Journal of Functional Analysis 283 (2022) 109622

Contents lists available at ScienceDirect

Journal of Functional Analysis

www.elsevier.com/locate/jfa

Full Length Article

Stability of entropic optimal transport and Schrödinger bridges ☆

Promit Ghosal [a], Marcel Nutz [b,*,1], Espen Bernton [c]

[a] *Department of Mathematics, Massachusetts Institute of Technology, United States of America*
[b] *Departments of Statistics and Mathematics, Columbia University, United States of America*
[c] *Department of Statistics, Columbia University, United States of America*

ARTICLE INFO

Article history:
Received 17 June 2021
Accepted 29 June 2022
Available online 11 July 2022
Communicated by K.-T. Sturm

MSC:
90C25
49N05

Keywords:
Entropic optimal transport
Schrödinger bridge
Stability
Sinkhorn's algorithm

ABSTRACT

We establish the stability of solutions to the entropically regularized optimal transport problem with respect to the marginals and the cost function. The result is based on the geometric notion of cyclical invariance and inspired by the use of c-cyclical monotonicity in classical optimal transport. As a consequence of stability, we obtain the wellposedness of the solution in this geometric sense, even when all transports have infinite cost. More generally, our results apply to a class of static Schrödinger bridge problems including entropic optimal transport.

☆ The authors are grateful to Julio Backhoff-Veraguas and Giovanni Conforti for insightful discussions that greatly contributed to this research.

* Corresponding author.
E-mail addresses: promit@mit.edu (P. Ghosal), mnutz@columbia.edu (M. Nutz), eb3311@columbia.edu (E. Bernton).

[1] Research supported by an Alfred P. Sloan Fellowship (FG20166282) and NSF Grants DMS-2106056, DMS-1812661.

https://doi.org/10.1016/j.jfa.2022.109622
0022-1236/

1. Introduction

Computational progress has led to manifold applications of optimal transport in high-dimensional problems ranging from machine learning and statistics to image and language processing (e.g., [2,15,48,1]). In this context, entropic regularization is crucial to enable efficient large-scale computation via Sinkhorn's algorithm, hence has become the focus of dozens of recent studies. We refer to [44] for a survey with numerous references.

Our main contribution is the stability of solutions to the entropically regularized optimal transport problem with respect to the marginals and the cost function. Parallel to the fundamental stability theorem in classical optimal transport, it justifies, for example, that approximations found by solving discretized problems indeed converge to the true solution when the cost function is continuous. Our results are stated in terms of cyclical invariance, a geometric notion inspired by the c-cyclical monotonicity property in classical optimal transport. When the entropic transport problem has finite value, a coupling is cyclically invariant if and only if it is an optimal transport. Our stability theorem entails a general wellposedness result beyond the realm of optimization: cyclical invariance singles out a unique coupling even if the transport problem has infinite value—i.e., all couplings have infinite cost—and therefore the paradigm of cost minimization does not differentiate couplings from one another.

For ease of exposition, the Introduction focuses on entropic optimal transport. More general results are stated in Section 2 using the language of Schrödinger bridges that turns out to be natural for our approach. Given a measurable cost function $c : \mathsf{X} \times \mathsf{Y} \to \mathbb{R}_+$ on Polish probability spaces (X, μ) and (Y, ν), we consider the entropic optimal transport problem with regularization parameter $\varepsilon > 0$,

$$\inf_{\pi \in \Pi(\mu,\nu)} \int_{\mathsf{X}\times\mathsf{Y}} c \, d\pi + \varepsilon H(\pi|P), \quad P := \mu \otimes \nu, \tag{1.1}$$

where $\Pi(\mu, \nu)$ is the set of couplings and $H(\cdot|P)$ denotes relative entropy (or Kullback–Leibler divergence) with respect to the **p**roduct P of the marginals, defined as $H(\pi|P) := \int \log(\frac{d\pi}{dP}) \, d\pi$ for $\pi \ll P$ and $H(\pi|P) := \infty$ otherwise. If the minimization (1.1) is finite; i.e., if

$$\text{there exists } \pi_0 \in \Pi(\mu, \nu) \text{ with } \int c \, d\pi_0 + H(\pi_0|P) < \infty, \tag{1.2}$$

then it admits a unique minimizer $\pi \in \Pi(\mu, \nu)$ and moreover $\pi \sim P$.

Definition 1.1. A coupling $\pi \in \Pi(\mu, \nu)$ is called (c, ε)-*cyclically invariant* if $\pi \sim P$ and its density admits a version $d\pi/dP : \mathsf{X} \times \mathsf{Y} \to (0, \infty)$ such that

$$\prod_{i=1}^{k} \frac{d\pi}{dP}(x_i, y_i) = \exp\left(-\frac{1}{\varepsilon} \left[\sum_{i=1}^{k} c(x_i, y_i) - \sum_{i=1}^{k} c(x_i, y_{i+1}) \right] \right) \prod_{i=1}^{k} \frac{d\pi}{dP}(x_i, y_{i+1})$$

for all $k \in \mathbb{N}$ and $(x_i, y_i)_{i=1}^k \subset \mathsf{X} \times \mathsf{Y}$, where $y_{k+1} := y_1$.

By way of a factorization property that is equivalent to cyclical invariance, known results imply the following relation to the optimization (1.1).

Proposition 1.2. *Let (1.1) be finite. Then $\pi \in \Pi(\mu, \nu)$ is the minimizer of (1.1) if and only if π is (c, ε)-cyclically invariant.*

See Section 5 for details and references. We are mainly interested in optimal transport problems on Euclidean spaces X, Y. However, the only particular property of such spaces that plays a role for our analysis is that Lebesgue's theorem on the differentiation of measures holds. Thus, we postulate that property (see Assumption 2.3) and otherwise allow for a general Polish setting. We can now state the aforementioned wellposedness result.

Theorem 1.3 *(Wellposedness). Let $c : \mathsf{X} \times \mathsf{Y} \to [0, \infty)$ be continuous, $\varepsilon > 0$ and $(\mu, \nu) \in \mathcal{P}(\mathsf{X}) \times \mathcal{P}(\mathsf{Y})$. There exists a unique (c, ε)-cyclically invariant coupling $\pi \in \Pi(\mu, \nu)$. If (1.1) is finite, π is its unique minimizer.*

Uniqueness follows from known facts and does not require the continuity of c. On the other hand, existence beyond the framework of finite cost is a completely novel result. Rather than using convex analysis or variational arguments, it is based on the subsequent stability theorem for cyclical invariance. One example where wellposedness with infinite cost is of interest, is the statistical notion of rank recently proposed in [18]. Multivariate ranks have been defined in nonparametric statistics through Brenier's optimal transport map to extend the usual scalar notions and tests; see [15,19,20,28]. Leveraging the same idea but computationally less expensive, entropic optimal transport is used in [18] to define "differentiable ranks." Theorem 1.3 allows one to naturally define such ranks for arbitrary distributions—like in the scalar case—without imposing a second moment condition.

Theorem 1.4 *(Stability). For $n \geq 1$, let $(\mu_n, \nu_n) \in \mathcal{P}(\mathsf{X}) \times \mathcal{P}(\mathsf{Y})$, let $\varepsilon_n > 0$ and let $c_n : \mathsf{X} \times \mathsf{Y} \to [0, \infty)$ be measurable. Let $\pi_n \in \Pi(\mu_n, \nu_n)$ be (c_n, ε_n)-cyclically invariant. Suppose that μ_n, ν_n converge weakly to some limits μ, ν, that $\varepsilon_n \to \varepsilon > 0$ and that c_n converges uniformly on bounded sets to a continuous function c. Then π_n converges weakly to a limit $\pi \in \Pi(\mu, \nu)$ and π is (c, ε)-cyclically invariant.*

If the involved optimization problems are finite, the theorem states the stability of the (entropic) optimal transport couplings. A simple yet important application is when the marginals μ, ν are approximated by discrete measures, as it would be in a computational implementation. Even for this particular case, we are not aware of similar results in the literature. We mention that continuity for the limiting cost c in Theorem 1.4 is a sharp

condition: Proposition 5.3 will demonstrate that any nontrivial discontinuity in c leads to a failure of Theorem 1.4, for suitable marginals.

A noteworthy application of the stability theorem was presented in the follow-up work [40] where it was observed that convergence of Sinkhorn's algorithm can be seen as a stability problem. In that context, the marginals μ_n, ν_n are produced by the algorithm and known to converge to μ, ν in great generality, while the iterates of the algorithm correspond to π_n. Theorem 1.4 then implies the weak convergence to the correct optimizer π. Its geometric approach completely avoids the difficulty of establishing the integrability properties of μ_n, ν_n or even the finiteness of the associated entropic optimal transport problems.

The general existence result of Theorem 1.3 is a consequence of Theorem 1.4 applied with $c_n = c$, $\varepsilon_n = \varepsilon$ and approximations $(\mu_n, \nu_n) \to (\mu, \nu)$ where μ_n, ν_n are discrete measures with finite support. To solve the problem with marginals (μ_n, ν_n), one could use Proposition 1.2, but this particular case is a finite-dimensional minimization problem that can also be solved by standard calculus arguments. In particular, Theorem 1.4 yields an approach to construct cyclically invariant couplings which is novel even when the optimization problem is finite. This approach does not use the (classical but nontrivial) arguments of convex analysis and density factorization behind Proposition 1.2 (see [12,13,17,25,49,50], among others). It is also quite different from the iterative method of [26] which uses another finiteness condition; see [35] for a modern presentation, analysis and extension of that method. Instead, our approach is close in spirit to the construction of c-cyclically monotone couplings that is standard in classical optimal transport; cf. [52, pp. 64–65].

The analogy with classical optimal transport extends in several directions. McCann showed in [36] that cyclical monotonicity singles out a particular transport map for quadratic cost c on $\mathbb{R}^d$ even if the optimal transport problem (here the 2-Wasserstein distance) is infinite, thus extending Brenier's map to this setting; see also [52, pp. 249–258] for more general results. Here, the analogy becomes precise in the limit $\varepsilon \to 0$: the extended Brenier coupling is the weak limit of the couplings $\pi = \pi_\varepsilon$ established in Theorem 1.3 (this follows from [10]). Another important parallel occurs at the technical level. Working with cyclical invariance, we proceed in a local fashion and focus on finitely many points (x_i, y_i) at a time, rather than working with global objects like the Schrödinger potentials and their function spaces. For instance, non-compact marginal supports do not cause any particular difficulty in this approach. We emphasize that the novelty of the present study lies in how cyclical invariance is exploited and proved for the limit; the invariance property itself is merely an equivalent way of stating a factorization property of the density that is well known (see [12] or Proposition 4.4 below). It turns out that, once the line of argument is found, remarkably general results can be obtained with fairly concise proofs. We may hope that the techniques developed here can yield further insights into asymptotic questions on entropic optimal transport, and thus view the technique itself as a central contribution. Such questions may include quantifying the

speed of convergence in our stability theorem or finding analogous results for dynamic Schrödinger bridge problems.

The companion paper [10] illustrates the use of cyclical invariance for the limit $\varepsilon \to 0$. In this degenerate asymptotic, the limiting object is classical optimal transport as characterized by c-cyclical monotonicity. The latter property describes the shape of the support of the coupling and, therefore, is readily amenable to weak convergence arguments via Portmanteau's theorem. The same fact is often exploited in classical optimal transport theory, for instance in the standard proof of its stability theorem [52, p. 77]. In the present study, the limit is entropic optimal transport. Being a property of the density, the relation of cyclical invariance with weak convergence is less direct (especially as the measures in Theorem 1.4 may well be mutually singular). Our general principle is to blow the points (x_i, y_i) in Definition 1.1 up to small balls, pass to the weak limit, and then recover information about the limiting density by shrinking the balls, via differentiation of measures. This technique appears to be novel in this area.

Starting with [38,39], a number of works examine the degenerate asymptotic $\varepsilon \to 0$ where the limiting problem is classical optimal transport. The fact that weak limits of entropic optimizers are optimal transports was established by [33] using Gamma-convergence arguments in a more general context of Schrödinger bridges; see also [14] for the case of optimal transport with quadratic cost. As mentioned above, [10] extends this result to transport problems with infinite value by way of cyclical invariance and c-cyclical monotonicity; moreover, a large deviations principle quantifies the local rate of convergence. Related results can be found in [16,42] where the expansion of the optimal cost as a function of ε is studied. We remark that Gamma convergence seems difficult to use in the context of Theorem 1.4 due to the reference measures changing along the sequence. The limit $\varepsilon \to 0$ can also be analyzed in the associated dual problem, here the solutions are called potentials. Convergence of potentials was shown in [29] for quadratic costs and compactly supported marginals, and recently in [41] for a general Polish setting. Closer to the problem occurring in computational practice as well as the present question of stability, [9] studies the convergence of the discrete Sinkhorn algorithm to an optimal transport in the joint limit when $\varepsilon_n \to 0$ and the marginals μ, ν are approximated by discretizations μ_n, ν_n satisfying a certain density property. Explicit error bounds are derived, for instance for quadratic cost on the torus, to establish near-linear complexity of the resulting algorithm. For more on the computational challenges and remedies in this regime, see for instance [51] and the references therein.

While we are not aware of general stability results for the nondegenerate limit $\varepsilon_n \to \varepsilon > 0$ in the literature, the sampling complexity of entropic optimal transport (with fixed ε) can be seen as a particular form of stability with respect to the marginals. Indeed, [27,37] study how the empirical entropic Wasserstein distance, obtained by optimally coupling i.i.d. samples from the marginals, converges to the population version. The results are based on global arguments exploiting the regularity of the Schrödinger potentials, which, in turn, is achieved by imposing compactness and decay conditions on the marginals. See also [30] which studies a related asymptotic regime for a differ-

ent regularization of optimal transport. Related to the present work at least in spirit, there are several areas where analogues of c-cyclical monotonicity have recently lead to breakthroughs, including martingale optimal transport [7], optimal Skorokhod embeddings [5,8] and weak transport [4].

The remainder of this paper is organized as follows. Section 2 details the setting and main results in the language of Schrödinger bridges. The first step towards the stability theorem is reported Section 3 where we establish that weak limits of cyclically invariant couplings remain absolutely continuous. This is based on comparing measures of rectangles, an analysis that may be of independent interest. Section 4 continuous the main proof by showing that limits of cyclically invariant couplings are again cyclically invariant. It comprises of two steps; the aforementioned principle of blowing up points and passing to the limit first yields a weakened version of the invariance property, and then measure-theoretic arguments can be used to show that the (proper) invariance property already follows. The concluding Section 5 collects the arguments to prove the main results and their ramifications, including that continuity of the cost function is necessary for stability.

2. Main results

Let (X, d) be a complete, separable metric space; we write $\mathcal{P}(\mathsf{X})$ for the space of probability measures on the Borel σ-field $\mathcal{B}(\mathsf{X})$ endowed with weak convergence (induced by bounded continuous functions). The same is assumed for the second marginal space (Y, d), and we equip $\mathsf{X} \times \mathsf{Y}$ with the metric $d((x,y),(x',y')) = \max\{d(x,x'), d(y,y')\}$. Throughout this section, two measures $(\mu, \nu) \in \mathcal{P}(\mathsf{X}) \times \mathcal{P}(\mathsf{Y})$ play the role of given marginals for the static Schrödinger bridge problem

$$\inf_{\pi \in \Pi(\mu,\nu)} H(\pi|R) \tag{2.1}$$

where $R \in \mathcal{P}(\mathsf{X}\times\mathsf{Y})$ is a given **r**eference measure. We refer to [24,34] for extensive surveys on Schrödinger bridges. The entropic optimal transport problem (1.1) can be recovered (up to constants) as a special case for R defined by

$$\frac{dR}{dP} = ae^{-c/\varepsilon}, \quad P := \mu \otimes \nu \tag{2.2}$$

where $a = (\int e^{-c/\varepsilon}\, dP)^{-1}$ is the normalizing constant. In particular, $R \sim P$, which will also be an important condition in many of our results for (2.1). By way of (2.2), the following generalizes Definition 1.1.

Definition 2.1. Let $\pi \in \Pi(\mu,\nu)$ and $R \in \mathcal{P}(\mathsf{X} \times \mathsf{Y})$. We call (π, R) *cyclically invariant* if $\pi \sim R \sim P$ and there exists a version $d\pi/dR : \mathsf{X} \times \mathsf{Y} \to (0,\infty)$ of the relative density satisfying

$$\prod_{i=1}^{k} \frac{d\pi}{dR}(x_i, y_i) = \prod_{i=1}^{k} \frac{d\pi}{dR}(x_i, y_{i+1})$$

for all $k \in \mathbb{N}$ and $(x_i, y_i)_{i=1}^k \subset \mathsf{X} \times \mathsf{Y}$, where $y_{k+1} := y_1$.

The analogue of the finiteness condition (1.2) is that

$$\text{there exists } \pi_0 \in \Pi(\mu, \nu) \text{ with } H(\pi_0|R) < \infty.$$

We summarize the pertinent facts; see Lemma 5.1 for detailed references.

Proposition 2.2. *Let $R \sim P$ and let (2.1) be finite. There exists a unique minimizer $\pi \in \Pi(\mu, \nu)$ for (2.1), it satisfies $\pi \sim R$, and it is the unique coupling π such that (π, R) is cyclically invariant.*

In the remainder of the paper we assume that the underlying spaces allow for differentiation of measures in the following sense.

Assumption 2.3. Given $\rho, \lambda \in \mathcal{P}(\mathsf{X})$ satisfying $\rho \ll \lambda$, there exists $\mathsf{X}' \subset \mathsf{X}$ of full λ-measure such that

$$f(x) := \lim_{r \to 0} \frac{\rho(B_r(x))}{\lambda(B_r(x))}, \quad x \in \mathsf{X}' \tag{2.3}$$

defines a version of the Radon–Nikodym density $d\rho/d\lambda$. The analogous property is assumed on the space $\mathsf{X} \times \mathsf{Y}$.

For Euclidean spaces X, Y, the assumption holds by the standard differentiation theorem [21, Theorem 1.32, p. 53]. More generally, it holds in the context of so-called Vitali covering relations; the classical reference is [23, Theorem 2.9.8, p. 156]. For example, Assumption 2.3 holds when X and $\mathsf{X} \times \mathsf{Y}$ are compact subsets of Riemannian manifolds (due to the "directionally limited" property established in [23, Section 2.8.9, pp. 145-146]), or more generally, countable unions of such sets. See also [31, pp. 4-8, esp. Example 1.15] for an accessible introduction. For our purposes, the main restriction is that differentiation of measures generally fails on infinite-dimensional spaces; see [47] for a counterexample and [46] for a related result on coverings. An alternative to Assumption 2.3, making our results slightly more general, is to impose a doubling condition on the specific marginal measures (μ, ν); cf. Remark 5.2.

We have seen in the Introduction that continuity of c is essential for the stability of entropic optimal transport. In view of (2.2), it is then clear that the regularity of dR/dP is pivotal. The following generalizations of Theorems 1.3 and 1.4 are our main results.

Theorem 2.4 *(Wellposedness). Suppose that $R \sim P := \mu \otimes \nu$ and that the density dR/dP admits a continuous version. Then there exists a unique coupling $\pi \in \Pi(\mu, \nu)$ such that (π, R) is cyclically invariant. If (2.1) is finite, π is its minimizer.*

Theorem 2.5 *(Stability). For $n \geq 1$, consider $(\mu_n, \nu_n) \in \mathcal{P}(\mathsf{X}) \times \mathcal{P}(\mathsf{Y})$, $R_n \in \mathcal{P}(\mathsf{X} \times \mathsf{Y})$ and $\pi_n \in \Pi(\mu_n, \nu_n)$. Let (π_n, R_n) be cyclically invariant and suppose that μ_n, ν_n, R_n converge weakly to some limits μ, ν, R, where $R \sim P := \mu \otimes \nu$. Writing $P_n := \mu_n \otimes \nu_n$, suppose also that for some versions $\frac{dR_n}{dP_n}, \frac{dR}{dP} : \mathsf{X} \times \mathsf{Y} \to (0, \infty)$ of the respective densities and some constants $\alpha_n > 0$, it holds that for any fixed $z \in \operatorname{spt} R$,*

$$\frac{dR_n}{dP_n}(z') = [1 + o(1)]\alpha_n \frac{dR}{dP}(z), \tag{2.4}$$

where $o(1)$ stands for a function $\phi_z(z', n) \to 0$ as $d(z', z) + 1/n \to 0$. Then π_n converges weakly to a limit $\pi \sim R$ and (π, R) is cyclically invariant.

Schrödinger bridges are closely related to so-called Schrödinger systems (also called Schrödinger equations). For instance, Theorem 2.4 entails the following wellposedness result.

Corollary 2.6 *(Schrödinger system). Let $(\mu, \nu) \in \mathcal{P}(\mathsf{X}) \times \mathcal{P}(\mathsf{Y})$ and let $f : \mathsf{X} \times \mathsf{Y} \to (0, \infty)$ be continuous with $\int_{\mathsf{X} \times \mathsf{Y}} f \, dP = 1$. There exist Borel functions $\varphi : \mathsf{X} \to (0, \infty)$ and $\psi : \mathsf{Y} \to (0, \infty)$ such that*

$$\int_{\mathsf{Y}} f(x, y)\psi(y) \, \nu(dy) = \varphi(x)^{-1}, \quad \int_{\mathsf{X}} f(x, y)\varphi(x) \, \mu(dx) = \psi(y)^{-1} \tag{2.5}$$

for μ-a.e. $x \in \mathsf{X}$ and ν-a.e. $y \in \mathsf{Y}$. The pair (φ, ψ) is a.s. unique up to a multiplicative constant.[2]

As above, the uniqueness follows from known results. Existence for continuous functions f such that f, f^{-1} are uniformly bounded was first shown in [11]. Under the boundedness condition alone, existence is due to [32]. Using the connection with Schrödinger bridges, [49] relaxed the boundedness to a condition of finite entropy, corresponding to the finiteness of (2.1) in our setting. We refer to [34] for a more complete review of the literature which dates back to Schrödinger. In Corollary 2.6, we reintroduce the continuity condition of [11] but avoid any condition of finite entropy or boundedness. Of course, the stability result of Theorem 2.5 also has an analogous corollary for Schrödinger systems.

3. Absolute continuity of limits

In this section we show that if (π_n, R_n) is cyclically invariant and $R_n \sim \mu_n \otimes \nu_n$ holds in a locally uniform sense (to be made precise), then any weak limit pair $(\pi, R) = (\lim_n \pi_n, \lim_n R_n)$ must satisfy $\pi \ll R$.

[2] I.e., any solution (φ', ψ') satisfies $\varphi' = a\varphi$ μ-a.s., $\psi' = a^{-1}\psi$ ν-a.s., for some $a > 0$.

As the method of proof is novel, we first sketch the line of argument. We shall be comparing the measures of rectangles $F_i \times G_i \subset \mathsf{X} \times \mathsf{Y}$ for $i = 1, 2$ with their permutations $F_i \times G_{i+1}$.[3] Consider first the trivial case $R_n = P_n := \mu_n \otimes \nu_n$, then cyclical invariance of (π_n, R_n) implies

$$\pi_n(F_1 \times G_1)\pi_n(F_2 \times G_2) = \pi_n(F_1 \times G_2)\pi_n(F_2 \times G_1). \tag{3.1}$$

This will of course no longer hold if $R_n \neq P_n$, but the equivalence $R_n \sim P_n$ suggests that the two sides of (3.1) should still be comparable. We will quantify the equivalence $R_n \sim P_n$ and assume it to hold uniformly in n. Then, we prove that the two sides of (3.1) are comparable in the sense that their quotient remains bounded uniformly in n. For suitable rectangles, the bound propagates to the weak limit (π, R), accomplishing the first step of the proof. The second step is to argue by contraposition that this bound implies $\pi \ll R$. Indeed, we establish that if $\pi \in \Pi(\mu, \nu)$ is singular wrt. $\mu \otimes \nu$, then there exist F_i, G_i such that $\pi(F_1 \times G_1)\pi(F_2 \times G_2)$ is above a threshold whereas $\pi(F_1 \times G_2)\pi(F_2 \times G_1)$ is arbitrarily small.

For ease of reference, we first record two measure-theoretic facts.

Lemma 3.1. *Let ρ be a σ-finite measure on a Polish space (Ω, d). We say that $C \subset \Omega$ is* ρ-continuous *if its boundary ∂C is a ρ-nullset.*

(i) ρ-continuous sets form a field; i.e., unions, intersections, complements, differences of ρ-continuous sets are again ρ-continuous.

(ii) For fixed $z \in \Omega$, the open ball $B_r(z) = \{z' : d(z, z') < r\}$ is ρ-continuous for all but countably many values of $r > 0$. In particular, given $r > 0$, there exists $0 < r' \leq r$ such that $B_{r'}(z)$ is ρ-continuous.

(iii) If $F \subset \mathsf{X}$ is μ-continuous and $G \subset \mathsf{Y}$ is ν-continuous, then $F \times G$ is π-continuous for any $\pi \in \Pi(\mu, \nu)$.

(iv) Given $A \in \mathcal{B}(\Omega)$ and $\varepsilon > 0$, there exists an open set $B \subset A_\varepsilon := \{d(\cdot, A) < \varepsilon\}$ with $\rho(\partial B) = 0$ and $\rho(A \Delta B) < \varepsilon$.

Proof. Statement (i) is verified directly; (ii) holds because a σ-finite measure admits at most countably many disjoint sets of positive measure; (iii) follows from $\partial(F \times G) = (\overline{F} \times \partial G) \cup (\partial F \times \overline{G})$. Let A, ε be as in (iv). By interior and exterior regularity of ρ, there is a compact set $K \subset A$ with $\rho(A \setminus K) < \varepsilon$ and an open set $A \subset O \subset A_\varepsilon$ with $\rho(O \setminus A) < \varepsilon$. We have $r(z) := d(z, O^c) > 0$ for all $z \in K$ by the closedness of O^c, and K is covered by the open balls $B_{r(z)}(z)$ with $z \in K$. After making $r(z)$ smaller if necessary, each of these balls is ρ-continuous. Choosing a finite cover $\{B_{r(z_i)}(z_i)\}_{i \leq N}$, the set $B := \cup_i B_{r(z_i)}(z_i)$ has the required properties. □

The second fact is a conditional version of the differentiation of measures, based on Assumption 2.3 for the marginal space X. While not widely known, this concept was

[3] We use the cyclical convention for $i \in \{1, 2\}$; that is, $i + 1 := 1$ for $i = 2$.

already established in [45], although the author defined differentiation of measures in a slightly different way. For the convenience of the reader, we detail the adaptation to our setting.

Lemma 3.2. *Let $\pi \in \mathcal{P}(\mathsf{X} \times \mathsf{Y})$ and let μ be its first marginal. Consider for $x \in \operatorname{spt} \mu$ and $r > 0$ the probability measure $\pi_x^{(r)} \in \mathcal{P}(\mathsf{Y})$ defined by*

$$\pi_x^{(r)}(C) = \frac{\pi(B_r(x) \times C)}{\mu(B_r(x))}, \quad C \in \mathcal{B}(\mathsf{Y}).$$

Under Assumption 2.3 on X, there exists $\mathsf{X}_0 \subset \operatorname{spt} \mu$ with $\mu(\mathsf{X}_0) = 1$ such that for all $x \in \mathsf{X}_0$, the weak limit

$$\pi_x := \lim_{r \to 0} \pi_x^{(r)}$$

exists. Moreover, π_x defines a regular conditional probability of π given x.

Proof. As X, Y are Polish, there exists some regular conditional probability $\hat{\pi}_x$; it suffices to show that $\lim_{r \to 0} \pi_x^{(r)} = \hat{\pi}_x$ weakly for μ-a.e. $x \in \mathsf{X}$. Fix a countable collection $\mathcal{C}$ of nonnegative bounded continuous test functions $\phi : \mathsf{Y} \to \mathbb{R}$ that determine weak convergence (cf. [43, Theorem 6.6, p. 47]). Then we need to show, for fixed $\phi \in \mathcal{C}$, that

$$\int \phi \, d\pi_x^{(r)} \to \int \phi \, d\hat{\pi}_x \quad \text{for } \mu\text{-a.e. } x \in \mathsf{X}. \tag{3.2}$$

As $\hat{\pi}_x$ is a regular conditional probability, it holds for μ-a.e. $x \in \mathsf{X}$ that

$$\begin{aligned} \int \phi \, d\pi_x^{(r)} &= \frac{1}{\mu(B_r(x))} \int_{B_r(x)} \mu(dx') \int \phi(y) \, \hat{\pi}_{x'}(dy) \\ &= \frac{1}{\mu(B_r(x))} \int_{B_r(x)} f \, d\mu, \qquad f(x') := \int \phi(y) \, \hat{\pi}_{x'}(dy). \end{aligned}$$

We now apply Assumption 2.3 to the pair $f \, d\mu \ll d\mu$ and deduce that the right-hand side converges to $f(x)$ for μ-a.e. $x \in \mathsf{X}$, which is (3.2). □

The next result is the main ingredient for the second step as sketched above: the sets $\{x_i\} \times U_i$ constructed in Lemma 3.3 will be "blown up" to rectangles in the proof of Proposition 3.5 and used to show by contraposition that $\pi \ll R$.

Lemma 3.3. *Let $\pi \in \Pi(\mu, \nu)$ and let $\pi = \mu(dx) \otimes \pi_x(dy)$ be a disintegration. If $\pi \not\ll \mu \otimes \nu$, then*

$$\mu^2\{(x_1, x_2) \in \mathsf{X}^2 : \exists U_1, U_2 \in \mathcal{B}(\mathsf{Y}) \text{ with } \pi_{x_i}(U_i) > 0, \pi_{x_i}(U_{i+1}) = 0\} > 0.$$

In addition, the sets U_i can be chosen to be disjoint, of arbitrarily small diameter, and such that $\pi_{x_i}(\partial U_j) = \nu(\partial U_j) = 0$ for $i, j \in \{1, 2\}$.

Proof. Let $\pi \not\ll \mu \otimes \nu$; that is, there exists a set $A \in \mathcal{B}(\mathsf{X} \times \mathsf{Y})$ with $\pi(A) > 0$ and $(\mu \otimes \nu)(A) = 0$. Let $A_x = \{y : (x, y) \in A\}$ denote the x-section, then $\nu(A_x) = 0$ for μ-a.e. $x \in \mathsf{X}$. On the other hand, any $B \in \mathcal{B}(\mathsf{Y})$ satisfies $\nu(B) = \int_{\mathsf{X}} \pi_{x'}(B)\, \mu(dx')$. In particular, $\nu(A_x) = 0$ implies that $\pi_{x'}(A_x) = 0$ for μ-a.e. $x' \in \mathsf{X}$. Therefore,

$$F = \{(x, x') \in \mathsf{X}^2 : \pi_{x'}(A_x) = 0\} \quad \text{has full measure } \mu^2.$$

As $\pi(A) > 0$, the set $W = \{x : \pi_x(A_x) > 0\}$ satisfies $\mu(W) > 0$ and hence $\mu^2((W \times W) \cap F) > 0$. For $(x_1, x_2) \in (W \times W) \cap F$ we have $\pi_{x_i}(A_{x_i}) > 0$ and $\pi_{x_i}(A_{x_{i+1}}) = 0$. In particular, the disjoint sets $U_i' := A_{x_i} \setminus A_{x_{i+1}}$ satisfy $\pi_{x_i}(U_i') > 0$ and $\pi_{x_i}(U_{i+1}') = 0$. By intersecting with a suitable ball, the diameter of U_i' can be assumed to be arbitrarily small. Finally, let $\rho = \pi_{x_1} + \pi_{x_2} + \nu$ and choose ρ-continuous sets U_i'' for U_i' as in Lemma 3.1 (iv), with $\varepsilon > 0$ small enough such that the sets $U_i := U_i'' \setminus U_{i+1}''$ have the required properties; cf. Lemma 3.1 (i). □

The next lemma establishes that the two sides of (3.1) are comparable with a bound related to the equivalence $R_n \sim P_n$. The following notation is useful: when $k \geq 1$ and a k-tuple $(z_1, \ldots, z_k) \in (\mathsf{X} \times \mathsf{Y})^k$ are given, and

$$\text{if} \quad z_i = (x_i, y_i) \in \mathsf{X} \times \mathsf{Y}, \quad \text{we set} \quad \bar{z}_i := (x_i, y_{i+1}), \tag{3.3}$$

with the cyclical convention $y_{k+1} := y_1$.

Lemma 3.4. *Let $\pi \in \Pi(\mu, \nu)$ and $\pi \ll R \sim P := \mu \otimes \nu$. Consider rectangles $A_i \in \mathcal{B}(\mathsf{X}) \times \mathcal{B}(\mathsf{Y})$ for $1 \leq i \leq k$ and denote $\bar{A}_1 \times \cdots \times \bar{A}_k := \{(\bar{z}_1, \ldots, \bar{z}_k) : z_i \in A_i\}$. For some $\alpha, \bar{\alpha} > 0$, suppose that $dR/dP \leq \alpha$ on A_i and $(dR/dP)^{-1} \leq \bar{\alpha}$ on $\bar{A}_i$, for all i. Then*

$$\pi^k(A_1 \times \cdots \times A_k) \leq (\alpha\bar{\alpha})^k \pi^k(\bar{A}_1 \times \cdots \times \bar{A}_k).$$

Proof. Set $Z = d\pi/dR$ and note that

$$P(dz_1) \cdots P(dz_k) = \mu(dx_1) \cdots \mu(dx_k)\nu(dy_1) \cdots \nu(dy_k) = P(d\bar{z}_1) \cdots P(d\bar{z}_k). \tag{3.4}$$

Using the cyclical invariance of (π, R) and the rectangular form of A_i,

$$\begin{aligned} &\pi^k(A_1 \times \cdots \times A_k) \\ &= \int_{A_1 \times \cdots \times A_k} Z(z_1) \cdots Z(z_k) \frac{dR}{dP}(z_1) \cdots \frac{dR}{dP}(z_k)\, P(dz_1) \cdots P(dz_k) \end{aligned}$$

$$
\begin{aligned}
&\leq \alpha^k \int\limits_{A_1\times\cdots\times A_k} Z(z_1)\cdots Z(z_k)\, P(dz_1)\cdots P(dz_k)\\
&= \alpha^k \int\limits_{A_1\times\cdots\times A_k} Z(\bar{z}_1)\cdots Z(\bar{z}_k)\, P(d\bar{z}_1)\cdots P(d\bar{z}_k)\\
&= \alpha^k \int\limits_{\bar{A}_1\times\cdots\times \bar{A}_k} Z(z_1)\cdots Z(z_k)\, P(dz_1)\cdots P(dz_k)\\
&\leq (\alpha\bar{\alpha})^k \int\limits_{\bar{A}_1\times\cdots\times \bar{A}_k} Z(z_1)\cdots Z(z_k)\, R(dz_1)\cdots R(dz_k)\\
&= (\alpha\bar{\alpha})^k \pi^k(\bar{A}_1\times\cdots\times\bar{A}_k). \quad \square
\end{aligned}
$$

We can now prove the main result of this section.

Proposition 3.5. *Let $(\mu_n,\nu_n)\in\mathcal{P}(\mathsf{X})\times\mathcal{P}(\mathsf{Y})$, let $\pi_n\in\Pi(\mu_n,\nu_n)$ and $R_n\sim P_n:=\mu_n\otimes\nu_n$. Suppose that (π_n,R_n) is cyclically invariant for each n and that π_n converges weakly to some limit π. In particular, μ_n,ν_n converge to some limits μ,ν, and $\pi\in\Pi(\mu,\nu)$. Set $P=\mu\otimes\nu$ and suppose that (R_n,P_n) are uniformly locally equivalent in the following sense: there are versions $\frac{dR_n}{dP_n}:\mathsf{X}\times\mathsf{Y}\to(0,\infty)$ of the relative densities and constants $\alpha_n>0$ such that given $z\in\operatorname{spt}P$, there exist $r=r(z)>0$ and $n_0=n_0(z,r)$ with*

$$
\sup_{z'\in B_r(z),\, n\geq n_0}\left(\alpha_n\frac{dR_n}{dP_n}(z')+\alpha_n^{-1}\frac{dP_n}{dR_n}(z')\right)<\infty. \tag{3.5}
$$

Then $\pi\ll P$.

Proof. Note that the weak convergence of $\pi_n\in\Pi(\mu_n,\nu_n)$ implies the weak convergence of its marginals to some limits μ and ν, and then $\pi\in\Pi(\mu,\nu)$. By Lemma 3.2, there is a set $\mathsf{X}_0\subset\operatorname{spt}\mu$ of full μ-measure such that for $x\in\mathsf{X}_0$, the weak limit

$$
\pi_x(\cdot):=\lim_{r\to 0}\frac{\pi(B_r(x)\times\cdot)}{\mu(B_r(x))}
$$

exists, and $\pi=\mu\otimes_x\pi_x$ is a disintegration of π. In particular, for any $x\in\mathsf{X}_0$ and any π_x-continuity set U,

$$
\pi_x(U)=\lim_{r\to 0}\frac{\pi(B_r(x)\times U)}{\mu(B_r(x))}. \tag{3.6}
$$

Suppose for contradiction that $\pi\not\ll\mu\otimes\nu$. Then Lemma 3.3 yields $x_i\in\mathsf{X}_0$ $(i=1,2)$ and disjoint sets U_i such that

$$
q:=\min_i\pi_i(U_i)>0 \quad\text{and}\quad \pi_i(U_{i+1})=0, \quad\text{where } \pi_i:=\pi_{x_i}.
$$

Moreover, given $r_0 > 0$, the sets U_i can be chosen to be $(\pi_1 + \pi_2 + \nu)$-continuous and contained in a ball $B_{r_0}(y_i)$ around some $y_i \in \operatorname{spt}\nu$. Using also (3.5), we can choose r_0, n_0 such that for some $\alpha_* > 0$,

$$\alpha_n \frac{dR_n}{dP_n} \le \alpha_* \quad \text{on } B_{r_0}(x_i) \times U_i, \qquad \alpha_n^{-1} \frac{dP_n}{dR_n} \le \alpha_* \quad \text{on } B_{r_0}(x_i) \times U_{i+1}, \tag{3.7}$$

for all $1 \le i \le k$ and $n \ge n_0$. Writing $B_{i,r} := B_r(x_i)$ for brevity, (3.6) implies in particular that

$$\frac{\pi(B_{i,r} \times U_i)}{\mu(B_{i,r})} \to \pi_i(U_i) \ge q, \quad \frac{\pi(B_{i,r} \times U_{i+1})}{\mu(B_{i,r})} \to \pi_i(U_{i+1}) = 0.$$

That is, given $\varepsilon > 0$, choosing $r \le r_0$ small enough results in

$$\frac{\pi(B_{i,r} \times U_i)}{\mu(B_{i,r})} \ge q - \varepsilon, \quad \frac{\pi(B_{i,r} \times U_{i+1})}{\mu(B_{i,r})} \le \varepsilon, \tag{3.8}$$

and we may further choose r such that $\mu(\partial B_{i,r}) = 0$. Specifically, we choose $\varepsilon > 0$ such that $\varepsilon/(q-\varepsilon) < \alpha_*^{-2}$, then (3.8) implies

$$\prod_{i=1,2} \pi(B_{i,r} \times U_i) > \alpha_*^4 \prod_{i=1,2} \pi(B_{i,r} \times U_{i+1}).$$

Note that $B_{i,r} \times U_j$ is a π-continuity set for $i, j \in \{1, 2\}$; cf. Lemma 3.1 (iii). In view of $\pi_n \to \pi$, it then follows that

$$\prod_{i=1,2} \pi_n(B_{i,r} \times U_i) > \alpha_*^4 \prod_{i=1,2} \pi_n(B_{i,r} \times U_{i+1}) \tag{3.9}$$

for n sufficiently large. On the other hand, we apply Lemma 3.4 with $k = 2$ and $A_i = B_{i,r} \times U_i$. In view of (3.7), the condition of the lemma holds with $\alpha := \alpha_n^{-1}\alpha_*$ and $\bar{\alpha} := \alpha_n \alpha_*$. Noting that the α_n cancel to yield $(\alpha\bar{\alpha})^k = \alpha_*^4$, the lemma yields the inequality opposite to (3.9), a contradiction. □

4. Cyclical invariance of limits

In this section we aim to show that limits of cyclically invariant couplings are again cyclically invariant, under suitable conditions. We proceed in two steps. First, we establish that limits are weakly cyclically invariant as defined below. Second, we show that weak cyclical invariance already implies cyclical invariance. This second step has little to do with the passage to the limit; rather, it settles some measure-theoretic aspects to get rid of pesky nullsets.

The "weak" notion is introduced mainly to disentangle the proof of the main result. It weakens in two ways the cyclical invariance of (π, R) as stated in Definition 2.1: the

equivalence of π and R is reduced to absolute continuity and the cyclical relation only holds for points from specific sets. We recall the notation $\bar{z}_i$ from (3.3).

Definition 4.1. Let $\pi \in \Pi(\mu,\nu)$ and $R \in \mathcal{P}(\mathsf{X}\times\mathsf{Y})$. We call (π, R) *weakly cyclically invariant* if $\pi \ll R \sim P := \mu\otimes\nu$ and there exist

(i) a version $Z : \mathsf{X}\times\mathsf{Y} \to [0,\infty]$ of the density $d\pi/dR$,
(ii) $\Omega_1, \Omega_0 \in \mathcal{B}(\mathsf{X}\times\mathsf{Y})$ with $\pi(\Omega_1) = R(\Omega_0) = 1$ and $0 < Z < \infty$ on Ω_1

such that for all $k \in \mathbb{N}$,

$$\prod_{i=1}^{k} Z(z_i) = \prod_{i=1}^{k} Z(\bar{z}_i) \quad \text{for all } (z_i)_{i=1}^k \subset \Omega_1 \text{ with } (\bar{z}_i)_{i=1}^k \subset \Omega_0.$$

Proposition 4.2. *Let $(\mu_n,\nu_n) \in \mathcal{P}(\mathsf{X})\times\mathcal{P}(\mathsf{Y})$, let $\pi_n \in \Pi(\mu_n,\nu_n)$ and $R_n \sim P_n := \mu_n\otimes\nu_n$. Suppose that (π_n, R_n) is cyclically invariant for each n and that π_n, R_n converge weakly to some limits π, R. In particular, μ_n, ν_n converge to some limits μ,ν, and $\pi \in \Pi(\mu,\nu)$. Suppose that $R \sim P := \mu\otimes\nu$ and that there are versions $f_n, f : \Omega \to (0,\infty)$ of the densities $\frac{dR_n}{dP_n}, \frac{dR}{dP}$ and constants $\alpha_n > 0$ such that for any fixed $z \in \operatorname{spt} R$,*

$$f_n(z') = [1 + o(1)]\alpha_n f(z), \tag{4.1}$$

where $o(1)$ stands for a function $\phi_z(z', n) \to 0$ as $d(z', z) + 1/n \to 0$. Then (π, R) is weakly cyclically invariant. Specifically, the quantities Ω_1, Ω_0, Z of Definition 4.1 can be chosen as

$$\Omega_0 = \left\{ z \in \operatorname{spt} R : \ Z(z) := \lim_{r\to 0} \frac{\pi(B_r(z))}{R(B_r(z))} \text{ exists in } [0,\infty) \right\}$$
$$\cap \left\{ z \in \operatorname{spt} P : \ f(z) = \lim_{r\to 0} \frac{R(B_r(z))}{P(B_r(z))} \right\}$$

and $\Omega_1 = \Omega_0 \cap \{Z > 0\}$.

Proof. Assumption 2.3 for the space $\mathsf{X}\times\mathsf{Y}$ shows that Z is a version of $d\pi/dR$ and $R(\Omega_0) = 1$. As (4.1) implies (3.5), Proposition 3.5 yields that $\pi \ll P$, hence $\pi \ll R$ and the definition of Ω_1 implies $\pi(\Omega_1) = 1$.

Let $z_1, \dots, z_k \in \Omega_1$ be such that $\bar{z}_i \in \Omega_0$ and consider for $r > 0$ the balls $A_i = A_i^{(r)} := B_r(z_i) = B_r(x_i) \times B_r(y_i)$. To avoid unwieldy formulas, we use the vector notation

$$\boldsymbol{z} = (z_1, \dots, z_k), \quad \bar{\boldsymbol{z}} = (\bar{z}_1, \dots, \bar{z}_k), \quad \boldsymbol{A} = A_1 \times \cdots \times A_k, \quad \bar{\boldsymbol{A}} = \{\bar{\boldsymbol{z}} : \boldsymbol{z} \in \boldsymbol{A}\}$$

together with the convention that functions and measures are evaluated by multiplication over the components, for instance

$$f(\boldsymbol{z}) = f(z_1)\cdots f(z_k), \quad \pi(\boldsymbol{A}) = \pi(A_1)\cdots\pi(A_k).$$

As $d\pi_n/dP_n = (d\pi_n/dR_n)(dR_n/dP_n)$, we then have

$$\pi_n(\boldsymbol{A}) = \int_{\boldsymbol{A}} Z_n(\boldsymbol{z}')\, f_n(\boldsymbol{z}')\, P_n(d\boldsymbol{z}'), \quad Z_n := d\pi_n/dR_n.$$

We also have $\pi_n(A_i) > 0$ for n large as $Z(z_i) > 0$ and $\pi_n \to \pi$. On the other hand, $Z_n(\boldsymbol{z}')$ and $P_n(d\boldsymbol{z}')$ are invariant under $\boldsymbol{z}' \mapsto \bar{\boldsymbol{z}}'$ due to the assumption on π_n and the form of P_n; cf. (3.4). Thus

$$\pi_n(\bar{\boldsymbol{A}}) = \int_{\bar{\boldsymbol{A}}} Z_n(\boldsymbol{z}')\, f_n(\boldsymbol{z}')\, P_n(d\boldsymbol{z}') = \int_{\boldsymbol{A}} Z_n(\boldsymbol{z}')\, f_n(\bar{\boldsymbol{z}}')\, P_n(d\boldsymbol{z}').$$

Applying the assumption on f_n to each of the points $z_i, \bar{z}_i$ then yields

$$\begin{aligned}\frac{\pi_n(\bar{\boldsymbol{A}})}{\pi_n(\boldsymbol{A})} &= \frac{\int_{\boldsymbol{A}} Z_n(\boldsymbol{z}')\, f_n(\bar{\boldsymbol{z}}')\, P_n(d\boldsymbol{z}')}{\int_{\boldsymbol{A}} Z_n(\boldsymbol{z}')\, f_n(\boldsymbol{z}')\, P_n(d\boldsymbol{z}')} \\ &= \frac{[1+o(1)]\alpha_n f(\bar{\boldsymbol{z}}) \int_{\boldsymbol{A}} Z_n(\boldsymbol{z}')\, P_n(d\boldsymbol{z}')}{[1+o(1)]\alpha_n f(\boldsymbol{z}) \int_{\boldsymbol{A}} Z_n(\boldsymbol{z}')\, P_n(d\boldsymbol{z}')} = [1+o(1)]\frac{f(\bar{\boldsymbol{z}})}{f(\boldsymbol{z})}\end{aligned}$$

where $o(1)$ stands for a function of $(\boldsymbol{z}, \bar{\boldsymbol{z}})$ converging to zero as $r + 1/n \to 0$. For values of $r > 0$ such that $A_i, \bar{A}_i$ are continuity sets of P (and hence also of π and R), taking $n \to \infty$ yields

$$\frac{\pi(\bar{\boldsymbol{A}})}{\pi(\boldsymbol{A})} = [1+o(1)]\frac{f(\bar{\boldsymbol{z}})}{f(\boldsymbol{z})}. \tag{4.2}$$

On the other hand, $z_i, \bar{z}_i \in \Omega_0$ also guarantees that $\frac{R(A_i)}{P(A_i)} = [1+o(1)]f(z_i)$ and similarly for $\bar{z}_i$. Recalling $P(\bar{\boldsymbol{A}}) = P(\boldsymbol{A})$, we deduce

$$\frac{R(\bar{\boldsymbol{A}})}{R(\boldsymbol{A})} = \frac{R(\bar{\boldsymbol{A}})/P(\bar{\boldsymbol{A}})}{R(\boldsymbol{A})/P(\boldsymbol{A})} = [1+o(1)]\frac{f(\bar{\boldsymbol{z}})}{f(\boldsymbol{z})}. \tag{4.3}$$

Combining (4.2) and (4.3) yields

$$\frac{\pi(\boldsymbol{A})}{R(\boldsymbol{A})} = [1+o(1)]\frac{\pi(\bar{\boldsymbol{A}})}{R(\bar{\boldsymbol{A}})}$$

and then letting $r \to 0$ (along a sequence of r such that $A_i, \bar{A}_i$ are continuity sets of P) shows $Z(\boldsymbol{z}) = Z(\bar{\boldsymbol{z}})$, as desired. □

Remark 4.3. Assumption (4.1) on f_n, f in Proposition 4.2 is a sufficient condition for

$$\lim_{r\to 0}\lim_{n\to\infty}\frac{\int_{\boldsymbol{A}} Z_n(\boldsymbol{z}')\, f_n(\bar{\boldsymbol{z}}')\, P_n(d\boldsymbol{z}')}{\int_{\boldsymbol{A}} Z_n(\boldsymbol{z}')\, f_n(\boldsymbol{z}')\, P_n(d\boldsymbol{z}')} = \frac{f(\bar{\boldsymbol{z}})}{f(\boldsymbol{z})}; \tag{4.4}$$

it can be replaced by any other condition implying (4.4). We note that (4.4) can be seen as a differentiation of measures intertwined with a weak limit.

As mentioned above, the second step is to upgrade the weak cyclical invariance. Some of these considerations are similar to arguments in the proofs of [3], where it is shown by variational arguments that minimizers of certain static Schrödinger bridge problems admit a factorization.

Proposition 4.4. *Let $\pi \in \Pi(\mu,\nu)$ and let $R \in \mathcal{P}(\mathsf{X}\times\mathsf{Y})$ satisfy $R \sim P := \mu\otimes\nu$. The following are equivalent:*

(i) (π, R) *is cyclically invariant,*
(ii) (π, R) *is weakly cyclically invariant,*
(iii) there exist Borel functions $\varphi : \mathsf{X}\to(0,\infty)$ and $\psi : \mathsf{Y}\to(0,\infty)$ such that $(x,y)\mapsto \varphi(x)\psi(y)$ is a version of the density $d\pi/dR$.

The implications $(i)\Rightarrow(ii)$ and $(iii)\Rightarrow(i)$ are immediate. The fact that $(i)\Rightarrow(iii)$ is also well known, cf. [12], but will not be used directly. For the proof of $(ii)\Rightarrow(iii)$, a measure-theoretic fact will be useful. Given a set $A\subset\mathsf{X}\times\mathsf{Y}$ with $P(A)=1$, Lemma 4.5 below states that within any set B of positive measure we can find a point (x_*,y_*) which acts like an airline hub for A: any two points of A are connected through (x_*,y_*), modulo marginal nullsets. In particular, any trip can be achieved with at most one stopover, and we may stop within B. (The bound of one is optimal as the set A need not be a rectangle; in fact, A may fail to contain any measurable rectangle of positive measure [22, Exercise 5.4, p. 74].) Lemma 4.5 is refinement of [6, Lemma 4.3] which asserts the connectedness of A in the sense of [12]—in our analogy, connectedness means that any trip between two points of A can be achieved with finitely many stopovers at some points in A.

Lemma 4.5. *Let $P=\mu\otimes\nu$ and let $A,B\subset\mathsf{X}\times\mathsf{Y}$ be Borel sets with $P(B)>0$ and $P(A)=1$. There are Borel sets $\mathsf{X}_0\subset\mathsf{X}$ and $\mathsf{Y}_0\subset\mathsf{Y}$ with $\mu(\mathsf{X}_0)=\nu(\mathsf{Y}_0)=1$ such that setting $A_0 := A\cap(\mathsf{X}_0\times\mathsf{Y}_0)$ and $B_0 := B\cap(\mathsf{X}_0\times\mathsf{Y}_0)$, there exists a point*

$$(x_*,y_*)\in B_0 \text{ such that } (x,y_*),(x_*,y)\in A_0 \text{ for any } (x,y)\in\mathsf{X}_0\times\mathsf{Y}_0.$$

Proof. Let $C_x=\{y:(x,y)\in C\}$ denote the section of a set $C\subset\mathsf{X}\times\mathsf{Y}$ at $x\in\mathsf{X}$, and analogously for $y\in\mathsf{Y}$. Let $\mathsf{X}_1=\{x\in\mathsf{X}:\nu(A_x)=1\}$. In view of Fubini's theorem,

$P(A) = 1$ implies $\mu(\mathsf{X}_1) = 1$. Similarly, $P(B) > 0$ implies that $\{x \in \mathsf{X} : \nu(B_x) > 0\}$ has positive μ-measure. In particular, there exists a point $x_* \in \mathsf{X}_1$ with $\nu(B_{x_*}) > 0$.

Next, let $\mathsf{Y}_0 = \{y \in \mathsf{Y} : \mu(A_y) = 1\} \cap A_{x_*}$. Then again $\nu(\mathsf{Y}_0) = 1$, and in particular there exists a point $y_* \in \mathsf{Y}_0 \cap B_{x_*}$. Moreover, the set $\mathsf{X}_0 := \mathsf{X}_1 \cap A_{y_*}$ satisfies $\mu(\mathsf{X}_0) = 1$. By passing to Borel subsets of full measure, we may assume that $\mathsf{X}_0, \mathsf{Y}_0$ are themselves Borel.

Writing $A_0 = A \cap (\mathsf{X}_0 \times \mathsf{Y}_0)$ and $B_0 = B \cap (\mathsf{X}_0 \times \mathsf{Y}_0)$, we have by construction that $(x_*, y_*) \in B_0$ satisfies $(x_*, y) \in A_0$ for all $y \in \mathsf{Y}_0$ and $(x, y_*) \in A_0$ for all $x \in \mathsf{X}_0$. □

Proof of Proposition 4.4 $(ii) \Rightarrow (iii)$**.** Let Z, Ω_1, Ω_0 be as in Definition 4.1. Our aim is to find Borel functions $\varphi : \mathsf{X} \to (0, \infty)$ and $\psi : \mathsf{Y} \to (0, \infty)$ such that $Z'(x, y) := \varphi(x)\psi(y)$ defines a version of the density $d\pi/dR$. It is sufficient to construct φ on a Borel set $\mathsf{X}_0 \subset \mathsf{X}$ of full marginal measure, as we may then extend φ by setting $\varphi = 1$ on $\mathsf{X} \setminus \mathsf{X}_0$, and similarly for ψ. In particular, we may assume that $\mathrm{proj}_{\mathsf{X}} \Omega_1 = \mathsf{X}$ and $\mathrm{proj}_{\mathsf{Y}} \Omega_1 = \mathsf{Y}$.

Noting that $P(\Omega_1) > 0$ due to $\pi \ll P$, we can apply Lemma 4.5 to Ω_1 and Ω_0, and in view of the above observation, we may assume that $\mathsf{X}_0 = \mathsf{X}$ and $\mathsf{Y}_0 = \mathsf{Y}$ in its assertion. We then obtain a point

$$(x_*, y_*) \in \Omega_1 \text{ such that } (x, y_*), (x_*, y) \in \Omega_0 \text{ for any } (x, y) \in \mathsf{X} \times \mathsf{Y}.$$

Define $\varphi(x_*) := a > 0$ as an arbitrary number and $\psi(y_*) := Z(x_*, y_*)/\varphi(x_*)$. Given $(x, y) \in \mathsf{X} \times \mathsf{Y}$, we have $(x_*, y) \in \Omega_0$ and $(x, y_*) \in \Omega_0$, allowing us to define

$$\psi(y) := Z(x_*, y)/\varphi(x_*) \in (0, \infty), \quad \varphi(x) := Z(x, y_*)/\psi(y_*) \in (0, \infty).$$

The fact that Z is Borel readily implies that φ, ψ are Borel. Define $Z'(x, y) := \varphi(x)\psi(x)$ for $(x, y) \in \mathsf{X} \times \mathsf{Y}$. Clearly Z' is Borel and takes values in $(0, \infty)$. Let $(x, y) \in \Omega_1$. Then $(x, y_*), (x_*, y) \in \Omega_0$ and weak cyclical invariance yields

$$\varphi(x)\psi(y) = \frac{Z(x, y_*)}{\psi(y_*)} \frac{Z(x_*, y)}{\varphi(x_*)} = \frac{Z(x, y_*)Z(x_*, y)}{Z(x_*, y_*)} = Z(x, y).$$

That is, $Z' = Z$ on Ω_1. To prove the same relation on Ω_0, consider $(x, y) \in \Omega_0$. There are $x' \in \mathsf{X}$ and $y' \in \mathsf{Y}$ such that $z_1 := (x, y')$ and $z_2 := (x', y)$ are in Ω_1, and of course we also have $z_3 := (x_*, y_*) \in \Omega_1$. Thus the already established fact that $Z' = Z$ on Ω_1 implies

$$\varphi(x)\psi(y) = \frac{\varphi(x)\psi(y')\ \varphi(x')\psi(y)\ \varphi(x_*)\psi(y_*)}{\varphi(x')\psi(y_*)\ \varphi(x_*)\psi(y')} = \frac{Z(x, y')Z(x', y)Z(x_*, y_*)}{Z(x', y_*)Z(x_*, y')}.$$

On the other hand, $\bar{z}_1 = (x, y)$ and $\bar{z}_2 = (x', y_*)$ and $\bar{z}_3 = (x_*, y')$ are all in Ω_0, so that the invariance yields

$$0 < Z(x, y')Z(x', y)Z(x_*, y_*) = Z(x, y)Z(x', y_*)Z(x_*, y').$$

As a result, $Z'(x,y) = \varphi(x)\psi(y) = Z(x,y)$, showing $Z' = Z$ on Ω_0. Recalling $R(\Omega_0) = 1$, it follows that Z' is again a version of $d\pi/dR$. □

5. Proof of main results and ramifications

For ease of reference, we first summarize some known results.

Lemma 5.1. *Let $(\mu,\nu) \in \mathcal{P}(\mathsf{X}) \times \mathcal{P}(\mathsf{Y})$, $R \in \mathcal{P}(\mathsf{X}\times\mathsf{Y})$ and $R \sim P := \mu\otimes\nu$.*

(i) If the static Schrödinger bridge problem (2.1) is finite, it admits a unique minimizer $\pi \in \Pi(\mu,\nu)$. Moreover, (π, R) is cyclically invariant.
(ii) Let $\pi \in \Pi(\mu,\nu)$. If (π, R) is cyclically invariant and (2.1) is finite, then π is its minimizer.
(iii) There exists at most one $\pi \in \Pi(\mu,\nu)$ such that (π, R) is cyclically invariant.

Proof. Recall that cyclical invariance of (π, R) is equivalent a factorization of the density $d\pi/dR$ into strictly positive functions; cf. Proposition 4.4 or [12]. Taking that into account, (i) and (ii) can be found in [40, Theorem 2.1] in the stated generality. (The original results are due to [12,13,17,25,50], among others.) Finally, (iii) follows from (ii), as was also noted in [40, Corollary 2.9]: if $\pi, \pi' \in \Pi(\mu,\nu)$ have positive densities $d\pi/dR$, $d\pi'/dR$ admitting factorizations, then $d\pi/d\pi'$ also admits a factorization and now (ii), applied with π' as reference measure, implies that π is the unique minimizer of $H(\cdot|\pi')$. As $\pi' \in \Pi(\mu,\nu)$ is itself a coupling, this minimizer is π'. □

Proof of Theorems 1.3, 1.4, 2.4 and 2.5. Given data as in Theorem 2.5, the sequences (μ_n) and (ν_n) are tight, which readily implies the tightness of (π_n); cf. [52, Lemma 4.4, p. 44]. In view of Propositions 4.2 and 4.4, any cluster point π is such that (π, R) is cyclically invariant. The uniqueness of cyclically invariant couplings, see Lemma 5.1 (iii), shows that all cluster points coincide and hence that the original sequence (π_n) converges. This proves Theorem 2.5.

To deduce Theorem 1.4 from Theorem 2.5, we choose the reference measure as in (2.2); i.e.,

$$\frac{dR_n}{dP_n} = a_n e^{-c_n/\varepsilon_n}, \qquad \frac{dR}{dP} = a e^{-c/\varepsilon},$$

where a_n, a are the normalizing constants. Combining the uniform convergence $c_n/\varepsilon_n \to c/\varepsilon$ on bounded sets with the continuity of c, we see that (2.4) holds, for instance with $\alpha_n = a_n/a$.

The uniqueness part of Theorem 2.4, as well as its last assertion, is stated in Lemma 5.1. To deduce existence from Theorem 2.5, we consider the constant marginals $(\mu_n,\nu_n) := (\mu,\nu)$ and define approximating reference measures R_n via

$$\frac{dR_n}{dP} = a_n \left(\frac{dR}{dP} \vee \frac{1}{n} \right),$$

where a_n is the (finite) normalizing constant. As $\frac{dR}{dP}$ is continuous and positive, hence bounded away from zero on small balls, the condition (2.4) is satisfied with $P_n = P$ and a function $o(1)$ independent of n. The static Schrödinger bridge problem (2.1) for R_n falls into the classical setting of Lemma 5.1 (i) because the product coupling $\pi_0 := \mu \otimes \nu$ satisfies $H(\pi_0|R_n) < \infty$. In particular, the associated cyclically invariant couplings $\pi_n \in \Pi(\mu, \nu)$ exist, and now Theorem 2.5 implies the existence of $\pi \in \Pi(\mu, \nu)$ such that (π, R) is cyclically invariant. Finally, Theorem 1.3 is a direct consequence of Theorem 2.4 via (2.2). □

Proof of Corollary 2.6. Define $R \in \mathcal{P}(\mathsf{X} \times \mathsf{Y})$ by $dR/dP := f$ and let π be as in Theorem 2.4. As (π, R) is cyclically invariant, a version of the relative density admits a factorization $d\pi/dR(x, y) = \varphi(x)\psi(y)$ into positive Borel functions; cf. Proposition 4.4. The fact that π has marginals (μ, ν) then translates to the fact that (φ, ψ) solve the Schrödinger system. Uniqueness of φ, ψ up to a constant follows from the uniqueness of π (here the fact that $R \sim P$ is particularly important). □

Remark 5.2. Assumption 2.3 can be replaced by the assumption that the marginals (X, μ) and (Y, ν) satisfy the so-called doubling property. The latter assumption is structurally different as it refers to the specific measures rather than the metric spaces. Indeed, let (X, μ) be doubling; i.e., there exist $C > 0$ such that

$$\mu(B_{2r}(x)) \leq C\mu(B_r(x))$$

for any ball $B_r(x) \subset \mathsf{X}$. This ensures that differentiation of measures (in the sense of Assumption 2.3) holds for measures $\rho \ll \mu$; cf. [31, Theorem 1.8, p. 4]. In particular, Lemma 3.2 holds as stated, and then so does Proposition 3.5. If (Y, ν) is also doubling, then so is $(\mathsf{X} \times \mathsf{Y}, P)$ where $P = \mu \otimes \nu$, showing that differentiation wrt. P holds for measures $\rho \ll P$, in particular for $\rho := R \sim P$ in the context of Proposition 4.2. As $dR/dP > 0$, it follows that differentiation also holds wrt. R. This ensures that the proof of Proposition 4.2 remains valid, and hence the main results.

5.1. Instability for discontinuous costs

We show that stability of entropic optimal transport fails as soon as the cost function has an "essential" discontinuity. To see why the qualifier is necessary, consider a cost function c of the form

$$c(x, y) = \hat{c}(x, y) - f(x) - g(y)$$

for some possibly discontinuous functions $f : \mathsf{X} \to \mathbb{R}$, $g : \mathsf{Y} \to \mathbb{R}$ and a continuous function $\hat{c}$. It is immediate from Definition 1.1 that the marginal functions f and g do

not affect cyclical invariance. The cost c is equivalent to $\hat{c}$ from this perspective; in terms of optimal transport, the interpretation is that f, g change the cost of any coupling by the same constant, and therefore do not change the optimizer. As a result, a discontinuity in c can only be relevant for stability if c cannot be written in the above form with a continuous $\hat{c}$, which is the condition below.

Proposition 5.3. *Let c be a cost function such that $(x, y) \mapsto c(x, y) - f(x) - g(y)$ is discontinuous for any functions $f : \mathsf{X} \to \mathbb{R}$ and $g : \mathsf{Y} \to \mathbb{R}$. Then the stability of entropic optimal transport fails for c. That is, there are marginals $(\mu_n, \nu_n) \to (\mu, \nu)$ and $(c, 1)$-cyclically invariant couplings $\pi_n \in \Pi(\mu_n, \nu_n)$ with $\pi_n \to \pi \in \Pi(\mu, \nu)$ where π is not $(c, 1)$-cyclically invariant.*

Proof. Fix arbitrary $(x_0, y_0) \in \mathsf{X} \times \mathsf{Y}$. As $(x, y) \mapsto c(x, y) - c(x, y_0) - c(x_0, y)$ is discontinuous, there is a sequence $(x_n, y_n) \to (x_\infty, y_\infty)$ such that

$$c(x_n, y_n) - c(x_n, y_0) - c(x_0, y_n) \not\to c(x_\infty, y_\infty) - c(x_\infty, y_0) - c(x_0, y_\infty). \tag{5.1}$$

Consider the marginals $\mu_n = (\delta_{x_0} + \delta_{x_n})/2$ and $\nu_n = (\delta_{y_0} + \delta_{y_n})/2$. Let $\pi_n \in \Pi(\mu_n, \nu_n)$ be $(c, 1)$-cyclically invariant; that is,

$$\frac{\pi_n(x_0, y_0)\pi_n(x_n, y_n)}{\pi_n(x_0, y_n)\pi_n(x_n, y_0)} = \exp\big[c(x_0, y_n) + c(x_n, y_0) - c(x_0, y_0) - c(x_n, y_n)\big]. \tag{5.2}$$

After passing to a subsequence, π_n converge weakly to a coupling $\pi \in \Pi(\mu, \nu)$ of $\mu := (\delta_{x_0} + \delta_{x_\infty})/2$ and $\nu := (\delta_{y_0} + \delta_{y_\infty})/2$. Suppose for contradiction that π is cyclically invariant, then

$$\frac{\pi(x_0, y_0)\pi(x_\infty, y_\infty)}{\pi(x_0, y_\infty)\pi(x_\infty, y_0)} = \exp\big[c(x_0, y_\infty) + c(x_\infty, y_0) - c(x_0, y_0) - c(x_\infty, y_\infty)\big]. \tag{5.3}$$

As the left-hand side of (5.2) converges to the left-hand side of (5.3), the convergence of the right-hand sides follows, contradicting (5.1). □

References

[1] D. Alvarez-Melis, T. Jaakkola, Gromov-Wasserstein alignment of word embedding spaces, in: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, 2018, pp. 1881–1890.
[2] M. Arjovsky, S. Chintala, L. Bottou, Wasserstein generative adversarial networks, in: Proceedings of Machine Learning Research, vol. 70, 2017, pp. 214–223.
[3] J. Backhoff-Veraguas, M. Beiglböck, G. Conforti, A non-linear monotonicity principle and application to the Schrödinger problem, preprint, arXiv:2101.09975v1, 2021.
[4] J. Backhoff-Veraguas, M. Beiglböck, G. Pammer, Existence, duality, and cyclical monotonicity for weak transport costs, Calc. Var. Partial Differ. Equ. 58 (6) (2019) 203.
[5] M. Beiglböck, A.M.G. Cox, M. Huesmann, Optimal transport and Skorokhod embedding, Invent. Math. 208 (2) (2017) 327–400.

[6] M. Beiglböck, M. Goldstern, G. Maresch, W. Schachermayer, Optimal and better transport plans, J. Funct. Anal. 256 (6) (2009) 1907–1927.
[7] M. Beiglböck, N. Juillet, On a problem of optimal transport under marginal martingale constraints, Ann. Probab. 44 (1) (2016) 42–106.
[8] M. Beiglböck, M. Nutz, F. Stebegg, Fine properties of the optimal Skorokhod embedding problem, J. Eur. Math. Soc. (JEMS) 24 (4) (2022) 1389–1429.
[9] R.J. Berman, The Sinkhorn algorithm, parabolic optimal transport and geometric Monge-Ampère equations, Numer. Math. 145 (4) (2020) 771–836.
[10] E. Bernton, P. Ghosal, M. Nutz, Entropic optimal transport: geometry and large deviations, Duke Math. J. (2022), in press.
[11] A. Beurling, An automorphism of product measures, Ann. Math. (2) 72 (1960) 189–200.
[12] J.M. Borwein, A.S. Lewis, Decomposition of multivariate functions, Can. J. Math. 44 (3) (1992) 463–482.
[13] J.M. Borwein, A.S. Lewis, R.D. Nussbaum, Entropy minimization, *DAD* problems, and doubly stochastic kernels, J. Funct. Anal. 123 (2) (1994) 264–307.
[14] G. Carlier, V. Duval, G. Peyré, B. Schmitzer, Convergence of entropic schemes for optimal transport and gradient flows, SIAM J. Math. Anal. 49 (2) (2017) 1385–1418.
[15] V. Chernozhukov, A. Galichon, M. Hallin, M. Henry, Monge-Kantorovich depth, quantiles, ranks and signs, Ann. Stat. 45 (1) (2017) 223–256.
[16] G. Conforti, L. Tamanini, A formula for the time derivative of the entropic cost and applications, J. Funct. Anal. 280 (11) (2021) 108964.
[17] I. Csiszár, I-divergence geometry of probability distributions and minimization problems, Ann. Probab. 3 (1975) 146–158.
[18] M. Cuturi, O. Teboul, J.-P. Vert, Differentiable ranking and sorting using optimal transport, in: Advances in Neural Information Processing Systems, vol. 32, 2019.
[19] N. Deb, B. Sen, Multivariate rank-based distribution-free nonparametric testing using measure transportation, preprint, arXiv:1909.08733v1, 2019.
[20] E. del Barrio, J.A. Cuesta-Albertos, M. Hallin, C. Matrán, Distribution and quantile functions, ranks and signs in dimension d: a measure transportation approach, Ann. Stat. 49 (2) (2021) 1139–1165.
[21] L.C. Evans, R.F. Gariepy, Measure Theory and Fine Properties of Functions, revised edition, Textbooks in Mathematics, CRC Press, Boca Raton, FL, 2015.
[22] K.J. Falconer, The Geometry of Fractal Sets, Cambridge Tracts in Mathematics, vol. 85, Cambridge University Press, Cambridge, 1986.
[23] H. Federer, Geometric Measure Theory, Grundlehren der mathematischen Wissenschaften, vol. 153, Springer-Verlag, New York, 1969.
[24] H. Föllmer, Random fields and diffusion processes, in: École d'Été de Probabilités de Saint-Flour XV–XVII, 1985–1987, in: Lecture Notes in Math., vol. 1362, Springer, Berlin, 1988, pp. 101–203.
[25] H. Föllmer, N. Gantert, Entropy minimization and Schrödinger processes in infinite dimensions, Ann. Probab. 25 (2) (1997) 901–926.
[26] R. Fortet, Résolution d'un système d'équations de M. Schrödinger, J. Math. Pures Appl. 19 (1940) 83–105.
[27] A. Genevay, G. Peyré, M. Cuturi, Learning generative models with Sinkhorn divergences, in: Proceedings of the 21st International Conference on Artificial Intelligence and Statistics, in: PMLR, 2018, pp. 1608–1617.
[28] P. Ghosal, B. Sen, Multivariate ranks and quantiles using optimal transport: consistency, rates and nonparametric testing, Ann. Stat. 50 (2) (2022) 1012–1037.
[29] N. Gigli, L. Tamanini, Second order differentiation formula on $\mathrm{RCD}^*(K,N)$ spaces, J. Eur. Math. Soc. (JEMS) 23 (5) (2021) 1727–1795.
[30] Z. Harchaoui, L. Liu, S. Pal, Asymptotics of entropy-regularized optimal transport via chaos decomposition, preprint, arXiv:2011.08963v1, 2020.
[31] J. Heinonen, Lectures on Analysis on Metric Spaces, Universitext, Springer-Verlag, New York, 2001.
[32] C. Hobby, R. Pyke, Doubly stochastic operators obtained from positive operators, Pac. J. Math. 15 (1965) 153–157.
[33] C. Léonard, From the Schrödinger problem to the Monge-Kantorovich problem, J. Funct. Anal. 262 (4) (2012) 1879–1920.
[34] C. Léonard, A survey of the Schrödinger problem and some of its connections with optimal transport, Discrete Contin. Dyn. Syst. 34 (4) (2014) 1533–1574.
[35] C. Léonard, Revisiting Fortet's proof of existence of a solution to the Schrödinger system, preprint, arXiv:1904.13211v1, 2019.

[36] R.J. McCann, Existence and uniqueness of monotone measure-preserving maps, Duke Math. J. 80 (2) (1995) 309–323.
[37] G. Mena, J. Niles-Weed, Statistical bounds for entropic optimal transport: sample complexity and the central limit theorem, Adv. Neural Inf. Process. Syst. 32 (2019) 4541–4551.
[38] T. Mikami, Optimal control for absolutely continuous stochastic processes and the mass transportation problem, Electron. Commun. Probab. 7 (2002) 199–213.
[39] T. Mikami, Monge's problem with a quadratic cost by the zero-noise limit of h-path processes, Probab. Theory Relat. Fields 129 (2) (2004) 245–260.
[40] M. Nutz, Introduction to Entropic Optimal Transport, Lecture Notes, Columbia University, 2021, https://www.math.columbia.edu/~mnutz/docs/EOT_lecture_notes.pdf.
[41] M. Nutz, J. Wiesel, Entropic optimal transport: convergence of potentials, Probab. Theory Relat. Fields (2021), https://doi.org/10.1007/s00440-021-01096-8, in press.
[42] S. Pal, On the difference between entropic cost and the optimal transport cost, preprint, arXiv: 1905.12206v1, 2019.
[43] K.R. Parthasarathy, Probability Measures on Metric Spaces, Probability and Mathematical Statistics, vol. 3, Academic Press, New York, 1967.
[44] G. Peyré, M. Cuturi, Computational optimal transport: with applications to data science, Found. Trends Mach. Learn. 11 (5–6) (2019) 355–607.
[45] J. Pfanzagl, Conditional distributions as derivatives, Ann. Probab. 7 (6) (1979) 1046–1050.
[46] D. Preiss, Gaussian measures and covering theorems, Comment. Math. Univ. Carol. 20 (1) (1979) 95–99.
[47] D. Preiss, Gaussian measures and the density theorem, Comment. Math. Univ. Carol. 22 (1) (1981) 181–193.
[48] Y. Rubner, C. Tomasi, L.J. Guibas, The earth mover's distance as a metric for image retrieval, Int. J. Comput. Vis. 40 (2000) 99–121.
[49] L. Rüschendorf, W. Thomsen, Note on the Schrödinger equation and I-projections, Stat. Probab. Lett. 17 (5) (1993) 369–375.
[50] L. Rüschendorf, W. Thomsen, Closedness of sum spaces and the generalized Schrödinger problem, Teor. Veroyatn. Primen. 42 (3) (1997) 576–590.
[51] B. Schmitzer, Stabilized sparse scaling algorithms for entropy regularized transport problems, SIAM J. Sci. Comput. 41 (3) (2019) A1443–A1481.
[52] C. Villani, Optimal Transport, Old and New, Grundlehren der Mathematischen Wissenschaften, vol. 338, Springer-Verlag, Berlin, 2009.