London Ltd., London (1993).
- [93] P. Monmarché. Wasserstein contraction and Poincaré inequalities for elliptic diffusions with high diffusivity. *Annales Henri Lebesgue*, vol. 6, pp. 941–973 (2023).
- [94] P. Monmarché, W. Songbo. Logarithmic Sobolev inequalities for non-equilibrium steady states. *ArXiv:2309.07858* (2023).
- [95] J. Nash, Continuity of solutions of parabolic and elliptic equations, *American J. Math.* vol. 80, pp 931–954 (1958).
- [96] F. Nielsen. On a variational definition for the Jensen-Shannon symmetrization of distances based on the information radius. *Entropy*, vol. 23, no. 4 (2021).
- [97] M. Nutz. Introduction to entropic optimal transport. Lecture notes, Columbia University (2021).
- [98] F. Otto, M. G. Reznikoff. A new criterion for the logarithmic Sobolev inequality and two applications. *Journal of Functional Analysis*, vol. 243, no. 1, pp. 121–157 (2007).
- [99] F. Otto, C. Villani. Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality. *Journal of Functional Analysis*, vol. 173, no. 2, pp. 361–400 (2000).
- [100] A.A. Pooladian, J. Niles-Weed. Entropic estimation of optimal transport maps. *ArXiv preprint arXiv:2109.12004* (2021).
- [101] H. J. M. Peters, P. P. Wakker. Convex functions on non-convex domains. *Econ Lett*, vol. 22, no. 2, pp. 251–255 (1986).
- [102] G. Peyré, M. Cuturi. Computational Optimal Transport. *Foundations and Trends in Machine Learning*, vol. 11, no. 5-6, pp. 355–607 (2019).

- [103] L. Saloff-Coste, J. Zuniga. Merging for time inhomogeneous finite Markov chains, Part I: Singular values and stability. *Electronic Journal of Probability*, vol. 14, no. 49, 1456–1494 (2009).
- [104] L. Saloff-Coste, J. Zuniga. Merging for time inhomogeneous finite Markov chains, Part II: Nash and Log-Sobolev inequalities. *Annals of Probability*, vol. 39, no. 3, pp. 1161–1203 (2011).
- [105] L. Saloff-Coste, J. Zuniga. Merging and stability for time inhomogeneous finite Markov chains. In *Surveys in stochastic processes 2011 Aug 10* (pp. 127–151). European Mathematical Society-EMS-Publishing House GmbH (2011).
- [106] L. Saloff-Coste, J. Zuniga. Convergence of some time inhomogeneous Markov chains via spectral techniques. *Stochastic processes and their applications*, vol. 117, no. 8, pp. 961–979 (2007).
- [107] T. Séjourné, G. Peyré, F.X. Vialard. Unbalanced optimal transport, from theory to numerics. *Handbook of Numerical Analysis*, vol. 24, pp. 407–471 (2023).
- [108] R. Schoen, S. H. Jonsson. Estimating multistate transition rates from population distributions. *Demographic Research*, vol. 9, pp. 1–24 (2003).
- [109] R. Schoen. Multistate transfer rate estimation from adjacent populations. *Population Research and Policy Review*, vol. 35, no. 2, pp. 217–240 (2016).
- [110] R. Sinkhorn. Diagonal equivalence to matrices with prescribed row and column sums. *American Mathematical Monthly*, vol. 74, pp. 402–405 (1967).
- [111] R. Sinkhorn. A Relationship Between Arbitrary Positive Matrices and Doubly Stochastic Matrices. *The Annals of Mathematical Statistics*, vol. 35, no. 2, pp. 876–879 (1964).
- [112] R. Sinkhorn, P. Knopp. Concerning nonnegative matrices and doubly stochastic matrices. *Pacific Journal of Mathematics*, vol. 21, no. 2, pp. 343–348 (1967).
- [113] J. Sohl-Dickstein, E. Weiss, N. Maheswaranathan, and S. Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *International Conference on Machine Learning*, pp. 2256–2265 (2015).
- [114] Y. Song, D. P. Kingma. How to train your energy-based models. *ArXiv preprint arXiv:2101.03288* (2021).
- [115] Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, B. Poole. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations* (2021).
- [116] G.W. Soules. The rate of convergence of Sinkhorn balancing. *Linear Algebra and its Applications*, vol. 150, pp. 3–40 (1991).
- [117] N. Th. Varopoulos. Small time gaussian estimates of heat diffusion kernels. II. The theory of large deviations. *Journal of Functional Analysis*, vol. 93, pp. 1–33 (1990).

- [118] C. Villani. Topics in Optimal Transportation, volume 58. AMS Graduate Studies in Mathematics (2003).
- [119] C. Villani. Optimal transport: old and new. Vol. 338. Springer (2009).
- [120] P. Vincent. A connection between score matching and denoising autoencoders. Neural Computation, vol. 23, no. 7, pp. 1661–1674 (2011).
- [121] T. Wang. Sinkhorn Divergence, Wasserstein Distance, and Their Roles in Generative Modeling: A Comprehensive Review of Optimal Transport and f-Divergences in Machine Learning. Authorea Preprints (2025).
- [122] M. Yan. Extension of convex function. J Convex Anal, vol. 21, no. 4, pp. 965–987 (2014).
- [123] J. Zhang, W. Zhong, P. Ma. A review on modern computational optimal transport methods with applications in biomedical research. Modern Statistical Methods for Health Research, pp. 279–300 (2021).